The Expert Panel met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Michael Lieberman, Chair, presiding.

PRESENT:
MICHAEL LIEBERMAN, MD, MS, Chair, Oregon Health and Science University, Chair
HOWARD BREGMAN, MD, MS, Epic
ZAHID BUTT, MD, Medisolv
KERI CHRISTENSEN, American Medical Association
JOSEPH JENTZSCH, Kaiser Permanente
SAUL KRAVITZ, Mitre
JINGDONG LI, MD, Lantana Consulting
CATHERINE EIKEL MAJOR, MBA, Booz Allen Hamilton
RUTE MARTINS, MS, The Joint Commission
GINNY MEADOWS, RN, McKesson Corporation
J. MARC OVERHAGE, MD, PhD, Siemens Medical Solutions, Inc
MARTHA RADFORD, MD, NYU Langone Medical Center
SHANNON SIMS, MD, PhD, Rush University Medical Center
ALDO TINOCO, MD, MPH, National Committee for Quality Assurance
PAUL TANG, MD, MS, Palo Alto Medical Foundation
NQF STAFF:

HELEN BURSTIN, MD, MPH
BETH FRANKLIN, MS, RN
ANN HAMMERSMITH, JD

ROSEMARY KENNEDY, PhD, RN, MBA, FAAN
KATHRYN STREETER, MS
REVA WINKLER, MD, MPH

ALSO PRESENT:

ALYSSA CRAWFORD, Mathematica Policy Research
# Table of Contents

<table>
<thead>
<tr>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welcome by Michael Lieberman. .................</td>
</tr>
<tr>
<td>Introductions and Disclosures</td>
</tr>
<tr>
<td>by Ann Hammersmith .......................</td>
</tr>
<tr>
<td>Project Goals and Meeting Objectives</td>
</tr>
<tr>
<td>by Dr. Winkler ................................</td>
</tr>
<tr>
<td>Environmental Scan of Approaches to</td>
</tr>
<tr>
<td>eMeasure Feasibility Assessment ..........</td>
</tr>
<tr>
<td>Discussion: eMeasure Feasibility Assessment</td>
</tr>
<tr>
<td>Principles and Guidance for eMeasure Feasibility Assessment</td>
</tr>
<tr>
<td>Recommendations for eMeasure Feasibility</td>
</tr>
<tr>
<td>Scoring data element feasibility ........</td>
</tr>
<tr>
<td>NQF Member and Public Comment ..........</td>
</tr>
<tr>
<td>Adjourn .....................................</td>
</tr>
</tbody>
</table>
P-R-O-C-E-E-D-I-N-G-S

(8:40 a.m.)

DR. LIEBERMAN: Okay, I think we'll go ahead and get started. Welcome, everybody, to the NQF Panel on eMeasure Feasibility Assessment Expert Panel Meeting rescheduled from last, around Halloween, thanks to Superstorm Sandy.

I'm Mike Lieberman, I'll be the Chair today. And I'm going to really keep the introduction or welcome to a minimum here and turn it over to Ann to go through the disclosures. And then we'll go around the room for introductions. Thanks.

MS. HAMMERSMITH: Good morning, everyone. I'm Ann Hammersmith, I am NQF's general council. I see a few familiar faces so I think some of you know the drill about disclosures, but I'll review it briefly before we go around the table.

When we had the call for nominations, we asked people to fill out a
detail conflict of interests disclosure form.

So what we do at the beginning of these meetings is ask you to orally disclose.

We are not looking for you to tell us everything that's in your CV. Please don't, because we'll be here for a very long time.

We only ask you to mention things orally that you believe are relevant to what's before the committee today. Specifically we're interested in grants, research or speaking engagements that may be relevant to what the committee's going to look at today.

Also want to remind you that you serve as an individual on the committee. You don't represent your employer, you don't represent any group that may have nominated you or supported your nomination for the committee.

Sometimes committee members will say, I'm so and so and I'm here representing the American Society of fill in the blank, and
actually you are not. You are here as an individual expert; that's why we chose you for service on the committee.

The last thing that I want to remind you of is that disclosures don't necessarily turn on money changing hands. Sometimes people will say, I have no financial conflict of interest. Financial conflicts of interests are important, financial disclosures are important. But, because of the world that you all work in, often you'll serve as volunteers on committees and things of that nature. So we're also looking for you to disclose any relevant volunteer service as well.

So with that, we can go around the table and combine disclosures and introductions. Tell us who you are, who you're with and then if you have anything that you would like to disclose.

And Michael, since you're the Chair, you get to start.

DR. LIEBERMAN: All right, Mike
Lieberman from Oregon Health and Science University and I have no disclosures to make.

DR. KENNEDY: Rosemary Kennedy, Vice President of Health Information Technology at NQF.

MS. RUBINI: Juliet Rubini, Senior Project Manager, NQF.

MR. JENTZSCH: Joseph Jentzsch, I'm with Kaiser. I have no disclosures.

MS. MARTINS: Rute Martins with the Joint Commission, no disclosures.

MR. KRAVITZ: Saul Kravitz, Mitre, no disclosures.

DR. TANG: Paul Tang, Palo Alto Medical Foundation, no disclosures.

DR. LI: I'm JD Li from Lantana, no disclosures.

MS. MEADOWS: I'm Ginny Meadows from McKesson and I have no disclosures.

DR. SIMS: I'm Shannon Sims from Washington University Medical Center. I've been the primary investigator on sub-level of
grants and contracts for measure development organizations and CMS contracters, but I've not received any direct compensation for any of those.

DR. TINOCO: Good morning, Aldo Tinoco. I work with NCQA. I participate in work with developing quality measures under CMS-funded contracts and AHRQ-funded grants. Otherwise no disclosures.

MS. CHRISTENSEN: Hi, Keri Christensen from the AMA-PCPI, also grant work developing measures.

MS. MAJOR: Catherine Major with Booz Allen Hamilton and we do measure-development contracts with CMS and with ONC.

MS. JAVELLANA: Minet Javellana, CMS. No disclosures.

MS. KRAUSS: Good morning, Debbie Krauss, CMS. No disclosures.

DR. BREGMAN: Howard Bregman, I'm a physician and I work for Epic, the EHR vendor in Verona, Wisconsin. And I have no other
disclosures.

DR. RADFORD: I'm Martha Radford. I'm a physician. I am Chief Quality Officer at NYU. And I'd like to disclose that I'm a member of the ACC NCDR Management Board. That's volunteer, I have no financial disclosures which means I live below the Manhattan poverty line.

(Laughter.)

MS. FRANKLIN: I'm Beth Franklin, Senior Project Director at NQF. I can't remember.

DR. WINKLER: And I'm Reva Winkler, I'm the Senior Director here at NQF in Performance Measures.

MS. HAMMERSMITH: Okay, thank you, everyone, that was a world record for disclosures. You did very well. Do you have any questions or anything that you want to discuss with each other based on the disclosures this morning? Okay, thank you.

DR. WINKLER: Okay, thank you all
very much for being here today. We have a
couple of little introductory things before we
get into the meat of the conversation and let
you all talk it out.

It's important that we'll go through
the, sort of the project goals as well the
objectives for today's meeting. Next slide.

Just so everybody remembers, I think
we talked about this on the conference call:
the background for this effort. From NQF's
perspective, here in Performance Measures, we
look at all measures that come through for
potential endorsement for several criteria.

But two that are particularly
pertinent to what we're talking about today
are scientific acceptability. In other words
reliability, validity as well as feasibility.

And so that's going to be sort of
the focus of your discussion today. Several
previous efforts that NQF has done has touched
tangentially on these topics but, if anything,
probably prompted more questions then provided
And so those were our testing task force report that identified criteria for evaluating EHR measures for reliability and validity. And then there was, last spring, there was a draft proposal on how NQF would review and assess eMeasures that prompted comments from stakeholders who were identifying a need to have greater clarity and more direction around assessment for feasibility.

So the recommendation from the comment period on that effort was that NQF criteria should incorporate feasibility of data capture for the data elements utilized as well as looking at reliability and validity. So that is really the purpose of what we're here today to do, is to respond to that need out there as we're trying to develop and implement the EHR measures. Next one.

So specifically this project, to address those issues, has a couple of goals.
The first one is the environmental scan that Beth, I think, described in great detail, when we had our conference call on October 30th.

We asked a series of questions of EHR vendors, developers as well as some providers. Since that call, we've had some updates as well as we do want to hear some input from some of the measure developers on the Panel, to make the environmental scan a little more robust and be sure that we have representation from as wide a group of folks as possible.

The discussion memo that we sent to you earlier in the week is meant to be the starting point for your discussions to help us pull together what will ultimately be the deliverable from this effort, which is a series of recommendations.

We've given you some draft thoughts around principles and guidance for eMeasure feasibility assessment that should prompt the major part of today's discussion. And then
hopefully that will also lend itself to
identifying some starter set of criteria that
NQF could use to further look at feasibility
when we're reviewing eMeasures for potential
endorsement.

So those are the goals of this
particular project. Does anybody have any
questions about any of those? Okay, next one.

Okay, so our objectives today with
this particular meeting, in other words when
everybody starts to peel out and leave this
afternoon, what do we hoped to have done by
then, is having you review or discuss any
additions to the environmental scan. But also
the discussion to help us pull together what
should be the principles and guidance for
eMeasure feasibility that will become part of
the report, any additional recommendations.

During your conference call in
October you talked about a potential scoring
system, so talk about what that might look
like, as well as any starter set
recommendations for criteria for NQF to use when evaluating eMeasures.

So those are the discussion points and we'll be trying to maybe perhaps bring the conversation to some decisions around how we're going to characterize these different elements in the draft report deliverable that's the outcome of this project. Next slide.

And just the project timeline going forward from this point, from our meeting today, is as soon as we finish, our job here at NQF over the holidays will be to draft that report and pull together the discussion and the recommendations that you all are discussing and helping us formulate today.

After that the first couple weeks of January, happy new year to all, we'll ask you to review that draft report and give us your feedback and any revisions or suggestions.

After that, that report will go out for a 30-day public comment, which is likely
to generate some recommendations and comments.

And so after that we will spend some time
looking at those comments to make final
revisions to the draft report.

That draft report deliverable will
then go to both HITAC, CSAC and finally to
NQF's Board of Directors. So that's
essentially what we're looking to do over the
next couple of months as a result of this
project.

So we will be coming back to you on
a couple of occasions after this discussion,
as we try to pull the report together and be
sure that it represents the thinking and
response to stakeholder input around the topic
of eMeasure feasibility.

So any questions from anybody about
what we're doing, where we're going and what's
your role in the activities is going to be
going forward?

Okay, I'd like to just briefly let
my colleague, who arrived, introduce herself.
Helen?

DR. BURSTIN: Good morning, everybody. I'm Helen Burstin. I'm the Senior VP for performance measures. I arrived last night, as I suspect Howard did, from our meeting we were at in LA for the IOM.

So I got up at 1:30 in the morning, it was a little hard to get the children to accept the fact that, yes I've been home for 15 minutes and I'm leaving. So anyway, good morning, add my welcome, thank you all for coming and I think it will be a great day.

Thanks.

DR. WINKLER: We still have a couple of the other committee members who have yet to join us. Did anybody new join us who hasn't introduced themselves? I don't see anybody.

Okay, with that, I would like to turn it over to my colleague, Beth Franklin, to talk about just some updates to the environmental scan and we'll be asking some of our committee members to offer their
contributions as well.

MS. FRANKLIN: Good. Thanks, Reva.

Yes, so since we meet by conference call in October, we had two additional people, groups if you will, submit their environmental scan results. One was Abt Associates and the other was Siemens. They were in your packets and hopefully you've had a chance to read them.

And then we have three groups or three representatives today who are going to talk. And they're making my life easy because they're all sitting right in a row to my left.

So we're going to start with Katherine. Sorry, Katherine, to put you on the spot, to talk about -- Katherine Major from Booz Allen, to talk about measure development and what they're doing with measure development, what their role is.

MS. MAJOR: Sure, good morning, everyone. And we actually sat in a row on purpose, because we all knew that we'd been asked to speak a little bit about our
development and testing approaches and they're very similar. And so we'll kind of maybe tag-team it a little if that's all right.

So as I mentioned, Booz Allen helps support measured development contracts to do electronic clinical quality measures for meaningful use. Including the feasibility testing portion.

And our approach to the feasibility testing I think is similar to a lot of what is included in the environmental scan so far. And again is similar to what I think Keri and Aldo will talk to as well.

So we've collaborated with both Keri and Aldo and their organizations on our approach and have leveraged a lot of the AMA-PCPI approaches, in terms of doing an assessment around technical feasibility, implementation feasibility, mostly through test sites.

We're trying to get a variety of test sites to participate in filling out a
data-collection tool that helps understand sort of what data elements are currently available that would support the various measures that are being tested, and at what rates those are collected, where their collected, where the data are stored, things like that.

And then also included in the data-collection tools are sort of questions around, not only can my EHR do this, but do my work flow processes support this kind of measure and support collecting the data and at what, how frequently and how embedded are those kind of workflow processes to support collecting the information?

So there's a combination of sort of the more quantitative information: which data elements are present at what rates and where are they. And then a little bit more of the qualitative of, what are my workflow and what would need to change. And with free text boxes please explain. That sort of thing.
And then the analysis of that is a, again sort of a combination of taking a look at the quantitative data, what's there and at what rate. And then how do the various test sites sort of feel about their ability to implement the measures?

I mean that's just a very high level summary. But again, I don't think it's that different from a lot of what's in the current-version environmental scan, so I don't know.

MS. CHRISTENSEN: I will add just some history on why we started doing that the way that we're doing it right now. Which is probably not how we'll do it exactly in the future. It changes a little bit every time we do it.

But what we were finding, our primary way of testing measures for reliability and validity in EHRs is to run a report out of the EHR and then go back in and manually extract and compare the two results. And we were finding that there was some pretty
systematic or maybe systemic problems that we were seeing with where things were not matching, where data was often being found not in a way that the EHR was able to capture it. So we wanted to find that a little bit sooner then when we were doing to extraction. I'm sure many of you are aware that extraction can be very expensive.

So that's kind of where this feasibility assessment methodology came from so that we could kind of know in advance where we were likely to see problems. And if we were going to see a lot of problems we did want to both to do the extraction, we wanted to fix the problems first.

So like Katherine said, we're looking at very specific pieces of information and we just use an Excel file right now. It's nothing fancy, it's just a systematic way to collect information. Is it discrete fields, what code sets are you using, what can you map to if you're not using a code set
individually.

And then that concept of the technical feasibility. Can you capture the field and are you able to do the calculations that are necessary for the measure versus a, we're calling it implementation feasibility which is, do I really think that clinicians are putting the information in that box. Just because you have the box, you all know, doesn't mean that you're using it.

And I might just throw out there, I know at least one person at the end of the row here has actually used that as a site, so maybe he might want to say a couple words too.

But to Aldo.

DR. TINOCO: So to be clear to those on the call, I'm not the one that Keri was pointing to. That gentlemen sits to my left.

But let me try and build upon what's already been said and already documented. Because you are all hearing common themes. At NCQA, we are very fortunate to
have some clinicians, users of the EHR systems in the past. Non-clinician, health care professionals who have been tasked with extracting data from these systems and also people who've worked with health IT modules that are designed to actually calculate these measures and report them out.

So to make this interesting, I think, under our grant work we've used a slightly different approach. We went earlier and we started asking the potential test sites and people who would be willing to respond to a survey, what are your EHR systems capable of doing. Secondly, during configuration and during implementation, what did you enable or disable? Thirdly, can you get the actual data from the data source itself or are there tools or APIs, Application Programming Interfaces, that are required to get access of that data? So how hard is it to get that information out? And we've been able to do that under our granted-funded work for
development of pediatric quality measures.

So we are -- based on experiences learned on other projects, we are looking at feasibility earlier, we trying to tease out some of those nuanced issues that impact data quality as we try and calculate a quality measure.

And we are trying to understand how do we both use attestation-based methods, but also direct queries-based methods from the EHR data based such as count queries or understanding the prevalence that a particular value exists in a data field. So those are just trying to build on previous comments.

MS. FRANKLIN: So can I ask you a quick question, the three of you? The data you're extracting, although first the data we're extracting, is it from inpatient and ambulatory systems? And is it multiple EHR vendors you're looking at?

DR. TINOCO: Thanks for asking. So within our previous experience in the past
year, we have focused on the outpatient setting because these are outpatients mentioned. Should mention that. And the second part of your question?

MS. FRANKLIN: Are you using multiple vendors?

DR. TINOCO: Thank you, appreciate that. So of approximately ten respondents, we had three vendor products represented.

MS. FRANKLIN: Okay. And Shannon, can I put you on the spot and ask you since I believe you're the one that they were pointing to, to talk a little bit about having used the tool that they, the process they were describing. Just your experience.

DR. SIMS: Yes. So the DT is a Excel spreadsheet that asks you to fill out various information about each data element for a given measure. I think I filled it out for three separate CMS projects. So I've done it dozens of times.

And it's an interesting process
because it forces you to delve into your EMR and think about where you collect this information, can it be collected in multiple ways. An example would be, I didn't do this for the DT process but I was thinking about it as Keri was talking on a prior project.

   We were looking at whether an EKG had been done in the emergency room or not. And as an example you would think that would be simple, so you can look at an order and that caught most of them but not all of them.

   Then I pulled data from our ECG system, which is called MUSE, it happens to be separate from our EMR. And that had additional patient in there that had gotten it.

   And then a third way was that the physician could actually not order it and it doesn't end up in MUSE. I have no idea how that doesn't happen, we're investigating that now.

   But documenting in our EMR and flow
sheet that it had been done. So in terms of
feasibility, it seems like a home run. Was an
ECG done: I mean that should be a basic
fundamental component of EHR and it turned out
that there were three, I'm sure there's more
that I'm missing, ways to do it.

So those are the kinds of things
that you have to think about. And then what
I do is you do that and someone like me kind
of geeks it out and thinks through everything
and then documents it.

And then I go talk to my clinicians
and I talk to them, what's the work flow
impact. And that gives us kind of a big-
picture feasibility.

We are an Epic client and I don't
know if this is an endorsement or a
lamentation, but you can make the doctors, I
can torture them to collect any piece of data
that I want, but they won't necessarily do it.
But that's where you kind of come into this
kind of gray area of feasibility.
It's kind of like pornography, it's harder to find but you know if it's feasible or not at this point. I do at least, as soon as I look at a measure.

So anyway, but what that DT tool did was force you to kind of codify those thoughts that are bouncing around in my head into numbers and categorizations. I think you had an ordinal scale from one to five for feasibility and other issues and so I thought it was a pretty effective exercise for me and it really helped solidify and I think was a good way to approach it. So that's what we did.

MS. FRANKLIN: Great, thank you. Does anybody have any questions for the four who presented, clarifications? Paul.

DR. TANG: This may not be a question for what was presented, but I wonder if we're, just to understand the scope of our activity. So this sounded a lot like, how can we assess the past measures?
And it's part of our scope also to say, how could we prospectively describe things to make future measures feasible, even if they may not be in the EHR today, captured in it today? So it's a forward-looking.

Because you know we have the retool, the de novo. So I think this is a lot, the retool sort of approach.

I like Ginny Meadows description because it talked about some of the things that you like to have prospectively as you develop new measures, with the EHR in mind. And again, it doesn't anchor itself into saying, oh, let's only think about what we have today, but what would be needed to get what we really want from a measure.

DR. BURSTIN: I think both are definitely in scope. I think there's very different approaches to both of those. I don't think it's as simple distinction as retooled or de nova.

There maybe de nova measures that
you can test in today's EHRs as well. So I think it's more an issue of current tense versus things we don't think we can yet do in an EHR, but we think are going to be important and what's the approach for those?

DR. TANG: So I guess the catch is then, what we can do in EHRs today? And that's what I don't want to have us, I mean I assuming we don't want to be anchored there. So I wouldn't want to be weighed down with what's "feasible" in an EHR today.

MS. FRANKLIN: Keri.

MS. CHRISTENSEN: And if I can just add, that's a great point that I left out. We asked folks if it wasn't feasible today what would it take to get there?

Is it something as simple as adding that field, whether or not you want to do it, which Shannon alluded to. Or is this a full, we've got to talk to our vendor and do a full upgrade kind of thing?

So we tried to assess that as well.
And obviously it doesn't mean that you might not go there someday.

If it requires a full EHR upgrade, but we did use that as criteria how it would not be an appropriate measure for a program today. If that was going to be necessary.

MS. FRANKLIN: Yes, Martha was going and then I'll get you Saul. Martha.

DR. RADFORD: I'd like some clarification on whether we're talking about feasibility of elements or feasibility of measures? Because I think some of the issues that we're dealing with have to deal with how we design our measures so that we have to focus on a lot of different element issues.

But perhaps we need to rethink how we design measures to make them fit into the clinical work flow better. Because that's one of the big elephants in the room.

MS. FRANKLIN: That's a great point and I think that's something for this group to talk a little bit more about. But I'm going
to go to Saul and we'll put that on the list
of questions to discuss. Saul.

MR. KRAVITZ: In terms of the
approach to data element feasibility, I can't
speak at all to the clinical work load. I
would hope that we would get to a point where
someone would be able to triage and measure
without having to talk to people in the field.

In other words there should be some
catalog of data elements that everyone would
agree are accessible in some setting. It
would have to be qualified by some notion of,
everyone would agree that the patient's birth
date is fair game in a measure, for example.

Whereas when you get to other data
elements you're really going to have to do,
it's going to be a big question mark. Is this
data element developed, we don't know without
further investigation.

So if you stayed within the agreed
to catalog, the kind of it should be possible
to give a, to have high confidence that the
measure might work. Okay, I'm saying might
because you have to qualify it with the work
plus.

So I think if we're always in a
state where we have to go into the field to
collect information, no one's going to do that
early enough in the process to really cutoff
some of these measures that people considered
in the bud early, before we've invested a lot
of money in them.

MS. FRANKLIN: Good point. Paul,
Ginny I'll get you --

DR. TANG: Maybe I sort of
complement that point because I don't know
that we can't be working in the field almost
from the start. It's sort of the agile
methodology.

Because I don't know that you can,
with rare exception, like date of birth,
there's almost nothing that you can't benefit
from how it's done in the field. How it's
input.
The other point I want to ask is, is this EHR feasibility or this HIT feasibility? So the point would be the PROs, the Patient Reported Outcome measures.

Those aren't typically captured by EHRs. I think they should be fair game in terms of what can HIT make possible. And of course that opens up a whole new set of measures and data elements.

DR. BURSTIN: Any eMeasure of any kind, yes.

MS. FRANKLIN: Thanks Paul, Ginny?

MS. MEADOWS: So the discussion about whether we're talking about data elements or measures? I mean that's a good point because one of the things we were actually in discussion last week with some folks from ONC, CMS about was the data element catalog that we received for Stage 2, which was really helpful.

But what we found is that we could look at that data element catalog and say, yes
we have those specific data elements in our
EHR, but without the measure and the measure
logic it leaves a lot of information out. And
when you look at the actual measure and the
intent of the measure, it can very
substantially change what has to be collected
versus just looking at a single data element.

So I think we really have to think
about the entire measure as well. And the
whole process, thinking about an agile
iterative process of working kind of down that
whole path is really I think very valuable.

MS. FRANKLIN: Great point. Aldo.

DR. TINOCO: Thank you. You know
I've heard a lot of discussion about clinical
workflow. I've heard a lot of discussion
about, let's build measures only around the
data that's currently captured in a prevalent
in structured fashion.

And I think to discuss feasibility
we've got to consider the quality measurement
development process in mind. Quality measure
development process as a whole. And I think Ginny was alluding to that as well.

So one question for example. Let's identify, as Keri suggested, a data element that's really important for the measure, but it's just simple not captured.

As a measure developer I'm learning that there are some things that are important enough to merit changes in workflow. And that speaks to the importance question in the NQF criteria.

So if we encounter feasibility issues at a data element or technical level, we ask ourselves, well what's the evidence saying? Is this a flag or indicator that the person responding to the feasibility questionnaire may not be actually capturing information that's allowing us to assess whether or not they're adhering to a best practice.

So hopefully today we'll discuss that. Because again, through many discussions
and meetings such as these, I keep on hearing workflow, workflow, workflow. And measurement is designed to change workflow to be better aligned with a best practice that's out there.

MS. FRANKLIN: I'm going to go to Shannon and then Debbie and then if you could just move, if you want to speak if you could just put your, so Shannon, go ahead.

DR. SIMS: Thanks. I just want to build on Martha's comment that as we think about data element versus measures, I think we also need to think about composite measurement and we need to put that in the parking lot thinking about implications, if most but not all elements are individual components of a composite are viable or not?

MS. FRANKLIN: Debbie.

MS. KRAUSS: So at the end of the day I'd like to see us give recommendations for a three categories, at least, of feasibility testing. One the structured data element that we've talked about.
Two, look into the logic. Look into, if it's a new measure how we're going to represent the logic. But clearly a lot of time for feasibility of that logic needs to be considered.

And also, as I believe Keri mentioned, feasibility related to the terminologies being used. Not that it would hold things up because we could or stop the measure development, but we could know early on that we need to make new value sets or request new terminologies and do our due diligences in research to make sure we're not duplicating value sets and things like that.

So I think It's an important check point to have early on as we're looking at the different types of feasibility that should be required throughout the development of the e-spec.

DR. BREGMAN: Can you give an example of a measure of logic that would not be feasible? Because I cannot, nothing comes
to mind of an example of that.

MS. KRAUSS: I bet Rute could give an example. I'm not, it may not, I'm not the expert on that but I know that we have struggled with the tools that we have representing logic based on availability of the QDM representation, the use of the tools that we have.

So it may be able to be represented, but the complexity produces a 25-page electronic specification for that measure. So you may want to step back and look at the measure and say, should we break this down into different categories, is it really reasonable to ask for this information, is there a simpler way to do it, can we represent it simpler, shorter, easier without knowing that there's going to be so many twist and turns that it may make it very difficult to assure the accuracy.

DR. BREGMAN: Okay. Well I think that's fair. I don't think that that's
something that the EHR vendors are going to
have an issue with.

There is no logic that you could
express in a pseudocode that we couldn't
handle.

MS. KRAUSS: I'm sure.

DR. BREGMAN: So it's, our issue is
not the feasibility of logic. And I do think
the QDM is an issue, but it's not really
something that we will address. But I don't
think that it's something that we will have an
issue with.

MS. KRAUSS: I guess it's looking at
everything that goes into making the logic
represented, with using the tools that we have
today. And I guess the main point that I want
to make is that it's not just feasibility at
the structured data element, it's feasibility
of every section of that measure that we're
representing and is it going to be useable.

DR. BREGMAN: Well I do think that
the logic does effect the validity of the
measure very much. Because you certainly want
the logic to represent the intent of the
measure from a clinical point of view.

You want it to have meaning and I
think there are examples in the current
measures where the logic, in my opinion, does
not reflect the intent of the measure. But I
might point that out, but it's not really
feasibility in terms of the EHR. It's a
different issue.

MS. CHRISTENSEN: Just to provide an
example from some of the testing we've done.
It's not that an EHR could never make it
feasible, it's that many EHRs currently on the
market wouldn't be able to make it feasible.

A lot of times it's with linking.
So something like, maybe I placed an order for
a consult and I have a result from the
consult, but I can't link those two up.

Some vendors can do that, some
vendors can't. Some vendors are working
towards that.
So right now it's a feasibility issue for a lot of people, but it wouldn't have to be a feasibility issue, you're right. Probably most vendors could tackle that eventually. Does that kind of make sense?

DR. BREGMAN: Yes, it does make sense. That is certainly and issue. I think it's more of an issue of the data structure.

I don't think of it as an issue with logic. I'm not trying to, I'm getting a little technical here.

I'm just saying of the feasibility, the parts of feasibility metric. I'm not sure logic is really an equal part to the data element, feasibility.

DR. LIEBERMAN: I'll give one other example for meaningful use. There's a measure on, was a medication reconciliation done after a transition of care and trying to determine a transition of care.

A meaningful transition of care is very, very difficult. I mean we come up with
ways of working around it saying, perhaps at any visit is a transition of care. But I don't think that's really the meaning of the measure.

MS. MARTINS: So I'll start with Aldo and then I'll move to Debbie and Howard and Keri. Because I have things to say about all of those.

I completely agree with Aldo that we need to keep the eye on the prize. We need to continue to measure what's important.

And certainly workflow needs to be taken into account. And that's typically something that measure developers haven't had to think about because there was someone looking for the information, wherever it was in the record.

So I do think that we need to consider workflow. But workflow cannot be considered to be static and something that cannot change, period.

So having said that, I think there
are feasible measures that are useless or that are no longer valid. So we need to retain the balance between feasibility and the validity and reliability and the usefulness of the measure for care improvement.

Moving onto feasibility of the logic. I do agree that there are issues and I wouldn't call it feasibility of the logic. Whatever is represented can probably be coded.

I don't think that's the problem. I think the problem is, is the logic representing the intent of the measure and do we have all the tools as measure developers to represent what we intend to represent.

And an example of that would be the discharge medications. The QDM couldn't handle certain data types so we couldn't represent it given the buckets of information that the QDM gives us.

We couldn't specifically and accurately represent that in the logic. There was some information that would have to go in
So I guess, and maybe I'm wrong, Debbie, but what you're talking about is more of the representation of the measure. Whether that's, I don't think that stands into yours. I think that's, what are the limitations of the QDM, what are the limitations of HQMF? And how does that pose issues in terms of if we're trying to represent something that doesn't exists? That's a problem.

And then I would also add that, if we can make a measure, if we decide that a measure is feasible or in order to decide whether a measure is feasible, the use of standards, such as HQMF, such as QDRA certainly needs to be a part of the discussion. How far into the weeds we go into that, in feasibility assessment, and who should be doing that assessment, is it the measure developers, is it a collaborative work with standards, consensus organizations and
that sort of thing is a question mark.

But if we are able to represent

something in QDM, in HQMF, but then there

isn't a QRDA bucket for it, we have a problem.

So I think that all of these things need to be

considered in terms of feasibility of any

measure.

DR. RADFORD: Listening to this
discussion, if I were an EHR developer, I mean

I would just be asking for a priority list

here. And I think we probably need to address

the fact that some workflow changes, both

human and electronic, which is what I'm going

to put in like the connection between ordering

a consult and actually getting it, it's sort

of an electronic connection.

I think we need to help prioritize

that work. It's all hard work. And the

expectation that it get done is becoming more

of a national kind of hoof beat. So that's

what I would do.

I'd like to also tell you a story,
it's going to take a couple of minutes. I may be the only person in this room that actually uses quality performance measurement to change for your workflow behavior, whatever you want. And I'm going to tell you a story about how that happened at my place. Because it might help us understand some of the issues.

So when I first came to NYU we were performing really terribly on the national measures, which everybody knows. AMI, all those things. Really terrible.

We gave about 55 percent perfect care, which I thought was pretty dismal. We then had a change of leadership, some of us called it a coup d'etat.

But anyway, I went to the new leadership and I said looked, this is terrible. We can't call ourselves any good and say we're good if we are actually performing in on this.

He said, yes you're kind of right.
This is necessary, maybe not sufficient to be truly excellent, but it is necessary. So he said, okay you can have your executive group.

So we got a group together and over the period of the first year we got to 90 percent. Second year 95 and now we're consistently, for the last couple of years, 99 percent.

Now how did we do it? We had a lot of PDFs, some successful, some not. One of the most successful was putting into our EHR, kind of standard order sets that involved following the rules for the national measures.

And that did seem to help and people like it. Well let me put it this way, some people liked it.

What we found after we looked at this, we did a little research project on it, was that although we were at 99 percent perfect care, we were doing all the right things, half of the residents didn't use them.
answered no, because they just didn't want to
be bothered.

But having read the change of
management literature, realizing there's only
two things that change physician behavior.
And that's prompts and audit and feedback.
Maybe money, but that's sort of audit and
feedback I guess.

It really was a prompt and even
though they weren't using the order set, they
were actually doing the right thing in giving
the right stuff. So I didn't care. I want
the patients to get the best care, I don't
care if they fill in all the forms.

So I'm just telling you this story
because to me it exhibits lots of the
challenges and it keeps my eye on the ball of,
that we are trying to change workflow with
these measures. We are. But we have to be,
changing workflow is hard.

DR. BREGMAN: Well to follow up on
Martha's point and also to respond to what
Rute said. And what I want to say is, I missed the opportunity to endorse what Aldo's comment about, that workflow in some cases, you do want to change workflow.

And I certainly, from my perspective, I don't think we have an issue with that. And I certainly don't have an issue with that. I'm speaking for myself and not necessarily for my employer.

But let me give you an example of that. An example would be, we don't currently have, in Epic, a way to store, we don't have a list of patient preferences.

So you can imagine patient has some claustrophobia, if they're going to get an MRI they want it to be an open MRI. And you can imagine if we had a list, just as we have a problem list, we have a medication list, we would have a preference list.

And the preference list would include, prefers open MRI. Or wants to be personally consulted before any blood
transfusion or no blood draws in the left arm.

You could imagine that it would be, and I'm saying this because we're actually actively talking about. You can imagine having this list.

It could be reconciled, it could be prominent, it could be available to the patient in the patient portal where they can edit it. And it would be just as prominent as any other list.

I think everybody in this room would probably endorse a tool like that which we currently don't have. The issue is, from our point of view is, if we were to go down that path, how long would it take us to get there to actually build a tool, test it, make sure it fits into everything else, functions correctly and then could be trained as it is rolled out?

How much time would that take to do it? If we were given a mandate with a six month period to say, you need to start
recording this information and we had to solve it in six months, we would basically jerry-rig a solution based on our other tools, which are kind of generic and not build for that purpose.

And that's what we would come up with because we wouldn't have the ability to build it in six months. But if we were told several years in advance, in Stage X, this will be required and then with some hint of specifications for it, then we can go down the path of building it.

And then in that case the quality community would be pushing us to create that solution. And I think that's great. It's great for the quality of care. But that's what will happen if we are only given a short time.

DR. LI: Agree with Howard said. Regarding the eMeasure of feasibility, I think there are two aspects related to the feasibility.
One is, some challenge caused by the eMeasure. For example, the eMeasure, that element, to define more clearly. The eMeasure logic has some certain level of ambiguity. So basically there is some quality issue about the measure itself.

Such as an improper code or code system was created, that criteria which is sometimes impossible to be captured in the EHR.

So that's the one aspect. So feasibility difficulty was caused by the eMeasure itself.

But I presume these aspects should be captured during the other testing. Such as validity testing and reliability testing, science acceptance testing.

So most of these issues should be detected and addressed during this testing. So really to my understanding, the feasibility testing is really, to answer the question, if we have a measure specification created,
approved by the steward and reviewed by the domain expertise and there's no logic, ambiguity. All the data criteria, data elements are clearly defined.

But is these measure feasible to be implemented by the majority of the EHR systems? So to me the feasibility is forced into the area of the EHR implementation.

So once the key factors related to the eMeasure feasibility from the EHR prospective, it's the essential aspect of the obvious data. These are the data element required by the eMeasures, are routinely collected, generated during the daily clinical walkthrough.

Number one question. Number two question is that these data are captured, but are these data captured in the proper structured format?

For the IOM, even they are captured in the proper structured format. Do they have the corresponding or proper code value set to
bound them or how hard, if they use the local
terminology how hard, how much the effort will
be for the mapping between local terminology
to the standard of terminology defined by
eMeasure?

The concern is that if this data not
captured today in the EHR, when will they be
captured? In six months window or a year in
12 months window or 18 months window.

And how these future captured time
window will be weighed in the overall
feasibility, I mean goal or scale? So I add,
so really my hope is that by the end of the
meeting we can come up with some reliable
quantified approach that if we see this
measure goes through all these criteria, we
come up total score like 80 points.

We see this measure is immediate
feasible to today's EHR implementation. If
the measure is like, the score is between 60
to 80, we see is maybe feasible to be
implemented in the next 12 months. Something
like that.

So by doing that we can have a consistent systemic standard approach to evaluate whether the measure is truly feasible to be implemented or not. It's just some thoughts.

MR. KRAVITZ: I'd like to build on a comment that Rute made during the phone call which is, you know, if we're trying to get to a score for a measure, a single score, I think that's kind of a fool's errand because clearly a measure that's feasible within an integrated delivery network's outpatient EHR may not be feasible within an individual provider sitting out on the prairie someplace who's not connected to anybody else.

Likewise, I think the point that Rute made in the phone call was that, you know, there's the, what is it? The HL7 maturity model for EHR?

MS. MARTINS: It's the HIMSS --

MR. KRAVITZ: The HIMSS.
MS. MARTINS: -- EHR Adoption Model

which already establishes --

MR. KRAVITZ: So I think the example
you gave in the phone call was that, you know,
an inpatient measure for a hospital where the
ER and the inpatient systems don't talk to
each other, that presents a different set of
challenges from a hospital where those two are
integrated.

And unless there's going to be some
kind of mandate from on high that everyone has
to achieve the same level of interconnection
and maturity, we're unlikely to have any
measure which works everywhere, right?

We're going to have some measures
that, you know, that don't require any
information from elsewhere that'll work
everyplace.

But any measure that requires any
kind of integration of information across
settings or across facilities is going to
require different scores.
I think we really need to think about not just a, even if you had a criteria for, you know, as people have seemed to be aspiring towards, you know, if we had a score that we could compute for a measure, I think we're really going to have to compute that score against a matrix of different assumptions in terms of interconnectivity and maturity of the EHR for it to be meaningful.

And then someone's going to have to make a decision. You know, if the government is choosing measures for inclusion in programs, someone's going to have to make a decision about whether a measure that's appropriate in three of the nine setting configurations that we do the scoring in, whether that's appropriate for inclusion in a program, but at least it would be an informed decision.

MS. MARTINS: Yes, and to build on that, I think that the question is what is the average EHR we are measuring feasibility
against? What is that reference?

And I do think that there are
different levels of maturity and I would even
go further in terms of interoperability across
facilities. That's wonderful but we're not
even there in a single facility.

The lab systems may connect to the
EMR. The ER systems are almost certainly
separate systems. Operating room systems, oh
my God, those are silos. You know, all of
those are silos within a single, or may be
silos, within a single organization.

So when we're developing a measure,
let's talk about the paper-based measure.
There are no concerns about that. The medical
record is the medical record.

Someone will go in the different
systems and a person is highly interoperable
because it can interact with all of these
different systems and paper and derive meaning
from all of it.

So the bar is really high in terms
of where the clinical quality measures are on
the paper-based side. They're very
sophisticated because they can be very
sophisticated.

And I do think that what we can do
given the frame of reference of what the
average EHR does today is pretty much did the
patient get this medication during this
hospital stay? And that's, I think, very
consensual.

Anything above that would really
need to go through, you know, are we capturing
it? And it's not just the time frame, again,
because the time frame will be different for
each EHR.

So it's the maturity level of each
EHR and what I was suggesting in the call
actually is do we have measures for different
maturity levels of EHRs and do we have
feasibility criteria for these different
levels?

DR. LIEBERMAN: It seems that we
have a framework in place that would try to address that, which is EHR certification and criteria around that.

So we know as of certain dates EHRs are supposed to be able to do certain things. Now, I don't know, you know, I know what my EHR can do.

But I don't know, you know, how well those certification processes are working in general to give you that kind of baseline that you can then create measures against, and perhaps we can comment on that as well. But let me go to Ginny first.

MS. MEADOWS: Thanks, so I have a couple of comments. One is actually back to the conversation that Howard started about how we could potentially build almost anything into our EHRs, and that's absolutely true.

But I think when we're thinking about feasibility, we really need to look at, as one of the aspects, the cost versus the value of collecting the data because we could
build anything in an EHR.

You know, pretty much the sky's the limit and providers could collect any information that we kind of force them to do. But I think we really have to think about how much is it taking to get there and is there true value in what we're asking them to do?

And I think that's part of the whole measurement of this feasibility of what we're asking for, so that was kind of my first comment.

The second comment goes back to the conversation about maturity levels of both EHRs and provider organizations.

I mean, that's kind of a tough thing to think about because there are certainly different EHRs that have a different level of sophistication and a lot of that depends upon which type of providers they're really targeted to.

So some of these ambulatory EHRs that are really targeted to the small
physician practices, the one-to-three doc
practices, are never going to have the
sophistication of a system like Epic that's
really working or a McKesson that's working
with large academic institutions.

So that's, I think, a little bit
more of a difficult thing to think about
because, you know, the question I guess is
whether we would want to have measures that
would only be targeted to those more mature
organizations? Would that be helpful or would
that not be helpful?

So I think it's something we really
have to think about when we really consider
that whole maturity aspect in both the EHRs
and the provider settings.

DR. TANG: I want to endorse what
Ginny said about the value of the measures.
We talk about value of health care and the
benefit over the cost. I think this is a
really important point.

I do want to respond to Howard's
point about the six months. Now in 2009, six
months was a legitimate complaint in terms of
time to develop.

I don't think in 2012 it's a fair
criticism anymore because we talk about so
we're doing Stage 3, which is 2016, so we've
been talking about it for a year so there's
four years and preference list is on that.

So if a vendor in 2012 is not
listening because we already have a track
record of how much the administration, you
know, pays to the recommendations, I don't
think that's an excuse anymore, which means I
don't think we should hold back for things
that are important, high value, sometimes
involving a burden of cost, if it's the right
thing to do.

So I think the game's changed and
similarly in our measure specification,
measure design, we shouldn't hold back on
yesterday's EHR either. So I think that's a
big -- this is an inflection point and I think
this committee is positioned to make advice.

So CMS has to go let out contracts and other organizations for measures three years from now. We can't wait any more and we can't be anchored by yesterday's EHR.

And there's plenty of advance warning now about what we're, you know, PROs is not, it's already in the vocabulary. It's not a dream, that kind of thing. And similarly preference list is already in the discussion.

So I would say to EHR vendors, you need to be working on those things. But from a quality measurement point of view, we should be assuming that we can have access to it if it's important.

So, I mean, again, I think this is really an important document that's going to come out that CMS will really take into account and hopefully will create a brand new set of measures, if we're providing some good advice anyway.
DR. LIEBERMAN: Along those lines though, it sounds like when you're talking about kind of concurrently coming out with the specifications for meaningful use Stage 3 and the measures, the quality measures associated with them, which is I think going to be a little bit difficult in that you want to implement something before you start measuring it.

I mean, just when we've tried to bring up different modules of EHR, there's always this issue of we want reports on that module.

And we know that there's going to be issues when it comes up and so that if we try to have a report ready at go-live it's very difficult and usually doesn't work.

It's once we have the information coming in for a little while, we can see what it looks like, we can understand changes in work flow, changes in where the data actually shows up. You know, I think that having the
idea that there will be broad categories of
information available is very valuable.

So patient-reported outcomes, I
mean, you know, having measures related to
that seems reasonable if we know that, if
that's part of the criteria for Stage 3.

Yet, you know, having very detailed
measures around new functionality, you know,
you may be better able to measure
functionality required for Stage 2 by the time
Stage 3 comes around.

DR. BREGMAN: So, Paul, you're
saying that we should infer, the industry
should infer, that we need to develop a
discrete preference activity of the kind I
described from looking at the Stage 3
measures? Do you think that it would be, well
--

MR. KRAVITZ: When you said, well
when, Paul, when you are talking about we are
predicting, you're talking about the HIT
Policy Committee?
DR. TANG: But the value of this process that was created, and it's not executed all the time, but the track record's already been that CMS, ONC takes about 90/95 percent of the recommendations of the HIT Policy Committee.

And the importance is not that this one -- it's that the body does all of its work in public. Right now there's a request for comment before the NPRM.

There's tons of opportunity for people and every comment's read and it makes a difference so it's not as if it's some secret deliberation and somebody says a word.

This is just like NQF. It's a totally public process. It shouldn't be a surprise to anybody that as an NQF measure goes through the pipeline, oh, I didn't know until the board approved it. It's gone through a process.

Similarly so has the recommendations, at least for meeting these
objectives and quality measures, so it's out there for a purpose.

And CMS is clearly interested in hearing the feedback and makes adjustments which is why I think this committee and the results of this committee is so important because it, I hope, will help the measure developers both be empowered to use things other than what you had before but also to think about -- so instead of referring to this as feasibility testing, I might look at our objective as feasibility design because we have to, again, work with the vendors and the providers all along the way, developing the concepts, testing it along the way, not just here's the thing to test, go at it. And we'll end up with far better measures I think that way.

What would be nice is if this report said what would you consider but, importantly, and who would you consider it with and at what time, which is at the beginning, not at the
quote "testing end."

But not thinking, oh, there's no preference list in EHR so there goes that. You know, we'll just do the same old same old.

MS. MARTINS: I actually wanted to comment on two things that you said. One of them is the certification criteria and that is a baseline and that certainly is a baseline. It's just not consistent with the eMeasures that are being used with the program.

So I think there's a huge barrier and a huge difference and gap between what the EHR functionality measures or metrics are and what the quality measures associated with programs are.

MALE PARTICIPANT: Can you give an example?

MS. MARTINS: Oh my goodness, I can't think of an example in terms of a data element. What I can think of is the compliance.

So for instance, if we're saying the
1 patient has at least one problem in the
2 problem list, and it's also a feasibility
3 issue but more of a reliability and validity
4 issue.

5 If we're just saying that 20
6 percent, 80 percent of the patients have at
7 least one problem in the problem list, do we
8 have the confidence in the data being in the
9 problem list to use in a measure? So, you
10 know, and that I think is another discussion.

11 But my point is the fact that we
12 have labs and the fact that all of those ONC
13 certification criteria may not necessarily
14 translate into all of the QDM categories and
15 what can be used for developing an eMeasure
16 specification.

17 And we know that certification right
18 now is requiring that all of the data elements
19 used in the eMeasures are collected, but that
20 is it and that's a sentence.

21 MS. JAVELLANA: Hi. Actually I just
22 wanted to echo Paul's comments because at CMS
we are always looking to better improve our process, especially getting measures ready for certification. I mean, there's a lot that's involved in that. It's multi-layered.

And so if this group can come up with how, let's say for example, vendors can get involved earlier on in the game, how we do test decks earlier on in the game, I mean, that's exactly what we're looking for so that would be great.

MS. MEADOWS: So I'd like to speak to both your comments and to Paul's about how we as vendors should be looking forward.

And Paul's absolutely right. I mean, we're following very closely the discussions for Stage 3. It definitely signals to us on some things that we need to be focusing on.

But one of the key areas that we've found takes the most effort is actually in the quality measurement implementation and we really didn't have enough to work with until
we got the measure specs at the end of October
which was a couple of months after the final
rule.

And we're still working with CMS and
the measure developers and they've been
fabulous and really having a lot of great,
open working sessions to work through some of
the issues we found as well.

So I would say that being involved
earlier and having more transparency and
ability to see what those measures are going
to look like is going to be a key to us being
successful and being able to get Stage 3 out
the door in a reasonable amount of time.

DR. TINOCO: Thanks. I'm hearing
this theme of transparency and one of the
reasons why I accepted the opportunity to be
on this panel is to try and lend a little more
transparency into how we approach measure
development.

We haven't really talked about what
we do very, very, very early in the measure so
much in process. We do take guidance from the ONC HIT Policy Committee.

And just like I did when I was working in the vendor space, we built our products based on what we learned from our customers. We went out to the field and you cannot imagine the amount of innovation that's happening at the local setting.

And I'm looking to Shannon, for example. You know, yes, Shannon, I think you said you have Epic but it's not just Epic. It's the thing that he does with Epic.

I know that vendors also go out to their reference sites and find out what are those fantastic things that are happening now and which one of those things should we consider offering in our standard products?

Same thing with our measure development. Our CEO and President Peggy O'Kane, one of our Senior Vice Presidents Phyllis Torda, Sarah Scholle, lots of folks who are involved with making these decisions
are out in the field seeing what kind of innovation's going on within the space of quality improvement and quality measurement.

These are inputs to our process early on to say is this worth improving? And when we ask about feasibility assessments early in the process, we don't really make things up. We learn from what's out there in the field and what these high-performing organizations are doing.

And our vision, in part, is to learn more about them and then to show other folks how they could reproduce that level of innovation and success and improvement through these types of programs.

So, again, I'm hearing themes here. When we talk about what an EHR vendor should or should not be able to provide in their products, we're also listening to what people are doing with those fantastic products already available.

DR. TANG: So it's a perfect set up,
these two comments. So, one, wanted to respond to Ginny's having to wait for the CMS measure.

Our goal, Policy Committee goal, is for that not to happen in the future, that there be a platform so that you don't have to program what is specified.

You know, you already have a platform and you just make available the data elements that are needed to calculate a certain measure so that we don't have a one-to-one tie between some decision and then some hardwiring of something because that's where we are now.

But we're trying to move in a different direction so we have the flexibility to not only do what a CMS measure is but also the things that would improve ourselves as providers, so that's where we're headed and that's in our RFC.

The other thing that's in our RFC is what Aldo had just mentioned which is how do
you take advantage of all the innovations already going on? CMS would like that too. Right now they only primarily can use NQF-endorsed measures.

And by the way, this is an NQF initiative too that Helen had where she tried to gather all the what's going out there anyway?

So in the RFC is a proposal to have an innovation track where if Shannon has something and it works for him, he did it because it works for him, he wants this information, submit that to CMS and that can be in lieu of one of the required measures.

And what happens is then that gives you credit for doing work that you already want to be done and it's a feeder system to, oh, what else could be of use or interest if somebody's finding this helpful?

So one of the pleas is just like NQF's public process. There's an RFC out here. If people don't respond and say, yes,
I'd love to do that, it'll fall on deaf ears.

We just had this in RFI and then only the vocal against won out because all of us, silent majority, didn't speak up and what happens is the silent majority doesn't rule.

So it's really important because these things are going to have a direct influence on Stage 3. So if like the platform is of use to the vendors, if the innovation track's of use, speak up because otherwise it won't come to pass.

But I think these are the kinds of ideas that I think, I mean, it'll be very empowering for measure developers and providers.

MS. MARTINS: So, Paul, I think that's really good because that speaks to what we've seen, is when we talk about these communication representation standards such as QRDA and HQMF which are new to us as measure developers, we're talking about standardizing the communication of information.
And what you're talking about is the step prior to that which is standardizing the representation of information and kind of getting a common ground across the EHR so that we can build from that, and I think that's really important.

DR. TANG: The more you give the justification and even concrete proposals, the more likely it'll make itself through the system, and basically because it'll take advantage of the experience you have, the experience that's already out there that you all have so that the right regs can go out and it'll just make the world a better place.

MS. MARTINS: The other comment that I would like to make, and it goes back to the whole issue that we're discussing here which is feasibility, is feasibility exists, at least in our view, in paper-based measures as well. So feasibility testing is not new in terms of the measure development process.

What I do believe is that the
stakeholders that have been involved typically in feasibility testing traditionally are not enough. So we sure need to involve EHR vendors in this process.

And while there is some sort of formal assessment after a draft of the specs have been put together, this needs to start even earlier for EHR-based measures for sure.

The other thing that I think is very important in terms of how different eMeasures are from traditional paper-based measures is that the cost of putting forward a measure and the impact of putting forward a measure that is not feasible has much greater impact.

It has vendors developing new functionality within their EHRs and it needs a lot more consideration in terms of whether we want to include this data element or not.

And one really good example of how that was managed in the environmental scan is the Yale example where they studied the incremental effect of adding a data element.
But that does bring a lot more burden to the feasibility testing piece which is there's going to be a lot of work, and I think that's important work though, prior to actually rolling out these specs for testing, for pilot testing.

But it does require a lot more thinking and perhaps a more quantitative assessment than we are used to in terms of feasibility or at that stage.

MS. FRANKLIN: I think we've all stated on the webinar we had and now the importance of the early feasibility testing.

And I think we're stating that there are different parts of feasibility testing throughout this cycle that we need to include, so I agree with all that.

But let me also put a plug in for so what do we have now that we can do as part of this feasibility testing?

And one is a request for folks to participate in pilots that are requiring the
electronic specification reporting because we can learn from that and make changes moving forward.

You know, there's the CMS EHR Incentive Program 2012 reporting pilot for hospitals, there's the one on the ambulatory side, there's The Joint Commission pilot and so I think we do look forward and we need to improve our processes.

But right now this is what we have and we can also learn from that, so we've requested vendors for participation in these pilots.

And we had one vendor participate with us and they successfully reported four hospitals, all 15 clinical quality measures for Stage 1.

So that was thrilling on one hand but disappointing that we didn't have more participation even on the test file.

So I just want to restate that in any pilot you can learn and you can learn as
we try to change things moving forward.

And we're also looking for test partners. So we're looking for the vendor community to work with us with current e-specs and with future e-specs and with de novo e-specs so that we can learn and not have any surprises at Stage 3, to receive your input all along the way, and so I think this is critical and welcomed.

DR. RADFORD: I would certainly endorse what I'm hearing as earlier feasibility testing and I would rephrase it a little bit as really earlier collaboration among the vendor developer and provider communities, you know, from the get-go.

And I think one example of this is the concept that the providers are actually doing the R&D on measure development, as you have mentioned, where we develop measures for our internal use and some of them may very well be generalizable to other providers.

It's all about creating win-wins in
a lot of ways. You know, okay, so I've
developed hundreds of measures for my
organization.

Have I ever put one through the NQF
process? No. And, you know, there's nothing
in it for me except headaches. So, you know,
how do you do that? And I think there's
probably ways to do it.

One of the forces at work that we
haven't even mentioned or marshaled yet is
this new ACGME business and maintenance of
certification issue about providers
participating in quality improvement.

What that means is aggregate
reporting on themselves and so if you provide
that, particularly in the form of friendly
decision support in your EHR, bingo, you've
got measures. You've got everything. You've
got improvement too.

So, you know, I think there are some
new strategies we might be able to recommend
that would kind of speed this process along.
DR. LIEBERMAN: I want to make just one comment on something you said earlier, Martha, about I think it was the core measures that you were discussing and about how you implemented workflow changes and tools in your EHR and you improved but not everybody was using the tools.

So we've had, you know, similar experience in meaningful use reporting where we realized we were reporting on the same measures that we were already doing through abstraction, through the core measure process. And there's quite a bit of disparity between the two, so we do very well when we abstract and we don't do as well when we look at the electronic measure.

And in your example where you had made a big improvement in your abstraction rates but probably wouldn't in the electronic rates because half the people weren't clicking the button.

And that's where you get into this
issue of you've made change, you think that your quality has improved, but it's not reflected in the electronic record.

So then you get into this issue of, again, the cost versus benefit. We've gotten a benefit from it, but in order to prove that benefit, there's going to be additional cost to make people use those tools that they've been choosing not to use. I don't know where we're going to go with that but it was a shared experience.

Where are we on our agenda? We could either --

DR. WINKLER: You can be flexible.

DR. LIEBERMAN: Okay. So should we take a short break here and then, because we've kind of been going on with our eMeasure feasibility assessment and whatnot.

But let's take a short break and then come back to the discussion. So about ten minutes, 10:15 or so that we'll get restarted.
(Whereupon, the foregoing matter went off the record at 10:03 a.m. and went back on the record at 10:33 a.m.)

DR. LIEBERMAN: All right, thank you. So I think we're going to have Reva talk a little bit about the guidance and principles now, to kind of steer some more discussion.

And then eventually I think we're going to have to get down to work about defining what the criteria associated with feasibility assessment should be. So I'll turn it over to Reva for now.

DR. WINKLER: In trying to help put some structure around this conversation, and sort of anticipating what we're going to need to put together in our report, we've drafted a set of guidance and principles essentially drawn from your conversation in October.

So that's where a goodly amount of this has come from. And so we tried to figure out a framework to organize the thoughts. And it is, you know, in the materials that you
were sent. And I hope everybody has access to that. Because that's what we're going to be referring to.

It would be a great help if we could go through this and, with the intent of, you know, I sort of envision taking this as being sort of the first draftiest part of the report and building it all out around that.

And so your help in understanding did we capture it right? Did we capture the right things? Are there additional things, you know, structural organization?

So if we could just go through this as a way of organizing the thoughts it would be very helpful for us as we're trying to put together your recommendations, you know, in some sort of report afterwards. And so this is kind of like the nitty gritty sort of work.

But just -- This is the memo we sent you on Tuesday. It's dated December 4th. eMeasure Feasibility Meeting, Additional Materials. It starts out Project Goals. And
Guidance and Principles for eMeasure

Feasibility Assessment.

But just, I realize, perhaps we have one more panel member join us since we did introductions. And perhaps see if you'd like to at least introduce yourself.

DR. BUTT: I'm sorry. I was caught on the Baltimore-Washington Parkway for two hours. It was probably easier to fly out from California than to drive from Baltimore.

But yes, I'm Zahid Butt. I'm CEO of Medisolv. I'm also a practicing physician, and quite involved in various activities pertaining to EHRs and quality measurement. And I do not have any conflicts. I guess that was one of the things that I'm supposed to say publicly.

DR. WINKLER: Thank you very much, and welcome. I'm sorry you had such a difficult morning. Don't envy you. So if everybody's got the document, we are able to
show it.

It was not easy to create slides that weren't anything but totally wordy. So we're trying to maximize it to be helpful. But again, as I said, this was put together based on your conversations. So these are the things, these are a lot of the things that you all talked about.

We tried to put some, turn them into kind of principles, guidance sorts of language. And we need to know if, one, we captured it correctly. Would you alter? Would you expand? Are there things that are missing?

We want to be able to present these to audiences about the general thoughts that this group has had. And then finally the recommendations around feasibility assessment. So that's how we get started.

You know, it's your choice if you want to -- Do you want me to read them? Or do you want to have the group look at each
section perhaps? I don't know which is more useful to the group, in terms of prompting discussion.

If everybody has a chance to look at it, perhaps there, you would want to identify bullets you'd want to discuss further. Or feel that are inappropriate or not captured correctly. That would be very helpful.

DR. BREGMAN: Can I ask if we can start with a understanding of the definition of feasibility?

DR. LIEBERMAN: You can ask. But I think that -- So let me throw out what I kind of see as an output from this meeting. Is that I think we've all talked about scoring the data elements.

But we realize that's not enough to adequately get an idea of the feasibility of a measure. So it seems that we, what we really are looking for is a feasibility report, a feasibility assessment report that can be used as a communication device between
the measure developers, measure sponsors, measure implementers.

So that really the idea being that we're going to expose or make as transparent as possible what the issues are with the measure. Or issues might be a little word, but just to get, expose kind of the, any issues with the measure up front.

So that we don't end up putting a lot of time and effort into a measure. And then once it's developed, implemented and out there, we realize that it actually was not very feasible to get good data, useful data from this measure.

And so I would see it as being a, some sort of potentially structured document that would be a living document that you could start with early on in the development process, that allows you to address kind of a set of criteria around that measure that makes people understand what needs --

You know, is this going to be, how
is this going to be implemented? Or is it, you know, what are the issues around implementation.

DR. BREGMAN: That would certainly be the definition of feasibility assessment. But let me suggest a definition of actual, just of feasibility.

And I think it is, a measure would be 100 percent feasible if all the data that's required for the measure can be captured in a valid way, meaning accurately. With using the existing workflow of the average user with no additional actions by the user other than what they're usually --

And this isn't very elegant to say. But what they're usually doing in their clinical workflow. So if they do what they usually do, they record the data they usually record. And all the data is validly captured in that way, and can then be used to calculate the measure, that would be a completely feasible measure.
DR. TANG: But I would like to speak against that as a goal. Because that's sort of regressing to the lowest common denominator. I'd like to piggy back on what Ginny talked about on value.

And so in a sense, if we're only talking about feasibility, it's just like talking about health care costs. We don't just want it lower. We want it to have the maximized value to health.

So similarly I think we want to look at the value of eMeasures. And we have this, what's its contribution to health, which is a big deal in the sense that we have -- It's not just because you can measure it. It's will the feedback change behavior?

And then look at the burden of achieving that. And it's a balance. So I think that may be something that we want to discuss. Whether that's our goal, versus just a "cost" side, the feasibility side.

DR. BUTT: So I think both those
concepts can be incorporated. Because within feasibility I guess what Howard, what I heard him say is that, what is feasible either today or at any given stage of whatever criteria?

Whether it's through the meaningful use certification, are determined to be feasible according to the data that's captured within that framework.

And then the other is, yes, it's not feasible today. But it should be feasible in the future because it's important. And the importance then comes in that people are asked to incorporate those data elements. And some of them might be easy. Some of them might be very difficult. And some of them might be impossible.

But within that scale of what needs to be done beyond what is already sort of in the pipeline, would be determined based on how important or relevant it is for the quality measurement. So in that sense it is really a sequencing and timing issue, more than whether
it should be one or the other.

DR. TANG: That's what I meant to say. The assessment is to say it's either feasible today, it could conceivably be conceivable in a short period of time or sometime in the future. But still, that's what feasible means.

DR. SIMS: Well I guess from my perspective I think we have to be cautious not to boil the ocean. I mean, I think I'm sensitive that NQF needs to have a deliverable in early 2013. I guess I hear what Paul's saying.

From my perspective I think we can push the ball forward with some technical parameters. Meaning that if we can tell people what data elements, and thus perhaps a roll up into what types of measures. And then rolling up further into composites are most feasible.

I think that allows policy makers to decide what's important or not. I think that
that's an important piece of work that would be nice for us to accomplish. Knowing that it's in the larger context of what brings value to healthcare.

And frankly, I think that's the payer prerogative. I mean, if you know that an element or a measure is going to be infeasible, and you're going to impose a huge burden on to your providers, that's your prerogative if you want to do it from my perspective.

So I'd like to see us design a more, take a more technical approach here. Because I think that's work that needs to be accomplished. From my perspective, I'll tell you my bias.

I mean, if you're going to think about feasibility you need a -- I mean, the quality eMeasurement world is not an infinite universe. It's a fairly -- There are a few hundred endorsed measures.

There's a quality data model. There
are other data models out there that provide
us a list of data elements. That might be
thought about as not comprehensive. But
certainly most of the way there.

And I think if we can find a way to
use something like that as a starting point,
I think that's a rational basis. I think we
also need to make some baseline assumptions
about, you know, I think use of --

Clearly feasibility is a moving
target. This is going to be the first
iteration of feasibility as we progress down
our HIT journey. Things are going to get
better.

But I think we need to have a
rational starting point. And certainly the
certification criteria for Stage 2 seemed like
a rational starting point to me.

So as we think about feasibility, is
that the right place to begin? Maybe it's
Stage 3, or maybe it's something else. But if
we get to far afield of that I think we're
going to lose ourselves, lose the forest for
the trees.

MS. MARTINS: And I would like to
completely resonate what you just said. What
I was going to say is really that, what is the
reference point? Feasibility against what?
To me feasibility is answering,
trying to answer the question within the
feasible, I would say the feasibility of the
process. So how deep can we go into
feasibility?

We could do everything before we
actually develop final eMeasure specs. Is
that reasonable? Probably not. So when do we
stop in terms of quantitative assessments?

Does feasibility involve
abstraction, human abstraction? Probably not.
It's probably more about structured interview
surveys and questionnaires, which is a more
qualitative assessment. But it still needs a
point of reference.

And again, is that the average EHR
today? Is that 80 percent of the EHRs? Is that 50 percent of the EHRs? So as Shannon was saying that the certification criteria may be a starting point.

I would argue though that the certification criteria tend to be general. And I'm out of my league here because I haven't read the rule. So this is from my sense.

But when you're talking about specialty measures it may provide you nothing. So I guess some expansion of what the certification criterias are, or criteria are, or a framework that would work for measures that are outside of meaningful use, should be thought of.

DR. LIEBERMAN: So I'll ask a question much like Howard's. You know, who are we -- Who are the consumers of this assessment? So when we, if we talk about develop -- I mean, it's two questions really. It's like who's going to develop
this assessment? Who's going to be responsible for it? But then also who is it going to be used and how is it going to be used? And I think that, I mean, that will go a long way to helping us determine what should be part of it.

DR. BUTT: Not -- I guess on the consumer side, I mean, clearly in the current context the consumption is both internal, for quality improvement, and for external, as in whether it's pay for performance activities, or the other mandated accreditation related activities.

And in future perhaps public reporting. So my guess is that those are the three or four consumers of this type of quality measurement.

But what I was going to comment on again, in terms of trying to sort of put the feasibility sort of framework discussion on the table. That perhaps one approach is to really use sort of a gated approach in
feasibility.

So in that sense the first gate is really the data availability. So the reliability and validity are actually irrelevant if the data's not even available. So you can't get to the second gate unless you get through the first gate.

So maybe some kind of sequential type of gated approach of feasibility might be useful. Because it sort of goes from a foundational layer to more refined things. All the way to field testing, which is sort of the end goal eventually, to make sure that you get the same results that we expect.

MR. KRAVITZ: I'd like to build on what Zahid was saying with the gated approach. I think in terms of who the customers of feasibility assessment, I think they're -- My perspective on this is all from within the meaningful use program. So I'll frame it within that. There's a whole process where a measure gets selected for e-
specification, and then selected for inclusion in the program.

And at several steps along the way there should be opportunities to kill off that measure as a candidate. And thus the gated approach.

You need to do an initial, you know, when the TEF is considering selecting a measure for inclusion you don't want them to think, to spend too much time thinking about measures that don't have a snowball's chance in hell of ever making it out into successful deployment.

So we need some sort of course filter at the beginning of that process, the first gate. You need some kind of guidance for the developers, while they're developing the measure, that will help them avoid stepping into areas that will cause their measure to become less feasible.

And then there's some final gate which says, yes this measure is sufficiently
feasible across the appropriate range of EHRs that, in this case CMS thinks is important, such that we want to foist it on the community.

So I think there are, you know, gates, like Zahid was saying, where we want to provide feasibility feedback to the sponsor of the measure development process. So they can allocate their resources appropriately.

DR. LIEBERMAN: And, you know, just to kind of build on what Shannon and Paul both said. I think that we don't want to -- I think we need to be clear on kind of what the role of this group is. In that determining --

We all agree that if a measure, if there's enough value in a measure, you know, the absence of data currently should not be enough to say you can't do it. Yet we, I don't think as part of the feasibility assessment are going to be the ones determining the value of the measure.

I mean, I think that has to come
from elsewhere. And I think this group, the feasibility that we're talking about really is, I think more technical feasibility throughout, that covers the life of the measure.

I mean, it can be data elements. It can be, you know, a number of different sources for the data, and that sort of thing. But it's not -- I don't think that we can really address kind of the value issue.

Other than to say, you know, that we need, that having a low, what could be a low score in terms of data availability now should not be a don't pass type of issue. Rute, did you have something to say?

MS. MARTINS: So to build on what Saul said regarding measure selection, and I think we need to think about two different types of measures.

The measures that are already out there that we can do feasibility assessments on the concepts that already exist in the
measures in order to decide whether they
should be re-tooled, for instance. Or do some
sort of high level re-tooling, and then do a
feasibility assessment on that.

And then there's the de novo
measures, for which, you know, maybe we're
just talking about the name of the measure.
And it doesn't really exist in the form of
specifications or measure concepts.

And that work needs to be done
before we can actually assess the feasibility
of the components of the measure. I mean, and
I guess there may be two different
assessments.

One that is a high level discussion
on the concepts that could be part of the
measure. And the other one is when we have
agreed on concepts that are important for the
measure, do a feasibility assessment on that.

And as we are talking about a gated
approach, which makes a lot of sense to me, I
think that we also need to think about how
providers that are not at the very, you know, at the tip of the blade in terms of using eMeasures in their systems.

Should we block them from using simple feasible measures? And the possibility of considering different levels of maturity in EHRs, and are there measures for different levels of maturity?

Maybe they're not the same measures, and I'm being borderline heretic here. But maybe there are versions of the same measures for different EHR maturities that could potentially be compared within that maturity level, but not across maturity levels.

MR. JENTZSCH: Even as a consumer the output of the feasibility, I think it's important that the output is enough for us to look at and say, is this enough information for us to provide public comment before it becomes something we actually have to implement?

So if it's not sufficient to do

Neal R. Gross & Co., Inc.
202-234-4433
that, or it's too vague, that's of no value to us. It has to be very specific so that we can do public comment.

DR. TANG: I'm trying to figure out how to put together some of these comments. One, it sounds like we do have to have, not a yes/no, but a score. But I wonder if we need multiple categories.

I would worry if -- So CMS is not going to be served well if we only use what's in the EHR. There's no way we can get the value based purchasing. Because we've never, ever contemplated value.

And for example, you couldn't do patient reported outcomes. So I wonder if we have to actually categorize two different, some of the future measures.

De novo doesn't mean necessarily new concepts. It's new ways of measuring what we really want it to measure. Not being tied to billing and clinics. So that's, the input is maybe we actually need two different scores
for a measure.

Considering what can we do now, versus what we'd like to do. Because otherwise we will penalize a lot of the forward looking measures we're thinking about.

DR. BREGMAN: Well I appreciate everyone's comments. But I want to go back to my original question. Because I don't think anyone addressed it at all.

And that is, I simply asked whether we can define feasibility. And I proposed a definition. And I would like to come up with a clear definition of feasibility.

Again, the definition I proposed is just that the data's available in the current workflow without any additional input from the user required.

Now the process by which we come up with an assessment, the fact that we may have different categories in the assessment, the fact that we might say not feasible today, but potentially feasible two years from now.
Or potentially feasible if we do A, B and C, you know, all those can be part of the feasibility assessment. But I think just the, you know, I would like to just arrive at a definition of feasibility.

MS. KRAUSS: Howard, I agree with part of what you said, that part of the feasibility definition should include capturing the data. I don't think we need to assign a timeline yet. But we could talk about that in the report.

I disagree within the existing workflow. I think we've had good discussion to show that we can't stick to existing workflow. Because existing workflow, hospital A versus Hospital B, it can be completely different.

And so I think we need to be agile with the workflow and how hospitals implement it. I've helped to implement many systems and we've had to change the workflow. And we've had to twist some arms. But we've always
found that we can come up with a more efficient process.

So I think we all need to consider workflow, but not put it within the existing workflow. But I think the second component that we've talked about is that the ECQM is implementable in the EHR.

And that I think speaks to a number of points that people mentioned. So I'd like to definitely see those two points captured.

DR. BREGMAN: I actually think that workflow is an important part of feasibility. And I think the problem with, you know, when I look at Epic, and when you ask, how does Epic do feasibility testing?

The answer is generally not that we have to survey a large sample of customers and come up with an answer. Because in fact, for the most part, and I speak of course on the average.

But for the most part there are only a few ways that something can be captured.
And maybe it's really just one way for a lot of situations. But in other cases it may be just two. It's not like there are really six ways to capture it, and then we have to evaluate all of the six ways.

And the issue with workflow is, if we have a standard workflow to do something, and then the alternative is to change workflow, that makes it less feasible. Because yes, it can be done. Yes, it is possible. But the cost to the organization.

And we are trying to represent the organization in our responses. To get users to do something differently, to change the appearance of the software, to do that. And to train them. And then to get compliance is very high. And that would factor into our feasibility scoring.

So it's not just a sense of yes, there's potential workflow. And therefore, it's absolutely feasible. Because you could do it that way.
We try to factor in the fact that a change in workflow is costly and difficult. And therefore, it lowers the feasibility score.

MS. MEADOWS: So I agree with Howard and many of the things he's saying. I think the whole topic of workflow is definitely a really interesting topic to think about.

Because as we've looked at some of the things that meaningful use has kind of forced our providers to start doing, that they weren't doing before. A lot of those things are very valuable.

And if we think about the future of where we want to be headed with coordination of care and all the things that we're trying to impress upon to improve the quality of our patient care, those are definitely aspirational aspects of workflow that we really have to get to.

I think one of the challenges that we've seen is trying to drive those workflow
changes by implementing technology solutions
before the actual process of care has really
been defined.

And I think that's where we struggle
as a vendor. Because we don't want to impose
workflow changes on our customers that don't
really have a well defined process of care
behind them.

So that's where I would think we'd
have to think about really -- As we think
about the workflow aspects, concentrating on
how we would influence the development of
those new process before we actually force the
technology on people.

DR. BUTT: So I think workflow is
extremely important in the data capture. But
in this context, the question we have to ask
is, what is the measure looking for? The
measure is looking for data, and not
necessarily the workflow in most cases.

So the question is, does it really
matter if workflow A, B, or C get to the same
data that is well defined in the standard,
according to the code sets?

    Yes, one might be less efficient.
And one might be a best practice. But from
the standpoint of measure consumption of data,
does it really matter?

    My concern is that the most
variability you'll find, even within the same
EHR implementation, same vendor EHR
implementation, is in the workflow
differences. And we could really get bogged
down in trying to define those workflows as
part of an assessment.

    And so I think that's one thing we
should probably discuss, in terms of how
important it is. As important as workflow is
for sure in data capture, in this context
we're looking for certain data elements. How
important is workflow as part of that
assessment?

    DR. LIEBERMAN: I think actually in
the initial sort of criteria we had for HITEP-
there was kind of — One of the criteria was captured in usual clinical workflow.

And I think that that resonates with what Howard is saying as well. It's not -- We know that we can always make changes in the EHR to capture something. I mean, we can always put it into the workflow.

And really, oftentimes the magic of the implementation, or how successful the implementation is, is how do you -- It's a combination of both workflow and back end things. I mean, we can -- In the system, yes we can, there are multiple ways of capturing it.

And you can always ask somebody to click a button. But that's not going to work very well. If you can figure out other ways where in their usual workflow they're already recording information, and get it from there, then that is preferable.

But I guess I hesitate to endorse Howard's definition in that again, it's not a
yes/no, this is what feasibility is. I mean,
I think we can talk about what a desired state
is, or what an optimal state is.

But then we have to measure, or have
to determine how well a new measure fits into
that state, to give us an idea of again,
whether the value of that measure meets the
cost. Or how do those two come together?

Let's see, I think -- I'm not sure which of
you two was first. Saul, go ahead.

MR. KRAVITZ: I kind of like
Howard's definition as a starting point.
Because I think we could all agree that if a
measure met Howard's definition, that it's
highly feasible, right.

So I think the concern is really
from the folks who really want to push quality
measurement to the next level. That if they -
- That Howard's definition handcuffs them,
such that no progress can possibly take place.

But I think if you look at -- If
you take the definition that Howard put
forward, and you qualify it just a little bit, I think you have the start of a really good definition.

So if you look at the, as Paul was saying, we have pretty good foreshadowing of what's going to be required in these programs as we move forward.

So for example, the -- What was the popular one? The patient, the PRO? That's something that's promised or threatened, or however you want to look at it. So you could -- Wasn't that funny?

So if we scored measures against -- If we're talking about scoring measures for Meaningful Use 3. If we said, okay, well we know what's expected -- Let's imagine we could say this. We know what's expected when a system, when a provider deploys an EHR that's certified for Meaningful Use 2.

That establishes some bar against which we could apply Howard's definition, right. So if I have a measure, and the data
elements that it requires are all ones that
are required to be captured in Meaningful Use
2.

And it doesn't violate any of the
other clauses in Howard's definition, we
should all be able to agree that that's highly
feasible in a Meaningful Use stage 2 EHR.

But that doesn't say anything about
what we're really trying to do, which is we're
trying to take up the process of developing
and updating measures for the next round of
changes, right.

We're trying to say, well if we're
investing now. If we're paying NCQA, or the
Joint Commission, or AMA or somebody to
develop a new measure that's to be deployed in
the Meaningful Use 3 time frame, we need to
predict. We need to assess the feasibility
against that time frame.

So I think you really need to -- I
like Howard's definition. But I would move
that it has to at least be divided into two,
which is clearly if it's against today's --

If it passes Howard's original definition, I think we would all agree that the measure's highly feasible.

But you'd also need to have a score which says, I project that against the background of the certification requirements for whenever this measure's supposed to be required, that it would pass the requirement then. So it's kind of a today and tomorrow kind of score.

MS. MARTINS: I don't agree with the definition because I think it's incomplete. And I think it's incomplete because data capture is only one aspect of feasibility.

I do think that workflow needs to be considered in terms of the data capture process. And I actually think there are multiple components to data capture, in terms of feasibility, that we should be looking at.

One of them is data availability.

Where is this data coming from? And
accessibility. Is it structured? Is it not structured?

And then the other one is data standardization. Are vocabularies being used to, standard vocabularies being used to represent the mapping? Are these mappings, are the data conformant with the QDM? So for instance --

And that bridges to eMeasure representation, which I think may also be a feasibility issue. Because if the data is available in the EHR, but we can't represent it in a standard information model, that's a problem.

And then also the data quality viability, which is the likelihood of documentation, how it feeds into workflow. Then the other aspect I think needs to be, or other aspects that need to be considered in feasibility, besides data capture and data representation, are the data extraction and transmission.
So the effort to pull the data out, you know. Do we need APIs or not to get this out of the system? Is there a single system versus multiple systems? That may complicate things there.

And also the transmission. So the transmission standard, such as QRDA, can they handle it or not? And I think that's all part of feasibility.

DR. BREGMAN: Well believe it or not, all of that was implied in the definition, which basically is about validity. You know, can the data be collected, transmitted, assembled, in a way that you're going to get it. And that's implied in saying the data needs to be valid.

PARTICIPANT: If not, what's the gap?

DR. BREGMAN: Yes, sure. And then it gets less -- If there's a gap then it's less feasible.

DR. SIMS: Yes, I mean, I don't know
how you disentangle feasibility and workflow.

I mean, for me at the provider level, when I'm sitting at a provider's level asking them, you know, explaining to them why we're doing this and how to collect the data, it's --

I'm trying to explain to them how to work it into their workflow. I mean, at the end of the day feasibility is a combination of, is it structured data? And is it, how much of a byproduct is it of the doc or the other providers daily workflow?

I mean, that's kind of, it's not that complicated I think. I guess, I think that we -- I understand that we need to improve health care. And that there's a lot we could do to do that. And there's some aspirational things we need to do.

But at the end of the day, I mean, and again, I'm speaking entirely for myself, not for my organization. My survey of the landscape is that many of us right now, for the existing measures, don't use them for
substantive quality improvement, because they

don't accurately reflect the quality that's

being provided.

And that's simply because they're

not capturing the right data at the right
time. So it's probably in a note, or it's

somewhere else, or whatever. So I think we

ignore workflow to our absolute peril.

I was having a conversation with my

colleagues on this committee. And we were
talking about this sort of novel notion of how

we're actually developing a strategy of using

the meaningful use and PQRS metrics to

actually drive quality improvement at our

institutions. And that's new.

So I think to ignore workflow, or to
try to disentangle it from feasibility, I
don't -- I think they're kind of the same

thing, frankly. So I think at the provider

level at least, I think that we'd be penny

wise but pound foolish to let go of that.

I have a straw man proposal for --
I think that Howard defined one end of a spectrum. I would propose that, two things. So two straw men I'll throw out there for the group to chew on.

One, I do think we need to start at the data element level. I think when you implement a measure, either at the vendor or at the provider, if you have one bad data element in a list of 25 data elements, you're in trouble on this measure.

So for example, if you are trying to capture patients with persistent asthma, and you only capture that they have a generic form of asthma, so you can't disentangle it and have intermittent or persistent, for which there are two different treatments, for which there's measures for, you're in trouble. So I think you need to assess at the individual data element.

And so my straw man proposal is that that's our starting place. And then we need to eventually move up to creating some sort of
a squaring methodology, which I'll have a proposal for as well, that goes to the individual metric, and then up to a composite at some point.

Because you're going to get in trouble if you have the annual adult wellness visit that has 12 components to it, and one of the measures has two bad data elements, and the other one has three. But the rest are good. What does that mean for feasibility? Think about that.

I would propose an ordinal scale, meaning from most feasible to least feasible. Perhaps with an addition of future stake considerations as Paul has alluded to. So I'm thinking three, four, maybe five categories of designation for each data element.

I would propose that the quality data model is a rational place to start to look at a list of quality data elements. I realize that that's going to be a big lift for some human beings. And I'm happy to semi
1 volunteer for that.

2 But I think at some point we're just
3 going to have to get down to brass tacks and
4 rate all of these data elements against some
5 sort of feasibility scale.
6
7 DR. TINOCO: Actually, I support
8 much of what Shannon said. I think I'd like
9 to challenge the rest of us, if we -- And
10 myself included.
11
12 If we talk about workflow let's get
13 concrete examples, particularly from those who
14 have lived and experienced workflow. Here's
15 a concrete example. We've got a measure that
16 assesses blood pressure improvement.
17
18 And I used to assess hypertension
19 amongst patients when I used to be in the
20 clinical setting. Our EHR systems right now
21 have a place to capture blood pressure. That
22 place is very structured. And that place has
23 LOINC codes.

When we actually went out and
22 assessed feasibility, sure, the EHR systems
can do it. When we spoke with our subject matter experts, we were in the field.

They said, oh yes, but actually what I do is, the medical assistant puts the blood pressure in there, that structured field.

But I think that patient has glycol hypertension. So I'm going to put my reassessment in my notes. So let's be very clear and careful about what we mean about workflow barriers, burden, using very concrete examples.

Secondly, my first task when I joined NCQA was, could you please come up with a way to assess feasibility for our proposed measures? And I went looking. And I found ways that were provided through NQF's measure evaluation criteria to say, here's feasibility, here's a definition, and here are four sub-criteria.

So as we continue to discuss what those feasibility criteria might be, let's see if we can draw upon what's already out there,
and what may already be implemented in
documents such as the CMS blueprint for
developing some of these eMeasures.

DR. RADFORD: I'm going to pause it.

First of all, I like Howard's definition too.
I think it's workable. And part of the reason
is that usual workflow is a little bit vague.

I actually think that it's not
incompatible with Rute's point of view. And
I also like Shannon's addition about the
different categories. I think there needs to
be some explanation about workflow though,
underneath of it. And it acknowledges the
fact that there's tremendous variability in
provider workflow.

And one EHR implementation is one
EHR implementation. And they all reflect
variations on the workflow themes that are
going to come out inevitably.

I personally believe, this is my
personal opinion, that providers own some
accountability for that variability. I'm fine
with there being variability.

But it's up to us to specify in this emerging world of accountability and transparency where we're going to park all this information in our EHR.

And that's part of an EHR implementation, is to develop that grid of where you're going to park what data elements. And have that accessible to those of us that want to measure stuff so that we can pull it.

And I think that that accountability at the provider level has not really been strongly acknowledged. We really kind of need to push that.

And I believe it needs to be pushed with the professional communities, with the specialty societies. The NQF might be a perfect platform for airing this accountability proposal.

DR. BUTT: So I agree with Shannon in terms of data limit level. I just would like to make sure that we specify that it's
data elements of codified data.

Because I can see a scenario where a measure, an eMeasure would consume non-codified data. Whether that data was capture structured or it was using some technical algorithm that was extracted.

But it still needs to be a codified data element for it to be consumed, I think. Unless someone thinks that there might be consumption of non-codified data within an eMeasure.

DR. LI: I agree Howard and Shannon's comments. At least we can use Howard's definition as a starting point to define what's the feasibility, what's the goal of that.

And also I want add, to echo Rute's comment that, what's the target of the feasibility? So you may come up consensus based framework, hoping to validate the eMeasure's feasibility, then what's the goal? Or what's the target?
If we see this eMeasure is feasible to be implemented we see, you know, a specific EHR, or most of the EHR, the ambulatory EHR, or even hospital EHR. So I think we should also cover the target part in the definition.

MS. MARTINS: So I agree with what Shannon and Aldo and Martha said, in terms of further specifying what workflow issues are. And that workflow really is a really important part of feasibility.

What I would like to say about workflow is that there are bad workflows out there, regardless of measures. And so we need to be very careful in terms of assuming that if this requires a workflow change, is it justified or not?

So I think that's an important aspect of workflow as well, in that there is good, meaningful workflow change. And there's bad workflow change from the perspective of a measure, if we're just pushing something into an EHR as a new workflow.
And it really isn't critical for the measure, for instance. You know, there's certainly discussion to be had around workflow changes and how meaningful they are. And then there are meaningful workflow changes that may be implemented in bad ways in EHRs.

And so I think we need to take into account all of these aspects of workflow when we think about feasibility and workflow. And then to Shannon's point in terms of having the QDM as a starting point for feasibility.

I think NQF already started that process with the QDM, was it a style guide? The QDM style guide, with kind of an assessment of how vendors see these QDM categories in states.

I would like to put a word of caution there though. Because QDM feasibility is not data element feasibility. And a good example of that is actually a data element that we've been struggling with, which is gestational age.
Probably all EHRs that deal or that deal with hospitals and providers that have a significant population of pregnant women and deliveries will have some sort of structured way to represent gestational age.

So under that assumption, when you start looking at the QDM -- And this is where representation I think needs feasibility. It's really hard to assign a QDM category in the state to gestational age.

And there are also vocabulary issues associated with gestational age. So while QDM feasibility could have been okay, when you start looking at specific data elements it all changes.

DR. LIEBERMAN: But if we're trying to develop a framework for eMeasurement, and we're saying that QDM is going to define the data elements, then something's got to give. We can't say that a measure's feasible, or that a data element is feasible if a current QDM data element does not exist.
Then it means that we need a process in place to make changes to the QDM. And at some level somebody has to, there has to be a decision as to whether or not there's enough value in that measure to make a change in the QDM.

MS. MARTINS: Let me just clarify that I don't mean that it can't be represented. In some situations you can represent it using the QDM. It's just that it doesn't provide, the QDM itself doesn't provide you enough detail on how that particular concept is feasible.

DR. LIEBERMAN: Are there additional comments? Or are these new comments from -- Martha, are you, you're done. And JD, are you done? Or did you have a new one. Okay, JD.

DR. LI: I want to comment on the workflow. So today's eMeasure are all retrospective to check the existing EHR data. So the EHR data availability is the key to the eMeasure correlation. Although really,
through workflow this data get captured.

And starting in EHR the sense is relevant. So my opinion is really, we should try to avoid to assess the workflow, various different workflow, the impact to the data capture.

You should leave it to the workflow assessment to the vendor and the clinician. Let them to decide what's the best workflow to capture the data. So the feasibility assessment part should really focus on the data. Not how the data gets captured.

MR. JENTZSCH: I go back to my original comments. And I'm going to kind of second Howard's concept of what we know. The output of this document, this feasibility document that it comes out.

When you do the measure it should have the assumptions of the workflow, or whatever we've all been discussing here. It should have those assumptions. Then the feasibility based on those assumptions.
And then there should be a whole other section that says, but to approve this measure, these are the things that need to be done, workflow changes, whatever the case may be.

I don't think you can just go to a score, or something of that nature, and say, this measure is worthwhile because it got a score of ten. I don't think that works. I think it has to be that robust to cover everything that's been discussed.

MS. KRAUSS: Just one final quick word that I have on workflow. First of all, I'm not saying workflow should not be considered. The only concern I had with the term "existing" workflow. But I think it should be a consideration.

And there should be some points to consider about workflow, if you could implement this ECQM. And not that it should hinder or drive completely the final decision whether or not to re-tool a measure. But it
should be one of those buckets that we consider rather early on.

And I feel important that technology should completely drive the process. But it should be a factor, just like workflow's a factor that we consider whether or not we proceed with it.

So I just wanted to make sure people didn't think I was against workflow. Because I'm one of the most proponent of assessing and evaluating.

DR. SIMS: I didn't mean to paint you as an evil anti-workflow person, Debbie. I knew in your heart of hearts we thought the same thing. So, Joe, I guess I have a question for you.

So in the example I was thinking of, it's just as easy for me to determine how many people are measuring blood pressure, as to determine how many people are controlling their patients with a blood pressure of under 140 over 90.
And clearly one of those metrics of blood pressure control brings a lot more value to the enterprise and to our country's healthcare. But I don't know that -- They're both equally feasible, right. It's the same data. So it's just a different calculation, different logic, basically.

So I guess the question I have is, I mean, are we trying to disentangle -- I mean, I guess what I hear from you about workflow, I mean, the obvious implication for workflow is if we're going to force people to change their workflow, it's got to be valuable to the healthcare system.

But again, I don't think that's our job. I think if we provide valuable information about just the technical clinical feasibility of things, I think that pushes the ball forward and gives decision makers the information they need.

When I think about this information, who's the consumers? I think that, I'm
thinking primarily of CMS and other payers, as well as the measure developers. That they have early information to think about, you know, as measure concepts come up.

How do we -- How viable is this, or not viable? And I think that the people probably who could provide the most information about feasibility is the vendors, probably have the broadest view I would say.

And us as providers who tend to have a narrower provider specific view. But is that a safe assumption? That we imagine the primary consumers of this to be of an output, whether it's an ordinal or binary scale, to be CMS and the measure developers? Is that a safe assumption?

DR. LIEBERMAN: Well, you mentioned two different people who are very different. So it is the, I mean, the measure developers, and then the measure, I mean, CMS or the payer, or whoever is going to decide which measures they use.
I think those are two of the customers. And then it's I guess the other -- but the implementers or the providers are the other ones.

Although I think, are you saying that they actually would not have as much need for this document?

DR. SIMS: Well, I guess what I'm saying is, depending on who we imagine to be looking at this feasibility, whether it's a score or a yes/no designation, I think that affects how, whatever scoring methodology we come up with, right.

I think simpler's always better. But if I'm CMS I may want to know on a scale from one to five how feasible it is. If I'm measuring health maybe I want one to 100. I don't know.

If I'm a provider organization thinking about implementing a random NQF measure, maybe some other designation, just a yes/no is what I want. So I think what I'm
trying to do is serve the greatest immediate
need to try and keep in mind things.

But I'm a, you know, a dark hearted
pragmatist at my core. And I just want to
make sure that we're bringing the most value
as soon as possible.

DR. LIEBERMAN: Yes, I would say
that, you know, that I foresee something where
yes/no, it might be what people want, think
they want. But that's not going to be
helpful. Because too many things would end up
probably being either no or yes. It wouldn't
mean enough.

So it's going to be some sort of
multi axial or multi category type of thing.
And it could be a combination of qualitative
and quantitative data.

I just wanted to make some
observations on this issue of workflow. So I
can think of two examples. Well, one example
has been given around blood pressure control.

And I think somebody mentioned that,
you know, the medical assistant puts in the initial one. And then the physician will re-take it at some point. And actually I think the measure's written for the lowest value during that visit should be the one that counts.

I mean, you know, my own experience I will re-take the blood pressure. And sometimes I'll put it in as structured data, and sometimes I won't. Depending on whether, you know, what my clinic data's like.

And that's going to be very difficult to overcome. I mean, the more that that information is being used, being circulated, whether it's the basis of bonus payments and whatnot, may steer me towards putting it in as structured data.

But it does take more time. And right now it's more just for my own, you know, clinical management of that patient. Just putting in my note is enough.

Likewise with diabetic foot exam. I
mean, I know I look at the feet of my
diabetics. But in the current system that I
use I need to go through another two or three
clicks to make it, put it in as structured
data.

And for me, most of the time that's
not worth it for me. So I don't do it. So
there's an issue. And that's where that
definition of feasible becomes very tricky.

I mean, there is a current work flow
that allows me to put in diabetic foot exam.
But it's not, you know, it's not usual. And
it's more difficult to make it structured data
as opposed to just data available for clinical
use. Minet.

MS. JAVELLANA: Sure. And I'm
already probably going to say things that have
already been talked about. So as far as
workflow goes, again it's more about, you
know, is it really like bad workflow, or
something that's really for quality
improvement?
You know, our goal is, you know, everybody, CMS, providers, is that we use these measures because we want to, not because we have to. And so that will take change in clinical workflow if we all want to get there.

But there were also a couple of other things. So now can we measure whether that workflow is for quality improvement? I don't know. Is there something we could add as a part of feasibility?

And the only way to do that really is to ask the providers. You know, is this something that would be useful to you or not? So those are just things to consider as far as, you know, where --

Because at CMS we do consider that. We do consider administrative burden. Because that's what EHRs are supposed to do, is to lessen that. So, you know, we don't --

We're a little bit conflicted I know in the conversation that we're going on here as well. But also, you know, to cast the
biggest net we could possibly cast, as far as
what --

And I know I'm kind of delving into
the actual, you know, quality measure
discussion. But, you know, as we were saying,
it's like, you know, what do most vendors
have? No, I should say, what do all vendors
have in their product? Not just most.

So something that would cross not
just the most providers, including
specialists. Because a lot of the comments
that we receive is, you know, these measures
aren't applicable to specialists.

So again, you know, if we want to
make it that basic. And also eventually take
it across all healthcare settings. So, you
know, just to kind of think of that as we're,
you know, looking at feasibility.

DR. BUTT: So I think in terms of
the workflow there's sort of two levels of
workflow that one can sort of try to further
define.
So in this blood pressure example,

no one argues that blood pressure can be
captured, and is captured. The issue is
whether it should be captured by a P.A., by a
nurse assistant, or by the front office
person, or the doctor.

The question would be, is it the
purpose of that assessment for us to define
who should do it, or how it should be done?
So it's that level of granularity and workflow
that one can get into.

On the other hand, I think the more
important workflow that perhaps is more
relevant here, is that no one is doing a
certain data element that Paul's committee
decides is very important for National Quality
Measurement for the National Quality Strategy.

And that's where the value question
comes in. That as Ginny was saying, you know,
one can, you know, beat up on the providers.
That they need to capture something which is
not just part of their normal routine of
patient care.

But it's an extra form that pops up that needs to be filled with five clicks. And it's very difficult for them to do as an extra burden, which is not part of -- But it's very important for the National Quality Strategy somehow. I'm just making this up hypothetically.

So the question then becomes, you know, how much do you flog the provider until morale improves, and they sort of, you know, start capturing it?

But, you know, that's where that question comes in. That yes, you know, it's a very difficult thing to accomplish. But maybe the value is not there. So let's not do it.

So I think that's where the feasibility is more relevant from a workflow standpoint, rather than getting into the granularity of that workflow that was just given as an example in the blood pressure.
MS. MEADOWS: So I wasn't going to so much speak to the workflow of discussion we've been having, but more to Shannon's question about who really is the end user of this feasibility evaluation.

And it does go into workflow a little bit. But you know, and this may seem pretty obvious to most people here, but just to kind of quantify what I'm thinking is that there's a couple ways to look at who the users are.

The people that should be completing this evaluation would be both the vendors and the providers. And the providers would be looking at it from the aspect of their workflow.

So it would actually be something that they could provide a lot of value to. So really, it would be a very collaborative process.

And the folks that would benefit from it, I think, would be the measure
developers and the measure stewards who could then, especially if collaboratively they were working with the vendors and the providers in completing this survey, would really understand how they should proceed with whatever they're trying to measure and really think about potentially even doing it in a little bit of a different way, different concepts, different ways of thinking.

So that was really kind of my answer to Shannon's question.

DR. TINOCO: So let me be concise and concrete again. I think Ginny, excellent point. I think quality measure developers are, in fact, consumers of the information derived from feasibility assessments at the data element level.

One point of guidance we can provide for quality measure developers, I think is, is it sufficient to allow providers to attest to feasibility data elements, or do we need to think about other methods such as
interrogating the data base and saying give me a count query, show me the prevalence of this data element is actually used in your organization.

And then a developer with a sponsor says oh, okay, yes, that's really good information. The slippery slope is well, the measure developer may start saying let's go find an organization that we know is doing this today.

You know, this is not research, this isn't randomized controlled trials, I know, but we have to be conscious of what people will do with the guidance that we provide in terms of the methods at which we start building this ordinal scale and how that impacts the quality of results of feasibility assessment.

DR. TANG: Going back to the customer, in principle, we would like to have metrics about a measure that says gosh, I would like to find something that has a high
impact, has a high scientific validity, and a high feasibility.

And without knowing all the things that go into those criteria, I would like to have scores of hundred, hundred, hundred. So it would really concern me if we used Howard's definition because that would calcify the current thought of what's possible today.

One of the reasons workflow is such an issue is because today's EHRs are not very usable. So most of the cost of the workflow is the work arounds of today's EHRs.

We can't calcify the benefit of the value of this measure based on today's work. And so that scares me to come up with something where I would score 100 percent if I just dealt with today.

So I think we have to think more about what would be in more of an ideal EHR, and the workflow of the clinician. The workflow of the clinician, does the clinician capture blood pressure? Yes, they have to.
So that's already in the workflow.

The fact that they have to stand on their heads to enter it into an EHR is not a penalty against meeting the capture of the blood pressure, or that it is not in the workflow of the clinician to capture blood pressure.

So we cannot penalize a measure, a good measure that produces good results by today's EHRs. That's sort of the concern I have with, you know, anchoring it there.

DR. SIMS: Paul, I mean, I agree with you. We're ideologically aligned. But where does pragmatism begin because, I mean, we think we know maybe what's coming for Stage 3 and that's really going to drive, really functioning a lot of what is available in Stage 3.

But how do you create a system that estimates the aspiration of what an EHR might be in the future right now? I mean, how do we do that?

DR. TANG: So Stage 3 is four years
away. I think the HR vendors, within the next four years, can find an easier way of capturing smoking cessation than five clicks away, as one example.

I mean, it's just one of those do we need to do it, do we want to do it? Yes. Is it hard to do in today's EHRs? Yes. In four years time, it shouldn't be.

That's what we should go towards instead of well, let's throw all that out because it's really too hard to do today.

DR. LIEBERMAN: Paul, what is your alternate definition?

DR. TANG: I think feasibility has nothing to do with anchoring it today. Everything that you said was what can I do today with no work and using today's EHRs.

The alternative is to take a lot of what Rute was saying. There's a lot of aspects to feasibility. We want to have something where you could put a high score, and that would say yes, it fits.
Blood pressure's a good one. It's a really important parameter. Everybody wants to do it. It's in the clinical workflow. It's just got to be a lot easier to put the doctor captured blood pressure than it is now in today's EHRs.

So if we put into these various aspects of "feasibility," I think that the word is maybe a problem. We will have something that a high score should be something, both the measure developer aspires to and a payer would say hey, if there's these three high scores, that seems like something that I should put out there.

And we should have that to be, it's balanced. That's the whole thing about the lead, and there's a lead time to what we're doing right here.

But CMS is working on the lead time. They have to work today for Stage 3, for a few years from now. But we cannot anchor it in the past. We will have lost a whole lot, I
think, if we anchor it in the past.

DR. BREGMAN: Well, how will I respond if you were to ask me to give a feasibility assessment, which is again, I think it's accurate to say it's either the vendors and the providers that are going to give this assessment, how would I respond to that request? On what basis would I respond? How would I generate a score on your paradigm?

DR. TANG: We're going to come up with a way. And I think that's our afternoon exercise. There'll be different weights to what, see CMS is looking for not today.

It's what would a measure be like if it was highly feasible in 2016, as an example. We can specify that, I think. If I were to say how easy is it, because that's the only dimension I see in yours to put it in your EHR? Not easy at all.

But that's not a property of the measure, that's a property of the EHR system, I think, largely. I mean, it's all many
things.

But we want to get to what's a good measure. That's what NQF is concerned about, and that's what CMS, for example, is concerned about.

DR. BURSTIN: Just to jump in for one sec, just a point of clarification and CMS could help here.

I think it's important to consider the fact that there are some of these measures in the here and now that as of 2014, if I'm correct, people could use their meaningful use measures, for example, to apply them to other accountability programs like value based purchasing, the physician modifier.

So I think it's really, I agree with you, Paul, we need to be future tense. But I think this has to be a stratified discussion. I think we need to have one that's feasible in the short term, and one that's more aspirational.

But I think if we blend them, we
loose the ability to say which of these measures in the short term are appropriate for accountability, because they're valid representations of quality.

And if we have systematic missing data, it's not a valid representation of quality. So I think we need to just have it as a stratified stream of both accountability.

And I would love to hear CMS' perspective on this because I think it's a really important real life consideration that's beyond meaningful use three.

DR. TANG: I know, but you were out of the room when we talked about having two categories.

DR. BURSTIN: Sorry, okay, okay.

DR. TANG: Just like you said, yes.

FEMALE PARTICIPANT: And Helen, I think that's a great consideration, and it's a great point that we have considered. And when the hospital side, and I read some of the other vendor methodologies when we did
feasibility testing for the structured data elements, we specifically asked the vendors that we had. We had a range of eight or nine vendors representing multiple systems.

And we looked at the data element feasibility capture to be captured today, and within the next 18 months and across. So certainly as we use them for currently retold measures, we want to be aware of what's going to be going on in the programs we're going to be implementing in the short term.

But we also want to consider, and everybody should consider in the feasibility testing, the future, the 18 months, the two years and how systems can adapt to capture that data.

So definitely it's a spectrum. That's why it's going to be difficult. We just need to have these various buckets of what we're going to include in feasibility in some ranges, in some multiple considerations.

DR. OVERHAGE: Dave, you already
talked about this morning, and I apologize for joining late, but to me, it seems like we're talking about measures and feasibility and workflows and all this stuff. But at some level, it comes down to what does it cost to capture a data element?

And it costs something different to capture different data elements in reality. And I think we, at some level, have some commonality of there's certain data almost that we think should be a low cost for an EHR or a provider to capture.

So a laboratory result, we would hope would be a relatively low cost thing for them to capture. Something else, like a psychosocial assessment of the patient's readiness to change might be a very expensive item to capture.

And so I think about things, the world more in terms of what are the data elements than I do about measures or workflows or whatever. So maybe this didn't make sense
to anybody else.

And I think about data elements in
at least three buckets. You know, there's one
bucket which are things that are sort of no
brainers, if you will.

Basic things, like I said, lab
results, maybe a diagnosis code at an ICD-9
level or ICD-10 level or whatever the current.
But just because that's something that you got
to do to survive.

There's another level of things, and
I agree with Paul, you know, that we don't
make it easy for providers to capture things.
But hey, that's a reality today.

So there's a real cost, at whatever
level it is, to get things out of the
provider's brain into a structured record cost
allot.

I mean, the example I always use is
they send you these stupid surveys in the
mail, and there's a $5 or $10 bill folded up
with a survey.
Well, they don't do that because they're nice. They do that because they know it takes your time to get some out of your brain into one of the five checkboxes for those ten questions.

And it's hard. You know, we don't have any magic to do that. I think EHRs could be much, much better. But even if they're perfect, it's still a very hard job.

So there's things that we have, as providers, captured, if you will, in our brains at some level. But translating them to structured data costs a lot.

So automated data, you know, pretty easy, should be. Even blood pressures, you could argue, should be pretty easy. And then there's the stuff that's really, really hard, you know, either because people aren't capturing it today, it's not available in the structure.

So it's maybe a different way of thinking about that, because at the end of the
day, we're talking about what the measure
should be.

It seems to me this is a diagnostic
test on our health care system. And how do
you look at a diagnostic test? It's a
screening test, right? We're saying how are
our providers doing with X, Y, Z?

So when you ask that question with
the screening diagnostic test, how do you
think about it? Well, you say how sensitive
and specific is it? And how much does it cost
to do?

So at the end of the day, a measure
that costs a lot to do, because it uses data
elements that are difficult to capture, and
I'll talk about today and tomorrow in a
minute, is one we may want to think real hard
about what's the benefit that we're gaining
from that diagnostic test.

And just the silly example, but you
know, the beta blockers after MI stuff where,
you know, how valuable is that going forward
given the level of performance? Do we want to keep testing every year for something that we know is stable and good.

If you're an internist, you know, like me, you do that. But in general we maybe don't want to do that because it costs something to do.

So the other related thought is to Paul's point. In the vein of trying to help people, you know, if I choose a stupid EHR, it costs me a lot. This is a huge point, you know?

If I can capture that data from a device or from unstructured data or something, I can be really clever and smart, or I can create a workflow where MA captures data and I validate it.

Good for me, it lowered my cost of doing it. And the analogy that came to mind for me was MPG, Miles Per Gallon goals. You know, maybe you have buckets for these different levels of goals.
There's stuff that should be free, right? Structured lab tests should be free, and the measure developer shouldn't have to worry about how much that costs.

You know, there's a next tier up of things that are, you know, they're $3 today to capture. And I'm not actually making up that number.

That's the number that Kaiser uses for what it costs them to capture a structured data element that they didn't previously, per data element.

So it costs, well for 2014, maybe what we're saying is, you know, that's a $2 data element, or $1.98 data element, or $1.10. I mean, it sounds silly, but it's sort of like so first, let the provider, you know, once you get sort of a today's assessment, to Howard's point about what is it today and is it a good and valuable thing to do?

But let's raise the bar by lowering
the price going forward. And then there's
stuff that just, do you really want to spend
$100 to capture the stage of labor, I mean,
I'm making it up.

But you know, some of these things
are really, really difficult to capture in a
clinical world. Is it worth doing the
measure? You know what I mean? You have to
ask that question.

So I hate to reduce it to dollars
and cents, but in some ways, you know, it's
like so many things, you need a common
currency.

What's the value of what we're doing
and what does it cost to do? And it makes it
a real easy decision. I'm oversimplifying,
obviously. Is this a worthwhile measure,
worthwhile diagnostic test to do given its
cost?

And can we simplify this assessment
of workflows and things by saying this kind of
data ought to be relatively cheap and easy.
This kind of data, you know, we could put tags and red, yellow, green in the QDM or whatever. You know, this kind of data more expensive than this kind of data. If you think about it, you better be ready for people to say wait a minute, that's going to be really hard and costly. And you better have a higher tier, or a higher standard for whether that's a good thing to do or not.

DR. LIEBERMAN: Ginny?

MS. MEADOWS: So I completely agree with those comments. That was actually great descriptions of what I was saying earlier about looking at the costs versus the value. But one thing, if we could kind of go back to Paul's comments about some of the difficulty, I think we have to be careful about mingling the usability of a system with the feasibility of collecting data because as we do know, I mean, EHRs have traditionally been used for very different tasks and reasons than we're now moving into, kind of this new
world of what we really want to be able to do
with our Electronic Health Records.

But the whole usability piece is a
whole separate topic that could take us a long
time to discuss. But we probably need to
separate that out a little bit because I don't
think it really impacts feasibility as we're
talking about it today.

DR. BUTT: So I think, based on all
the discussion, it looks like the goal for us
should be to have whatever assessment
methodology we sort of come up with later on,
that it be able to determine not just whether
it is feasible or not, but when is it
feasible.

In other words, is it feasible in
the short term, or is it feasible in the post
2014 world. And whatever methodology we come
up with, whether it's a single score, it needs
to be able to be able to discriminate between
those two states in a sense because, to me,
that's really the critical part of this.
So that's sort of something to keep in mind when we have the subsequent discussion later on.

MS. MARTINS: I echo Ginny's and Marc's comments in the sense that it's important for us to measure feasibility beyond now. I think it's more than just the timing issue.

It's not just timing. It's EHR capability because we're going to have EHRs at different stages at any given time. So it has to be anchored in what the EHRs can do in a set of criteria for what EHRs can do rather than, or not rather than, but in addition to how long a particular EHR may need to get to that state.

And then to further echo Ginny's comments on the blurry line between what are issues with EHR adoption versus what are issues with eMeasures in terms of workflow.

It's not only that there may be bad workflows out there in EHRs, it's that we're
asking providers to start using EHRs and to start collecting data for measures, based on EHR data.

So that's a huge barrier, I think, for providers. And it only started when incentives started being provided. So I think that the barrier to EHR adoption was workflow change in a sense.

And we're just asking for more workflow change. And I think from a provider perspective, the two may be confused. And there's a lot of change management that needs to be done around this, regardless of the measures.

DR. TANG: One comment about the usability, I don't think we ought to put that into the criteria. But I think it's a strong influencer over the workflow costs. That's how it's sneaking in here.

So let me go with what Marc was talking about, the dollar. It's a $3 effort now, it's a $1 in the future. And it maybe
still goes back to maybe there's actually a
stage of quality measure.

There's a 2012, there's a 2014 and 2016. And each of those had different "feasibility," I wish there was a better word, score because, ideally, a measure developer shouldn't have to assess the measure being posted, shouldn't have to assess, know the details of scientific rigor or how to even spell QDM.

They would just like to have a score that says somebody who does know all that stuff assigns it this. So in some sense, if we knew what it costs today, if it was $10 for 2012, then let's not make it at Stage 2, 2014 measure.

If it could be $1 and it's really important, in 2016 let's do that. I mean, that's the kind of information I would think the measure developer, particularly a consumer like CMS would want to know.

So we need to find the different
aspects that go into that dollar or $3 or $10 to really give a good rendition of what it means to CMS.

The way I would characterize Howard's definition is that's the easiness factor, and that's a factor. But it's not the whole game. I'm worried about us making it the whole game because I wouldn't want 100 percent easy. Well, you know, the free lunch thing.

DR. LIEBERMAN: Just one comment. You know, in trying to determine what the cost is in the future, or should be available in the future, I think it's valuable and we need to include it. There's always uncertainty, though.

And you have to factor that in, as well. So we can pretty easily say something's available now. We can even look a year ahead and say well, that should be available. When you start looking two, three, four years out, it's very difficult to, I
mean, we have some guidelines now with meaningful use criteria about what should be available, so that can help us.

But there's still going to me more less certainty about what the costs will be of a data element in the future.

DR. TANG: Let me give you an example. In the paper world, a patient reported out, come cost $100 per phone call. In today's world, it's probably still $10 or whatever it is, $10, $20.

We're expecting, we're hoping in 2016 that that will be down to 20 cents. So you can probably, if you know enough, figure out about that.

So yes, it's uncertain. But CMS and NCQ and PCPI should get some indication of what's reasonable in a sense in 2016.

DR. LIEBERMAN: And just the other point about assigning a cost to a data element. I like it a lot, and I think what would be, I don't know who would do that work,
but it sounds like Kaiser's already done some

of that.

But it would also provide some

feedback to individual institutions or clinics

or vendors to see that there's a data element

that is inexpensive. But in your system, it's

expensive.

And you have to ask the question,

why is that? And that's something to look at.

Howard?

DR. BREGMAN: I just want to respond
to Paul's comments. I do think, I'm going to
take credit that I think Marc kind of restated
my definition in his explanation.

Feasibility, in my mind, has nothing
to say about value. So that a measure could

be, if it has significant effect on quality of

health and it's deemed to be important, and it

scores a 20 in feasibility, I don't think the

EHR vendors are telling you not to do it.

For first of all, we shouldn't be
giving it a score of 20. I'm saying 100 is
perfect and zero is no feasibility. We shouldn't be giving you a score without an explanation of why we gave it that score.

We should be very specific, and we are willing to be very specific about why we would give it a certain score. But if it's determined by whatever stakeholders are involved, that that is important enough, then feel free to do it.

You know, go right ahead with implementing that measure. I would say that if you're going to pick ten measures, and they all score low, that's probably not a great decision because there's difficulty all around.

Now if you have ten measures and nine of them score very high and one of them scores very low, but you think that one is important, you know, this is all getting back to Ginny's original statement.

It's just a cost benefit decision.

The feasibility and, I think Marc restated
this, is a measure of cost. And if something
is low feasibility, it's just a statement by
us, an assessment of the cost.

And believe me, I do not think that
we have any interest in understating the
feasibility. So I think we really have an
interest.

And because we're not just in it to
get off easy, we're in it because we also have
an investment in increasing the quality of
care and the quality of outcomes that we
really do want to give you an accurate
assessment of the feasibility.

And if it's in the context, if it's
the feasibility today, then that is different
from what it could be, you know, in a couple
year interval.

DR. BUTT: Yes, just to clarify the
issue of the sort of milestones of the 2014
and beyond. I think just to clarify, I think
if the assessment tool or whatever we want to
call is designed with the necessary weighting
and so forth, because really what we're
talking about is data availability along those
milestones because there is a certain national
program that is pushing that availability
along those milestones.

So instead of focusing on the
milestone dates rather than focus on what is
it that would determine the score of that
availability that would automatically conform
to those milestones would be the way to look
at it, as opposed to try to sort of have
assessment that's geared towards the dates
themselves, but more towards what is
underlying those dates that would drive that
score is kind of what we need to keep in mind
when we develop the scoring.

DR. LIEBERMAN: I would assume that,
you know, the EHR certification process is
going to be ongoing, and will change over
times so there will always be a set of
criteria to move towards, that you should have
some expectation of what should be available
within what period of time. Am I correct in
that? More knowledgeable people in the room?

DR. OVERHAGE: I guess I would not
expect the certification process to be that
granular because it can't be that responsive.
I think the kind of capabilities that we'll
see, I agree, they will.

But I think for the purposes of our
discussion here, I don't think we're going to
see functionality specification at a level of
say, you must be able to record at a
structured level, the gestational age of -- I
mean, you know, that's sort of too low a
level. And that's where our challenges are,
I think.

MR. KRAVITZ: That's in there now.

Essentially, you have to be able to capture
all of the data elements that are in the
meaningful use two measures. That is a
certification requirement, irrespective of the
cost.

DR. OVERHAGE: That is sort of true.
I mean, it's back to Zahid's point that it
doesn't say you have to be able to enter them,
or you know, it doesn't specify how you have
to be able to capture them.

And it doesn't say that that's how
you have to use them to generate the measures.
It just says, so to Paul's point, the answer
is yes, you can stand on your left foot and
push with your pinky and you can make it
happen. That's certifiable, that's not feasible for our purposes.

MR. KRAVITZ: Okay, so that you
could be certified, but it doesn't mean that
there's any reasonable way that, so using
gestational age as the example, you could
certify that you could capture and compute off
of gestational age, but doesn't mean that
there's any meaningful workflow where that
data element would be entered into the --

DR. OVERHAGE: Well, certainly that.
Or it may be difficult, or you know, and
again, it's a disconnect. Right? They're not
connected in the sense that the certification
criteria are, like you say, abstract in the
sense of saying capture these data, only we're
not saying what they are.

And then somebody else is saying,
well here's this other data element. So data
elements are not data elements are not data
elements. It's very different.

And so there's this sort of
disconnect of saying well, you have to have a
way to do it. Well great, I can pop up a
form, you know, the answer.

So then I'm certified. Now you got
to figure out a way to do it realistically.
So I mean, I don't see those as connected, or
at least, not very well.

MR. JENTZSCH: Unless I'm missing
something, we're talking about eMeasures in
general, not just meaningful use. We keep
talking about meaningful use as certification.

I don't think certification has
anything to do with all the other 13 measures
that are not part of the MU, right? It's all
eMeasures. Am I wrong?

So the feasibility score is going to come in, it may not even be part of the
certification process. So I think that's kind of an irrelevant discussion.

I think it's important that it be certified, but I just don't think we should be focusing on just that. It's not really the vendor's responsibility to, the roll that I take, we do something called the Big Q.

It's a very large, essentially, database that shows all our measures from TGAC to NCQA, to a bunch of measures. And we go through a feasibility process all the time.

They may be measures that are specified by other people. We bring them in, we try to figure out what the feasibility is before we actually report it at a national level for that.

And that process is very difficult.

We may get a very low score on feasibility,
but it may be very important. It will show up on our dashboard.

And then we track it over time how well we are improving in the implement at the regional levels, different workflows, whatever they have to to improve their scores to do that.

DR. LIEBERMAN: I think that I wouldn't say that this is irrelevant in that what we've talked about is giving a baseline to assess feasibility against.

And that's where we do have national standards and national specifications about what should be expected, what an EMR or an EHR should be able to do.

DR. WINKLER: I actually wanted to ask a question based off of some of the comments here. We were talking about the potential customers for a feasibility assessment tool.

Someone said that the developers would benefit from the assessment. And I
wanted to ask the developers, the roll you see in this sort of an assessment tool in your development process very early on, because there's been a lot of talk about feasibility considerations have to be moved way up to the front of the whole process, which really starts with the developer creating some idea about a measure, and then you know, figuring out how you're going to do it.

So I would ask the developers, do you see this as a tool that you would, you know, would be using in your initial sort of formulative thinking about how this measure could be created, developed, constructed.

And then perhaps use it in an iterate fashion as you move through the next stages of development, really trying to figure out your specifications and really figure out how that measure's going to come together and ultimately, as you move into developing the logic phase.

So I just was wondering really how
the developers were seeing the potential use of this kind of a tool in their development processes.

DR. LIEBERMAN: We'll take more comments while you're thinking about it.

DR. BUTT: So I think that the certification standards are really very helpful at one level. So for instance, when the certification says that there must be a problem, there's that structured and in SNOMED, or ICD-9, that's very helpful because it's establish a certain floor, which is expected of all these EHRs.

Now what it doesn't do is then it has not the granularity that Marc was talking about, that okay so you have a diagnosis in a problem list, which few will smile here including Howard and Rute and Chris.

But the measure needs a principle diagnosis, the measure needs a discharge diagnosis, the measure needs all those attributes of a diagnosis, which generally are
not part of the certification requirements.

So that's kind of where the rub is
that those are very important for the measure,
but they're not.

So I think it's very good to know
that in every EHR, there will be a diagnosis,
and that it will be specified either in SNOMED
or ICD-9. But I think the eMeasures generally
need more than that.

DR. OVERHAGE: You said something,
and I think it's worth, you said that the
measure requires. And I think that's one of
the places where the rubber hits the road
here.

The measure doesn't require,
frankly, a principal diagnosis. That's a
mental model that has existed and we've used.
What does it mean to require?

You know, are there eight other ways
you could get at that? Probably. That's just
the way that we have done it in the past
because that fit our information model that we
had a primary diagnosis on a hospital discharge.

That's what we thought about.

That's not the only way to get there. And so that's where I think some of the tension comes in.

And as you said, moving this process up front so that you're not saying we assume that there's -- well, you don't have to have one. That's just something that we did because it was convenient.

DR. BUTT: Could I respond to that? I don't want to open a whole new discussion. But I think that some of it is, I guess, maybe required may be too strong a word, but some of it is quite important and necessary in some ways.

And we can have a whole discussion around that. But your point is well taken that we should look at everything that has been done up until now, or has been required to see whether the same intent can be achieved
by more sort of EHR based data capture methodologies.

But having been in the weeds of this stuff, I can tell you that some of it is necessary. And not all of it is something that can be gotten rid of.

MS. MARTINS: So I think that the measure, and to Reva's point, I think that the measure developers really need this feasibility assessment because some feasibility issues may be show stoppers for a measure, depending on how important the concepts are for that measure, how rare the event is for that measure.

How is it going to effect the rates in the end and make the measure meaningless? Well, maybe we need to reconsider this. So I think that's really important.

The other aspect of it is that traditionally, measure developers are thinking about what's in the record, period, and not at the cost of a particular data element that is
not captured in a structured way.

You know what? The information may be there in a free text field. It's not e-feasible. It doesn't mean that the measure isn't feasible.

So I think there is a gradient in terms of automated data capture from an EHR, and a measure that is fully implementable that way.

And then there is a decision that measure developers face which is, is it really cheaper to collect this measure as an abstracted measure?

Maybe it is, maybe we should stick with it for some of these measures. Maybe the value of collecting it electronically does not justify the workflow changes. It may not even be meaningful.

So if you're looking for a specific care plan with specific information for the patient, how can we get at that electronically right now?
So I think there is value for the measure developers to decide. You know, and if this really important, maybe it should stay as a paper based measure for now. Or maybe there is something that we haven't talked about really, which is a sort of a hybrid.

What is it that can be collected in an automated fashion, and then complimented by abstracted information in order to retain the validity of the measure and the buy-in from providers that are looking to their measurements and going ah, why are my rates so low?

Well, I did deviate from the guideline from this measure, but it was justified. And that's one of the hardest parts of capturing in a structured field, in a reasonable workflow, within an EHR.

How can we still account for that information where the providers feel that they need to?

MR. KRAVITZ: I think the cost
model, or talking about things in terms of
cost and value, I think it's really, really
helpful because I think, you know, if you look
at, and I want to come back to something
Shannon said a long time ago which is, you
know, one bad data element can kill the
feasibility of a measure.

When you're trying to translate the
science of the clinical quality of the
measurement into something that you can
actually do without a chart abstracter, you've
got to make some trade-offs.

You can either have a very expensive
measure that no one can implement that's very,
very accurate. Or you might have to back off
and say I need to drop that $15 data element.
And okay, my population shrinks by 50 percent,
but hey, I can actually get everybody to use
this.

So I think this notion of trading
off the cost of the data elements and the size
of the population that you can address, or the
clinical value that you can bring, I think that's a really good way to think about it.

DR. LI: To response to your question, the tool, the feasibility assessment tool, I think if there's a tool, it will be very useful, very helpful for the measure developer to determine what the measure candidate to be developed.

So personally, I really like such a tool available. So as of today, all of the eMeasures are developed based upon the QDM. So really, the tool should convert the conceptual QDM into a implementable representation.

Then, it's more like a NQF conduct semi-annual survey against all the QDM based implementable artifacts to majority of the EHR vendor. Is this element supported in today's, your EHR system?

What is the data format, structured or narrative? What is the vocabulary you are using? If not, what's the cost estimation to
support it in the next, you know, 18 months?

If we have such a comprehensive
survey in front of every measure developer,
that this survey results get updated every six
months, then we will have a much reliable
evidence to assess, to determine, you know,
should we retool this eMeasure, this paper
measure into eMeasure given the current state
of that availability?

DR. TINOCO: So thanks for the
additional time to think about the question.
I don't want to bandy with semantics. I think
the term tool can get people, I agree and I
know.

But what we need are standards of a
way of communicating the results of our
assessments. I believe that measure
developers, there are many of us out there,
and growing.

And you know, providers themselves
are developing measures, right? So with that
many players in this space, I think we do need
to think about the criteria, the uniform criteria that we should be communicating.

And how to communicate it such that the downstream consumers know what they're getting, and they can compare things across people that are actually building these things.

DR. SIMS: So one of the things I've been thinking about as we've been talking is I love the idea of costs. I have no idea how we would practically assess that against, you know, 10,000 different data elements.

But I'm wondering, something for you to chew on, I looked up a thesaurus. Feasibility, to me, when you say something's feasible, almost anything is feasible in my EMR.

So I'm wondering if a semantic word change to something like practical was the best thesaurus entry I could come up with. But maybe that implies incorporation of the workflow issues, the cost issues, the
technical feasibility issues.

I don't know that that's viable under your contract or whatever. But I do wonder if there's a better semantic choice. And then I want to talk about the use of this score.

So my dream scenario, and I used to be a measure developer way, way back in 2009 before EMRs actually were really on the scene, or electric quality measurement, I should say.

I mean, my dream would be that an organization like CMS would, I'm thinking of kind of a receiver/operator curve model where, you know, there's value and now this new feasibility.

And that when they choose measures to be in programs, they're maxing out the area under the curve of those two variables. And in the same way, I think the axes are different for our measure development colleagues.

But certainly the available
evidence, the need, and then also now this
feasibility could be an additional axis for
them to chew on.

It certainly would be my hope.

Based on the number of committees I'm asked to
be on, I'm assuming there's not a ton of
people with the expertise in this group
running around to staff all these various
initiatives, and certainly not at the provider
institutions.

Nobody at my institution speaks any
of this language at all, which is, I'm sure,
the case in most places. So that would be my
hope for the end users.

But at the end of the day, if we're
able to generate a score that can kind of be
the axes for these different permeations and
thinking, people in this space, I think that
would be incredibly valuable.

MS. CHRISTENSEN: I will step back
to all those comments from just a moment ago,
so we're kind of bouncing back and forth.
Sorry.

I would also support if we change the word tool to something more like framework, just from the perspective that our measure development activities are often very organic.

It would be great to be able to have some data out there to be able to go and show to the workgroup well, that's, to use your word, high cost data element.

Is that really the way we want to do this, because oftentimes, they are of differing opinions about the actual feasibility based on their own personal experience.

So that would be really good. What we wouldn't want to do is anything too prescriptive that we would have to fill out that would take away from the organic nature of the development.

DR. BUTT: Yes, so I was sort of intrigued about JD's sort of suggestion, or at
least, comment that potentially one way to try
to get a sense of what all the different major
EHR vendors are planning to do in the next six
to 12 months is through some kind of a regular
survey methodology, because I was actually
thinking about it myself the other day because
the only other thing that you can sort of
grasp on is the standards and certification.

And that, we know, is not granular

enough to know everything that we need to know
for eMeasurement.

And so my question to the EHR
vendors would be that how practical or
feasible would such a survey be if it is felt
to be an important component of what elements
can be expected to be had beyond the
certification level of granularity.

MS. MARTINS: Zahid, to your point,
I think that's a two way conversation. So I
think it has to do with the measures that are
on the pipeline and concepts that may be
important for the measures that we have no
idea how feasible they are and then what would be the cost of adding them to a measure.

But also, how workflows and EHR documentation evolves as EHR adoption evolves, as well. And even for your specific EHR installation, there may be customers who are asking for, you know what, I think I can document this now in a structured format.

And hopefully, that evolution will occur as EHRs become more and more mature and organizations use them more and more for their own internal purposes.

So when I say two way conversation, what is it that, on the measurement side, we're looking for and how does that fare against what EHRs do today versus also how are EHRs evolving and how can we leverage what EHR functionality and how it is evolving beyond the certification criteria.

DR. BUTT: I understand that this will be an ongoing process and different EHRs will be in different stages of it, as you
mentioned earlier.

My question is only that is there
some way to formalize the process of finding
out what elements are reasonably expected to
be available.

DR. BREGMAN: I'm probably going to
support Paul on this issue. I don't think
that the measure developers need to be
responding to what the EHRs are doing. I
think they can drive what the EHRs should do.

And I personally would not have a
problem if the certification criteria were
 stricter and more granular, because it would
basically say look, in the preference example
I gave before, you need to create a tool that
allows a structured, accessible, communicable,
patient accessible way to record patient
preferences and, you know, match to codes.

MS. MARTINS: And I agree with you.
I'm not saying that we should be driving
measure development by how EHRs are developing
and evolving.
But we certainly can take that into account, given what the priorities are in terms of measure development.

DR. BREGMAN: I think that that would be incorporated in the feasibility score.

MS. MARTINS: Yes, and --

DR. BREGMAN: When we say, you know, yes we're already working on that and yes, two years from now it will be available, that would be baked into the feasibility score from my perspective.

MS. MARTINS: What I would like to say in addition to that is in terms of the population, as measure developers, I don't think it would be meaningful enough to just reach out to EHR vendors because there are vendors with multiple systems.

So you have the product level in terms of feasibility. And then you have the EHR installation, which is I would guess, the same as provider level.
So when choosing a sample of institutions or vendors or organizations to work with in feasibility, besides the provider, or what we typically do in paper based is choosy, try to cover the population of organizations that we see as using these measures for quality improvement.

So we're talking, and typically on the inpatient side, you would be talking about hospitals. We want to take into account rural, urban, some of the demographics surrounding the hospitals.

In addition to that three measures, we certainly need to take into account different EHR vendors, so market coverage, different EHR vendors, different products within EHR vendors.

And then the different installations. And the different installations need to take into account the hospital demographics or the provider demographics. And that brings feasibility to
a very unfeasible level for measure developers.

So this is what I struggle with as a measure developer is how much do we bring upstream to feasibility so that we can kind of assess where we want to go with the measure and how we're going to move forward versus actually piloting this, which would validate the feasibility assessment and how we moved, how we decided to move after the feasibility assessment.

DR. OVERHAGE: This just builds a bit on your comment. You know, the EHR I know best today, you can enter any data element in it today.

I can say that categorically, there's no data element that you wanted to record that you can't record. It all gets into it.

So in some ways, you know, this sort of going and asking the vendors can they do it, to your point, I don't think helps us very
much because yes.

The question is how much work, how hard? And the variability, and you know, we talked a little bit before about, well we shouldn't let providers be too variable in their workflow or whatever.

And I just don't know how realistic that is. I mean, if you take a very simple, by my standards, anyway, observation like the fractional shortening or the injection fraction from an echocardiogram, there's these incumbent measures.

I know of at least 18 different ways that that data element gets captured today in customers I'm intimately familiar with. So I don't know that asking the vendors helps us very much, frankly.

I mean, they may be able to be a proxy for their customers, to a degree, but it's so variable. Like you say, urban, rural, size, what other systems they have, if they're legacy, on and on.
And that's why I was trying to suggest, I think to some level we're going to have to do some lumping here. I think if you think about at the individual data element level, you die.

And you've got to categorize them somehow as this type of data element, this type, this type. And then I think you can more realistically do either through a survey methodology or through other ways, start to assess the difficulty of that kind of data element.

So then you get into the issue about well, what is that taxonomy? And in some ways, I think that's what the QDMs sort of gets at, but not quite all the way there.

MS. MARTINS: Can I respond? I'm sorry. It's going to be really quick because I guess we go back to assessing what the QDM feasibility is versus what the individual meaning of the data element is.

And I do think that's dangerous
because a single measure may be built upon the
concept of gestational age we're looking at
term babies or term mothers.

If you don't get at the feasibility
of gestational age, you don't know how
feasible the measure is. So data element
specificity, from a measure development
perspective, is critical at the level of
specificity that the measure is looking for
it.

DR. OVERHAGE: Or putting it into
the right bucket.

DR. SIMS: So I'm the QDM
subcommittee, one of the co-chairs. Sorry,
that's a disclaimer, but I do think that
there's a rational way to do what you're
saying, Marc.

And the QDM, I realize it's not
perfect. And I know I'm mispronouncing your
name, pardon me. For doing that within the
QDM some lumping ability. And we've already
gone down this path with a little bit of the
style guide.

And the other thing is that the feasibility, this is one of the reasons I'm suggesting we think about a semantic change from feasibility to something like practicality because feasibility, yes, it's feasible. It doesn't mean it's practical or, you know, justifies the cost, et cetera.

MS. MAJOR: I just want to kind of echo a couple of the points that Aldo and Keri were making, but also to say, I mean, it's true that there is a lot of variability and you can't just go out and do a survey of one EHR vendors and you have to kind of get the context of all their various products and implementations and whatnot.

But I do think that having some level kind of setting of the available data elements and then making that kind of a dance, right, between the development process and what's currently available.

So it's not that what's currently
available dictates the development process,
but that there can be an opportunity for kind
of pushing to the next level within that, but
having some sort of baseline.

And the reason that what you were
saying kind of ticked me off on that or kind
of pinged it in my brain is that I think in
some ways, part of what we answer, or at least
in the processes I'm familiar with in
feasibility testing now is well gosh, can you
find a place that can do it, right?

Or can you find a couple of sites
that can do it with their vendor and with
their implementation and with their current
workflows.

And so, to Aldo's point, that's not
always, like, the best question you want to
answer, right? Not just that you have one
site that can do it.

So just again, to kind of reinforce
what Aldo and Keri were saying, that if
there's a framework, right, that can be used
so that that kind of measure development dance
can happen, if that makes sense, and that we
have a set of kind of how we're going to
describe the outcome in terms of a scoring one
through five or however, that is standard and
that we all know what the scoring outcome
means, we say it's feasible to an extent of
three, four, five, whatever, that we all
understand what that means.

And what you get out of one
feasibility testing process is the same that
you get out of another. We've all kind of
answered the same question.

DR. BURSTIN: Yes, it's interesting
because I sort of share your concerns about
the word feasibility, particularly as it's
tied to testing because I think it's when you
say feasibility testing that people really get
lost, because testing is really about the
reliability and the validity of the measure,
which is later.

But I do think there is something

about that very early stage when measure
development is happening and it's more
iterative where you could almost imagine some
sort of virtual marketplace, not the right
word, perhaps, but where the vendors and the
providers, the end users have some way to
provide input on how feasible currently or
practical it is to collect these data.

And the potential costs and all the
gradations we just talked about. And also at
the same time, have something that allows for
the developers to make the case for the value
of that data element.

So you begin to triangulate that,
and then the vendors get a good sense of this
data element's really important. I can't do
it now, but boy I had better get this one in
place over the next two years.

And the developers can kind of take
a step back saying you know, this measure's
not going to work for a while yet, it's clear
it's just not out there.
Let's rethink, you know, for example, do I really need to know if somebody, just to use Joint Commission example, you know, do I really need to know that somebody is on NIH protocol?

You know, how much of that is really small numbers that we could just move beyond knowing how incredibly difficult it is? But it just seems to me like a logical sort of coming together, of both sides coming together rather than it always being just the vendors here.

But it's really what's the value, what's the feasibility and bringing it together in some sort of shared space where people can comment and bring forward their thoughts I think might really add value, perhaps built off the backbone of the QDM.

DR. LIEBERMAN: All right, at this point, can we open it up for member and public comment?

DR. WINKLER: Arnika, operator?
Hello?

OPERATOR: At this time, I would like to inform everyone, in order to ask a question, press star then the number one on your telephone keypad. Again, to ask a question, press star, then the number one on your telephone keypad.

DR. LIEBERMAN: Okay, we do have a comment from the room.

MS. CRAWFORD: Hi. My name is Alyssa Crawford. I'm from Mathematica Policy Research. And I just wanted to make a really quick comment.

   Something Aldo mentioned before I wanted to reiterate and say that I think it's really important to consider, which is that I think the overall goal of doing these feasibility assessments and testing is to really document what the potential barriers to feasibility are so they're not surprises when we interact with them in implementation, and then to be able to identify potential work-
So I think in some ways we want to encourage measure developers and providers and vendors to consider all of the possible options and to look very broadly.

And I think the guidelines should reflect that because yes, we need to have some sort of easy way of saying, you know, how feasible is this on a spectrum.

But at the same time, the projects that actually go further and ask more people are going to identify more barriers. So that doesn't necessarily mean that the measures are less feasible. It just means that they've identified more of the potential problems down the line.

So I think this guideline is very helpful and the guidance that's going to come out of this project is very helpful, but it's about really helping measure developers and vendors and providers and all of the other stakeholders to start continuing to think...
along the line and not encouraging them to think restrictively within a certain number of sites. It's really about thinking farther.

DR. WINKLER: Operator, are there any other questions?

OPERATOR: At this time, there are no further questions.

DR. WINKLER: Okay, thank you.

DR. LIEBERMAN: Okay. So it looks like lunch is here. So I think we will break, is it for a half an hour? Half an hour, and then we'll get back together and, I think, start trying to actually come up with what this framework looks like.

And again, just to kind of quickly reiterate, I think what we heard was that this framework should be used as a communication vehicle amongst the various stakeholders, and using the generic stakeholder term there.

But I think it's useful there. So it gives people a way of talking about feasibility or of the measure. And for that
last comment, I think it's very important as well.

It exposes where the cost is in the measure to further that discussion about, you know, how to best do the measure. So we'll think about that over lunch. And then get back together at 1 o'clock.

(Whereupon, the foregoing matter went off the record at 12:34 p.m. and went back on the record at 1:14 p.m.)
DR. LIEBERMAN: All right, so welcome back from lunch. So we have a couple more hours, and I know that some people are going to have to start leaving to catch flights before the scheduled adjournment time of 4 o'clock, myself included. So we'll kind of get as far as we can in the time we have.

I foresee us coming up, again the task was to come up with a set of criteria around this feasibility report or feasibility framework or feasibility assessment. And as we heard right before the break, I mean I think the idea behind this is a communication device amongst stakeholders, a way of succinctly summarizing some of the feasibility of the measure and identifying issues with the measure and what might be impacting that feasibility.

So we're kind of tasked with coming up with what should be included in that
document or in that report. And just to give a little context, you know, I think it was Aldo mentioned this isn't a vacuum, work has been done here before.

But we're going to start with, I'm going over some ideas that have been bouncing around the NQF over the last couple weeks and months, and then we'll also, I think there's a slide in here on the HITEP criteria on the data element part, and then we'll go from there. So I'll let Reva talk through this.

DR. WINKLER: Okay. We put this up a little earlier. Again, with the report we're envisioning having, you know, you always need the picture or a schematic or something to keep it from looking totally boring, and trying to capture the who and what over the timeline, all these sort of multifactorial elements of the matrix.

And so we're interested in seeing your reaction to this. I can honestly say that right up front I noticed that the first
green box should have "developers" in it, at least, if not the first two green boxes.

But again we've been working on, you know, trying to figure out how do we pictorially present sort of way of conceptualizing all of these various elements to assist stakeholders in understanding what it is we're trying to accomplish here. So I'd appreciate any of your thoughts or feedback.

DR. LIEBERMAN: Yes, Paul?

DR. TANG: We've all been struggling with this word feasibility. I wonder when I look at this timeline it looks like it's the same thing happening in all phases. And I wonder if there is something that we do prospectively as you're even starting the measure concepts, we do something, the agile development, and then we do a feasibility testing or something test.

Do you see what I'm saying? There's different kinds of activities, and maybe that helps us, one, in describing what we would see
as the normative process to coming up with
developing new eMeasures. I don't know
whether I've said that right, but right now
they're undifferentiated and we might be
losing a little bit of what should happen.

Another analogy I'll use is user
interface usability. If it's usability
testing at the end then it's too late. And
there have been a number of, in the
environmental scan they mentioned this. If we
do the design and if the users and the user
interface experts are there at the design side
that's a critical role. We're trying to say
the same thing for feasibility.

DR. WINKLER: Any other thoughts?

DR. BREGMAN: Can you explain again
what the difference is between the two rows?

DR. WINKLER: Well, the first one
was talking about, you know, the data element.
What about the measure, what characteristic of
the measure is being addressed?

And the second is more about who's
working on it or involved in that part of the assessment because, you know, the sense of everybody is going to be involved at some point but maybe not everybody at all points, and the collaborative nature of it.

But again --

DR. BREGMAN: So does that mean the --

(Off microphone comments)

DR. WINKLER: Oh. Okay. In the first, the data level.

(Off microphone comments)

DR. WINKLER: Okay, underneath the Data Element, Workflow Processes, EHR Vendor and System Level, and it's obvious to me that that should also include Developer. Under Measure Logic it's EHR Vendor or Local Provider, and the question is that maybe Developer in that box as well. And then Measure Score is more at a Local Provider assessment.

DR. BREGMAN: I don't understand
that part.

DR. WINKLER: Well, I think that that's sort of the end of the road, being able to pull out a fully, you know, the measure result.

DR. LIEBERMAN: I think that's more under the reliability, validity, kind of feasibility assessment, tails off in that stage but that's in measuring those other properties of the measure.

DR. RADFORD: But then you also have the people that are using the score for accountability. They want that score too. So that's a product really.

DR. BUTT: So are they implying in Measure Score, the performance?

DR. WINKLER: That's the performance score.

DR. BUTT: Okay, so maybe if we, because we're talking about score in this context, perhaps if we could clarify that it might help.
DR. WINKLER: Sure.

MS. MARTINS: I think certainly the subject of what we're trying to assess in a feasibility assessment and in a reliability and validity assessment is largely overlapping. The way we go about and do that and the level of statistical validity of our conclusions are certainly different.

So I think the feasibility assessment is, and we've talked about this, more of a qualitative, may have some quantitative aspects, but to kind of steer us in the right direction where we're going with this measure and make sure that we're not creating a measure that in the end we've invested all of this time and effort and it's useless.

But then there is the assumptions that we work upon given the feasibility assessment need to be validated either both in the larger population of sites and how comparable the data may be across a larger
sample of sites, and also at the very specific
level of patient data from EHRs that are in
the real working world.

DR. WINKLER: As the conversation
goes on we'll keep tinkering with it and any,
you know, get your feedback as just a way of
schematically trying to describe this fairly
complicated and intricate process.

DR. LIEBERMAN: This is from ONC
trying to get a framework about thinking about
data elements as well, and really, you know,
hoping to find, upper right, all this in the
upper right box with high value.

So again, essential to quality of
care and that are structured and present in
the EHR, that would be a data element that
would score very well. And then on the other
end is the low value, elements not very
significant and that are unstructured are ones
that you would not find as much value in. Go
ahead to the next one.

So this is back to the work from
HITEP-I, where we started to look at trying to assess data elements, and these were the criteria that were used in that group. And you can see there's both, the scale is one to five for each of them, but they were weighted somewhat differently.

And I can read through them, but there was, you know, Authoritative/Accurate Source. Is the entry in the EHR from an authoritative data source? What is the accuracy of the data element in EHRs? And then Data Standards. So again is it using a nationally accepted terminology standard in a structured format?

And then Workflow Fit. This gets to, you know, a lot of the discussion this morning. So is it captured in a typical EHR workflow? And again we could look forward and say, will it be captured in a typical EHR workflow as well?

And then Availability in the EHR. So that one was just, is it currently there?
And then Auditable, can we look back and see whether or not it was accurate?

Paul, how many data elements did we score using this method? Do you have a --

(Off microphone comments)

DR. LIEBERMAN: Hundreds, yes.

Okay, and that was at the data element level for kind of what we felt were high value measures at that time. Okay, go ahead to the next one.

And then this is, again this is for more kind of from some thinking over the last couple of months about this as well. Really, and these were things that were called out. And you can see there's a good amount of overlap sometimes from them.

So again it's the, captured during the course of patient care, and I would say kind of routine course of patient care. Data found in structured data fields. Data element definition is precise and unambiguous with appropriate granularity to represent the
quality concept.

Data element and associated value set use standardized vocabulary. And then the last one, interoperability complexity, that's the one in reading through I wasn't as sure exactly what that was getting at. I think it was, how is that different than some of these other areas? Do you know, Reva?

DR. WINKLER: Yes. I mean we were having a hard time characterizing the concept that, Shannon, you bring up all the time is, certain things are easy and don't require a lot of interoperability, and sometimes the interoperability is so challenging that make it very, very difficult.

We didn't know how to characterize that concept, but the concept was if the data element typically is going to be challenging because of constraints around interoperability and data exchange that has impacts on feasibility. I don't know what to call it. This is the best we could come up.
We're open to any suggestion.

DR. SIMS: Well, I think what I was alluding to was that I think in the coming stages of meaningful use, not that that's the only reference point for how EHRs might grow, but I mean I think we'd all agree that we'd all like to get to the point where we can use, or we can develop and use quality measures that, you know, rely on exchange of valid data between institutions.

But that simply is not extant right now. It just doesn't exist. We don't have an HIE in Chicago, for example. Well, we do but nobody uses it. So I think that that's got to be a characteristic.

Now I think what I was alluding to is that, and maybe this is what Paul and I have been chatting about was, I think it's clearly going to be a future state but I don't know if it's two years or ten years, hopefully not ten. But I think that's got to be a realistic consideration when we think about
the feasibility of these data elements is,
what can we do right now or in the next couple
of years?

So I would say not so much
interoperability but data exchange, I guess,
is probably a better word.

DR. LIEBERMAN: So I would submit
that when we are looking at this there are,
data elements are definitely a key component
of feasibility, but it's probably not all of
the feasibility for measuring. It has to do
with, you know, you can have each individual
data element being available and feasible but
they could exist in different systems, so it
makes it a little more difficult or, I'm not
sure of workflow, but there may be additional
criteria beyond data elements and so we should
discuss that and whether that's the case.

But I think we could start with
talking about characteristics of data elements
both in kind of current EHRs and in some
future state to try to come up with a scoring
algorithm for that and a way of analyzing that information, and then move on to additional characteristics beyond data elements. And I'm open to other suggestions as well.

DR. BUTT: So I think that the issue of structured versus codified/standardized, I think that, you know, things could be structured but they're often not standardized or codified.

For eMeasures, really feasibility. That's the more important one, so perhaps we should incorporate that. Because even in that two-by-two that you showed for ONC, it only talks about structured being high value for quality. But if all of that structured data is based on nonstandard elements then it's not of as high quality.

DR. SIMS: Can I react to that? I'm sorry. So I totally agree with what you're saying. The problem is that, you know, just because something has a SNOMED code attached behind it doesn't mean anything, right?
I mean ultimately what you want to do to standardize data is ensure reliability, generalizability, external validity, right?

But if I'm doing a local mapping at my institution, which is going to be invariable here about what constitutes a blood pressure measurement and so forth given how customizable EMRs are, I'm not sure that necessarily having a standardized code behind it necessarily makes it more or less feasible or potentially reliable down the road. That's my --

DR. BUTT: I think the point I was making is that the more structured it is the more likely that it will be codified. However, you know, if, and I agree, Shannon, that there will be a lot of mapping, but that is the point. That mapping is necessary because the structured data is not codified.

So the mapping is to the code and that's what the eMeasure looks for. It's looking for the code sets or value sets,
however you specified them, whatever code
system you want to use, RxNorm, et cetera, et
cetera.

But at the end of the day it's going
to require that in some fashion, whether it
gets mapped or whether it gets structured,
whether it gets captured in unstructured, but
some mechanism to extract it with natural
language processing or whatever, but at the
end of the day the eMeasure is going to look
for some codified value set. So that's the
one thing.

And the other is that in terms of
the HITEP-I slide where there was a scoring
for the workflow part, how did they actually
get to the score for the workflow? Because
I'm interested in learning more about that.

(Off microphone comments)

DR. LIEBERMAN: People sitting
around saying, oh, I think it's about a four.

DR. RADFORD: I'd like to respond to
this issue of standardization.
I cannot believe I'm saying this
having been on multiple data standardization
workgroups, but I would argue to take data
standardization, which is standardized
definitions that everybody agrees to and
adheres to, off the table for this document.
Because it's really, in my view, kind of a
medical provider issue that we really haven't
dealt with optimally as a group.

I mean when you think about
different entities that have developed
standardized measures, I'm going to contrast
medicine with banking, I mean it's kind of
easy to codify what $1.00, meaning $1 that's
pretty easy.

But we really, in medicine, have a
lot of concepts that have a vocabulary around
them and nobody really knows whether what
heart failure really is and all that so I
really think that that's not something that we
should get at here.

We assume that it's correct or not.
We acknowledge that this is an issue. We acknowledge that the fact that data standardization doesn't really exist can lead to all kinds of interesting things to happen including gaming measures, but that this is really about automating measures such as they are.

DR. BUTT: So can I ask a follow-up question then? So maybe the measure developers can comment on that. Can you develop eMeasures that can use noncodified data but structured data and make it work?

DR. RADFORD: No. So let me just say --

DR. BUTT: Can we say, it depends?

DR. RADFORD: Yes. I mean I think that let's look to the future a little bit. When you talk about noncodified data you're not talking about nondiscrete fields, right, text fields?

DR. BUTT: Well, I'm saying that it needs to conform to some codification system,
and not necessarily but it's captured only as structured fields, because it'll not always be captured as structured data. Because if you get hung up on structured data only, we're going to miss the bigger point that is that we need codified data and accurate codified data however it gets to us in whatever form.

MS. MARTINS: The one comment I would make is that in order to, if you want to use a measure outside of a provider, a single provider, and even within a single provider if you want to compare different clinicians, for instance, if you want any level of comparability there certainly need to be standard definitions.

And I completely agree with you, Martha, that this is bigger issue than just the definitions that exist in the realm of a measure and that societies need to agree on what they mean by gestational age, for instance.

And I'm bringing it again, but that
work needs to happen with the clinicians, and the clinicians need to agree on what they're calling what, completely. We can't go there with eMeasures. All we can hope for is that the codes that are being used to define, because that's part of the problem is that in paper-based measures you could actually attach a definition to what your data element was or is, and in eMeasures you kind of have to, you move away from that.

And incidentally, vendors are trying to reverse engineer value sets to kind of try to understand what we mean. So I wonder if we don't need definitions with value sets as well. But that the codes are the only source for the meaning that we're trying to convey.

And so in that sense, I think we can't disassociate the representation of certain concepts using standard vocabularies for measure feasibility particularly.

And I'm not saying that they need to be, in order for a measure to be feasible that
the vocabulary needs to be used at the point of care and that it is consistently used at the point of care. I think that's more of a role of ongoing reliability, to be very frank, because we don't have the data yet to be able to assess that kind of reliability and validity.

But for instance, in eMeasure representation you may even have a situation where you don't have the codes to represent your concept, and this is due to the level of maturity and real-world experimentation with these standard vocabularies. They're not in widespread use, and so they need their own maturation or evolution I would say.

So there are issues in multiple fronts. But I do agree with you that just because we have a code it doesn't mean that the code is used in a standardized fashion, and that needs to be taken into account in validity and reliability testing and ongoing assessment.
DR. TINOCO: So I don't have a definitive answer and so I kind of agree, but I think it depends. So not all terminologies or nomenclatures are created alike.

Give me a LOINC code, I understand what you mean. Give me a SNOMED CT term, I don't really know what the definition is. I have to look up the definition in a medical dictionary to say, oh, that's what that diagnosis means. RxNorm also, I can navigate the hierarchy and I have an understanding of what it means.

So just to say that we should or should not use standard terminologies and make that a requirement, you definitely need measure specifications, maybe not a requirement of the EHR database itself. It's just not clear to me yet because it varies by subject domain.

And secondly, within some of our measures, and we're getting better, we do have to call out explicit definitions for our data
elements because of a couple things. Either
they're somewhat complex or we learn through
experience that when you say one thing like
cumulative medication duration, and Saul will
remember this one, it seems as if different
measure developers might have a different
understanding of what cumulative medication
duration is and how it should be computed by
an EHR system. So it's sticky.

DR. BREGMAN: I just want to return
to the question you asked that prompted this
discussion which I think you made the point
that, well, you asked whether how much, how
important the data elements and what else is
there other than data elements?

And I would make the case that
almost all the money is in the data elements.
When we sweat these measures out, from our
point of view it's always an issue of, where
are we going to get this data and where are we
going to find a structured format, is it valid
where we find it? What's the variability from
one place to another? Everything is just minor compared to that.

So I think when we talk about feasibility, you know, I love the HITEP paradigm, the HITEP-I paradigm. That really captures all of the issues with data elements and that's most of feasibility.

(Dr. Bregman's microphone is off)

DR. BREGMAN: Well, it was a very good paradigm to start with.

DR. BUTT: So I just again want to make sure that I'm very clear in what I said. I wasn't implying that codification means to select one of the existing code sets.

What it means is that even in all those definition, if they have to lock down a definition then that becomes their code system. A system needs to be in place that everybody can follow and the eMeasures can sort of understand what the data element is. And my comment was really only limited to what the eMeasure consumes, not what should be
captured and how it should be captured in the
bigger EHR.

So from the standpoint of making it
unambiguous, because there's not a human being
for interpretation in the middle, it's got to
have some kind of codification system be it
the existing systems or some new system. But
that's what's going to have to be done to
actually make it work.

DR. LIEBERMAN: So now I propose
that we spend some more time perhaps starting
with HITEP-I and look at those and determine
if we feel those are sufficient criteria by
which to evaluate individual data elements.
And we should also think about how we would do
that forward-looking.

So we've talked a lot about how
would we apply these criteria to what we
expect the state to be in 2014, or 2016, and
how would we work with that in our system. So
comments on these? Does anybody have
suggestions for how we might change these and
whether we want to get rid of any of these or
whether we want to add additional criteria
around data elements?

             MR. KRAVITZ: This may be beating a
dead horse, but the kind of availability, I
think it needs to be qualified somehow based
on some combination of the maturity of the EHR
and the kind of the IHE ecosystem that this
provider lives in.

             So my favorite example would be
outpatient measures that reference previous
inpatient encounters that the patient had. If
you're in an integrated delivery network where
the EHRs are integrated you might see it, but
from my understanding most providers today
wouldn't have visibility to that. And if they
did have visibility it may be stored as a PDF
attachment to the patient's record as opposed
to something you could query against.

             So again, some of the data elements
that are in the measures, you can't score them
in an absolute sense. You really need to
score them in some context. It's going to be
a more nuanced scoring than just yes,
inpatient encounters are available.

DR. BREGMAN: I think you're right,
but I would assume that that's part of the
evaluation of that part is considering that.
I mean when we are to provide this kind of
feedback from Epic we would have to think
about our entire user base, all the different
situations they were in, and we would have to
come up with an answer, a score, a composite
score.

And we would say for this, for
example, we would say, generally yes, because
most of our hospitals are integrated and they
have our system both in inpatient and
outpatient. However, in the case where they
didn't this would be a tougher issue and then
we would put that in the score.

MR. KRAVITZ: Sorry if I missed it,
but is this the questionnaire scale or is this
the final results scale? Because you're
saying if this is the questionnaire scale,
everybody who got the questionnaire has a
different perspective and some people might
grade that element a one if they're in a --

DR. BREGMAN: Well, sure.

MR. KRAVITZ: And some people might
grade it a five, and someone's got to roll
that up to a final score. So is this for
both?

DR. LIEBERMAN: I would say it's for
both in that I mean, I think, eventually we
want this to be a document associated with a
measure. So this would be, I would say we're
talking now about kind of a final score.

You could have different ways of
getting to that final score and part of that
could be surveys sent to EHR vendors or to
providers or something else on those measures,
or it could be, you know, discussion of an
expert panel, whatever that might be. So
there could be different inputs to it, but I
think we are talking about looking for that
(Simultaneous speaking)

DR. SIMS: I was going to say, you'll be surprised when I have opinions. But I think the first two again, to me, are more value based judgments. I mean the assumption for number one is that there's bad data in the EMR, and if that's the case then I don't know. I mean thus far in my experience doing core measures and PQRS and meaningful use and then internal efforts, I mean we consider most data points to be fairly equal. Clearly we know that if a med student writes down a diagnosis and it gets copied and pasted it's probably crap, or potentially crap, but at a certain point, I don't know, that feels like unnecessarily complex.

And on the data standard piece, I don't feel like that should be part of a feasibility assessment. And the reason that I say that is because actually in meaningful
use there were times when the value sets
actually excluded appropriate population.

So I think, I don't know, I just
feel like that mapping, the reason to have a
good data standard is obviously the measures
have to be calculated using that but currently
the way we do things is we map the data
standards if necessary.

And I think that there are bigger
issues about, I think, that the real value of
a standardized data vocabulary behind these
measures is that it improves comparability
between institutions and providers, which it
feels like outside the scope of this
committee. But I recognize a lot of really
thoughtful people put this together so I don't
want to speak out of turn.

DR. LIEBERMAN: Go ahead.

MS. MEADOWS: I was going to comment
to what Shannon just said. For the second
thing for data standards I do think that
that's something we should be thinking about,
because there are cases when we've seen things introduced that don't truly have any kind of standards behind them whether they're codified or could be mapped or not.

So we do have to think about whether there is a way to get some kind of codified standard even if it's through mapping or other capabilities. And there are some things I know that we've evaluated that don't actually have any kind of standard whatsoever. It could be anybody's list of things they think are important.

MS. MARTINS: And I would add on to that that from a measure developer's perspective, again comparability is an important part of assessing whether a measure is feasible or not. If we can't compare rates across institutions how is the measure even worth it? It may in some situations but it may not.

The other aspect of the data standards, and I do agree that we shouldn't
try to force data standards into EHRs, are they even going to get there? How? I don't know.

I think that from a measure developer's perspective I would want at least to gauge the field in terms of how the national standards are being used so that we can kind of start thinking about the potential issues in terms of reliability and all of these just different mappings that are happening and how do they impact the measure rates. So I agree it's important. Maybe it shouldn't have such a high weight.

DR. OVERHAGE: In some ways though this gets back to the whole reason for bringing this process upstream though, is just sort of fix, I mean to recognize and fix these issues, right. So as a provider or, you know, EHR implementer, so if there isn't a code for it don't put it in the measure because, you know, we've got to work back upstream and fix those things, otherwise it is going to break.
So part of me would say, yes, this ought to stay really high, because if it's not then it's not very implementable and we better go fix it.

DR. BREGMAN: I'm just going to propose a sixth criteria, and then that would be, I'm not sure of the right term but it would be likelihood of accuracy or validity, or someone can suggest a better way to describe it.

So an example I will give you is let's consider the sex of the patient. In these categories you would say, comes from an authoritative source. It's mappable to a terminology. It fits the workflow to enter the sex. It's certainly available in the EHRs. It's auditable, and it's very likely to be accurate, right.

Another example would be, when a physician receives an alert that says there's a drug interaction they have to consider and then they have a response. In Epic you have
a choice of responses. One common choice would be benefit outweighs risk. Another one would be, not relevant to the situation.

And there may be other choices the physician's supposed to choose, which one applies to the situation? And you would say, does it come from an authoritative source? Yes, it's coming from the physician. Is it mappable to a data standard? Yes. Does it fit the workflow? Yes.

A drug alert after ordering a drug, is it a standard part of the workflow? Is it available? Certainly. Is it auditable? Yes, as, you know, we can debate that one. And then is it likely to be accurate?

Well, the answer to that is I would say, no. It's very unlikely to be accurate because the physician is unlikely to be thoughtfully choosing the answer that really applies to the situation. So that would be an example of something you would mark low on, likelihood of accuracy.
DR. LIEBERMAN: I wonder if that could be included in Authoritative/Accurate because it's authoritative and accurate source, or maybe they are two different things.

(Off microphone comments)

DR. LIEBERMAN: Yes.

DR. TANG: What is the accuracy of the data element in EHR? I think that's the question we're asking, right? The second question, under description, first row --

DR. BREGMAN: Okay, well, I'm reading that to mean the source is likely. Well, I thought that was a comment on source, but yes, if that applies to that then that would be included in that.

But I think those are two separate issues. One is whether it's the right source, which in that case the physician was the right source, but even so it's unlikely to be accurate. So those kind of conflict.

DR. LIEBERMAN: And would
Authoritative Source, where would that score low? I mean --

DR. SIMS: I agree with that. An example of, you know, they're bantering around the notion of gender identity and sexual orientation and that's going to be collected, at least at our shop, by front desk staff which I want to be a fly on the wall for those conversations.

But I mean I don't know. I think if it's in the EHR it's certainly a legally discoverable element, and from our perspective legally at our institution we consider it to be true, knowing that.

DR. TANG: So that's an example of where that may not a score a five, that example? A back office lab would be an example because it's not the same thing as spitting out from an automated lab instrument. So that's where those are things that might, would not score a five, as examples.

DEBBIE: Another example could be a
result of a diagnostic test that is in a
separate system that had to be then the
results transferred into the EHR. So every
time another human touches something, you
know, there's more chance for error so that
could --

DR. TANG: So let me give a
counterexample to Shannon. So if the patient
was entering their gender identity that would
be called authoritative and accurate. So I
mean that's how that would work.

DR. SIMS: I agree with all that. I
just think we're losing the, I mean the issue
that we have in a quality measurement is not
these fields which we could clearly have some
issues with accuracy. There are keystroke
errors. At our house that's 20 or 30 percent
error rate we've learned.

But it's more of the things like,
was an asthma action plan created? Was the
smoking cessation counseling provided? It's
not stuff that's already mostly structured.
That's the problem. And that's where I think the measure developers need more feedback.

So those are the examples I think we should be focusing on perhaps more than some of the stuff that's already highly structured to begin with.

DR. TANG: I'll respond to that too, directly, like a smoking counseling checkbox. I would rate that pretty darn low which is a message to the measure developers saying, well, why should we do that? In the first place how useful is it in the scientific validity and how good is the data? How accurate is the data in the EHR?

So those are examples, actually, of where this should show up and influence your decision on including those.

DR. BURSTIN: Sorry to jump in, but I'm not convinced that's feasibility as opposed to validity. I don't know that you can, I don't want to confuse the data versus the validity of it. That's just the fact
that's just a nonvalid representation of smoking cessation. It has nothing to do whether it's a checkbox checked by a nurse, a medical assistant or anybody else.

DR. SIMS: I agree with that. I mean feasibility you're talking about a priori development of measures. Reliability you're talking about sticking those into the real world and seeing what comes out, comparing that to human chart review or to patient reported outcomes or whatever.

I guess you're really going to have to clarify what you guys want. Because everything Paul said is absolutely valid and I agree with, but if we're talking strictly feasibility you've got to disentangle things a bit.

DR. TANG: Well, I guess I would, okay. If I really wanted to know whether counseling was done then I would use the audio recording. If I wanted to figure out is there a better source, I'd look and say, hmm, is the
EHR a better source? Probably not. Go beyond that, you know, go next. Is the patient, asking the patient a better source? Probably better than the EHR. And that's how I would use this, quote, "score."

So I think those are some of the issues we're trying to get at. There's a lot, in some sense there's some anti-gaming. And you want to figure the merit of, hey, should I use the EHR to get this information, and this is how that would flow it into.

You won't see it under scientific validity or impact. You'll have to see, yes, this is where I, probably not going to get it from the EHR is, I mean that's the logical conclusion, getting a low score here.

MS. MARTINS: I wonder how that overlaps with availability in the EHR. So as I look at these two criteria I think we really need to clearly define. And I don't know that we need to do that today, but we want to make sure that they're at least mutually exclusive
in what we're trying to evaluate, although I think that's hard but the least overlapping possible. Let's just say that.

And that we're not evaluating the same twice, so that there can be an understanding from different measure developers and different respondents to these type of questionnaires on what the understanding is of what we're trying to get at.

DR. SIMS: I guess just in general I'm a fan of simplification and this feels like an awfully elaborate scale to apply to potentially thousands of data elements in a short course of time.

But I mean I think from a staff perspective I think you guys have got to tell us what you want. I mean everything that's being said is true, there's just so much opportunity to move the bar forward with good as opposed to perfect. I'd love to see that happen, that's all.
DR. WINKLER: In response I would say we've got different representation from the different stakeholders who are likely to be customers of this kind of a framework.

So I'm going to ask you, what is it that's going to be useful to you? How is this going to, you know, influence your work as a developer, as a vendor? What are the elements that are going to give you the most useful information that will help solve the problem that we've all been discussing going forward?

DR. BUTT: So I think that in the second section, the data standards, are we sort of saying that unless it's structured format that it's not going to score high? That's number one. So we should probably discuss that.

And then also, is it locking down terminology specifically as the only standard here? Because one could potentially be getting codified data from classification systems and so forth, and I'm not an expert in
that to say whether that falls under
terminology or not, but that may be another
thing to make sure that we're not
inadvertently locking down into something that
is very, you know, sort of narrow for codified
and standardized data.

DR. LIEBERMAN: Well, I think for
data standards with the QDM, I mean all the
data that's being used in electronic measures
is going to be structured. I mean the
measure's going to be evaluated.

Well, I guess it also depends on at
what stage we do this feasibility assessment,
because I would assume that if you're looking
at a QDM data element it's a piece of
structured data at that point. And the only
issue here is whether there's a terminology
standard associated with it or not.

If it's some type of information
that you're trying to get at that hasn't yet
been associated with a specific terminology
standard, would that score lower?
DR. BUTT: So, you know, when structured is used I assume it to mean that it is conforming to a certain structured form of data capture, right, or is basically, for example, just as a quick example, a diagnosis code, let's say it's SNOMED CT.

Could be in a problem list which is structured or could it be embedded in a note that is then pulled by whatever method, would that same diagnosis that's part of a note electronically extracted not be valid because it's not captured structure?

DR. OVERHAGE: No, I think I hear your question. Like I said, you clearly have to get to a point where you have a structured coded element to execute whatever kind of logic you're going to execute on it. I think we all kind of agree.

And I think your question, I think I got it, is what does it mean for that when we score this, let's say that you could extract the gender of the patient from the brain waves
of the person at the front desk and get it
available to do the measure, does that still
count? I mean that's sort of the extreme
eexample.

So in other words, how do you get to
the diagnosis that the patient has diabetes?
One way to get there is you require somebody
to enter the problem list or dictate into, or
I'm sorry, enter into a problem list as they
have diabetes.

But you could just as well extract
it from a note that they had written and turn
it into a structured coded element or have a
predictive model that takes into account the
patient's glycoside hemoglobins and their
pattern of care and their medication usage and
determine that they're diabetic, and that
should be just as valid.

DR. BUTT: I guess we can specify
exactly. Because my only concern is that in
general sort of usage the word "structured" is
interpreted in that context that it's got to
be captured in a certain sort of structured fashion, and it kind of limits it to how it has to be captured. And that's all --

DR. LIEBERMAN: Now potentially would that then end up in the workflow part? I mean if it's information that is captured somewhere in an EHR system it's not coded at that point. It's not in structured data element and you have to get it there. That's going to take a hit on the workflow part.

DR. BUTT: See, but I make a distinction being coded and structured, right. So it has to be coded. So that's what I'm saying that perhaps the emphasis should be on codified data. It could come from a structured capture source or non-structured capture source.

So like the example I give, that structured data capture is typically, for example, in a problem list. There's a certain way to capture that information, right, in a sort of a list fashion or in a pre-coordinated
fashion. But if the same data element were part of a progress note or something that was within the system, look, it would be codified.

Let's say if somebody has a, when they're making a note they have mapped their local term of diabetes to a code and the physician, and this is actually in practice in the EHRs where a note is being written and the selection of that code is being made within the note where there is a section of let's say diagnosis. Now it ends up from there into the problem list as well, but I'm just saying that maybe it's splitting hair, I don't know.

But that's one of the things that I think if it implies that it can be only captured in a structured fashion that's the part that a lot of physicians have the most trouble with when everything is structure, structure, structure, you have a checkbox or selections.

But there are some examples where in most cases -- you are correct, Helen. In most
cases the two are the same. That for it to be
codified generally it's captured in a
structured format.

But I'm just giving ourselves some
wiggle room in terms of the whole issue of
natural language processing and so forth where
data is extracted from notes and then some
algorithm converts it into a codified thing.

Is that going to be not possible to
be evaluated, that scenario, if we say that
that's not structured data capture, or would
that still be structured data capture if it's
converted into a structured? I guess it would
take unstructured data and convert it into
structure so maybe it would be okay.

DR. TANG: Correct. Because if you
read the, it says, is the data element coded
in a structured format? It doesn't say how
it's captured. It's just, does it exist in
the EHR in a structured format?

DR. LIEBERMAN: Aldo?

DR. TINOCO: So many reactions to
this. It's very interesting, because in my mind, in response to the prompt I'm trying to use this tool to assess a measure in my mind that we're working on right now.

First of all, I think it's very difficult unless we know how to assess data quality over the content of an EHR system as a measure developer. We know that that is an issue onto itself that affects feasibility but that's way over here. That's data provenance.

Assessing the accuracy of the problem list content, we know the reality of what's in the problem list, but we still refer to the problem list in some of our specifications, or at least diagnoses. So I'm a little wary of asking in the measure development process to assess data quality.

I'm not saying we shouldn't do it, but I think it could really distract us from identifying feasible data elements and then moving on to some of the harder work to say reliability and validity testing. Granted,
data quality issues will make their ugly heads known during reliability and validity testing, so there is this interplay we have to figure out. I just don't know if it's here.

And Shannon said, let's make it simpler. I agree. Scale of one to five, maybe. Scale of one to three, ah, now I know what we're talking about, you know, a little simpler. That's going to be helpful.

To be concrete for the data standards item, if something is available and coded using a nationally identified standardized terminology I'll give it a three. That's what I want in a structured field.

If it's structured but it doesn't use a LOINC or a SNOMED or an RxNorm but it's consistent, okay, that's a two. If it's in free text only that's a one. So that's how I would try and operationalize this tool as a measure developer making these decisions. And let me stop there, because I could go on.

DR. TANG: I was going to answer
Reva's question, but I'd love to hear the developers, whether this works.

So if you sat at your committee or this measure authoring tool and you were looking for a data element to fit in your definition and up popped this score, and you know that whether it's three or five, you know, whether it's 100 or 30 that this was a very high score, and you knew if I picked this to put in my definition the data in the EHR came from an authoritative source, it was very highly accurate because it got threes in your scoring, that it was coded in a nationally standardized format, that it generally fits the workflow of the provider so the burden was low, available today, which means you can use it today versus 2016, and you could audit it when you come in, would that be a use to you?

DR. TINOCO: Yes. It's a very important piece of the puzzle.

DR. TANG: I'm trying to answer Reva's question.
MS. MARTINS: That's nirvana. When we actually go out there and all of our data elements fit what you've just said, then our measure is completely feasible. It's easy. It's doable. It's perfect. It won't upset anyone downstream.

DR. BREGMAN: Well, that's great. But then Shannon said it was too complex, and then, Aldo, you made a reference that we should leave the accuracy of the validity testing at the end. Are you suggesting we take that out?

And Shannon, what would you take out if you wanted to simplify it?

DR. TINOCO: So if the information is available at accuracy or if we make an assumption about accuracy, I'll take it.

DR. BREGMAN: If we know, you know, that example I gave about the drug alert, we just know from experience that it's not going to be reliable information, I wouldn't want to not tell you that if knew it.
Now you could still want it to go through validity testing to confirm that but I certainly would want to let you know if we had concerns about its accuracy from the start.

DR. TANG: I can tell you an EHR's drug alert would not score anywhere close to a five or three.

DR. BUTT: So did the HITEP have the type of definitions for the scoring as Aldo was implying or suggesting? Because I think that may be a very important part of this to define what a one is and what a two is and what a three is for each one of these components.

DR. TANG: One to five was easier to get people to score. I don't think there's anything wrong with thinking about the three because that could also work.

DR. BURSTIN: Except we didn't define what a one, two, three, four or five meant, I think is what he's saying. Instead,
attach definitions to what a one is, a two is
and a three is. I like threes better --

        DR. TANG: It was easy for him to
use in that example of standard. Try one of
these others but it won't be as easy. But yet
there was fairly good agreement in terms of in
people's minds. You sort of looked across the
room and you could tell, because actually the
consensus process was very quick, which is a
signal to say people understood what the
difference between a five and three were. But
that was a good definition for --

        DR. LIEBERMAN: I'll just make one
comment. If we're looking at trying to
develop a score for a data element I think
that we, just from the previous discussions
here, workflow fit should be much more heavily
weighted.

        It sounds like that's what people
are really interested in knowing and that's
where a lot of the cost of the element comes
from. So right now, you know, it scores a
little bit. We kind of get lost. A low score in workflow should be very apparent.

DR. SIMS: So Howard, if I can, sorry. I guess from my perspective, and I know I'm a reductionist but I mean to me, data standards, workflow fit and availability are all kind of, they're related things.

I mean I think availability within the EHR is heavily contingent on if it's part of the workflow or not already. And when we say availability, what we really mean is structured data availability. So those would be the kinds of things that I might consolidate down. I mean if it's available in a free text, no. Not very many of us have natural language understanding and processing ability to translate that.

So I think for me at least we probably need to focus on things that are captured structurally, although certainly they would count if that's the way they had handled it.
So I like accuracy, I think that needs to be retained. But I think you could get down to, personally, I would think three, and they would be workflow fit, structured availability and accuracy.

DR. LIEBERMAN: Sounds like Shannon's making a motion to remove auditable.

DR. SIMS: How can we have a structured data element that's not auditable? I don't understand. Is there an example that comes to mind?

DR. TANG: There's replacing. Instead of tracking adversion there's non-version, so you have no idea what they saw at the time. That's a killer from a --

MS. MARTINS: Gestational age. You documented once, it keeps getting replaced and added. Who knows when it'll stop and how do you know at the point in time where you needed to know what the gestational age was, whether a specific course of treatment was appropriate or not? You lose that.
DR. TANG: And just a note on the fourth one, availability in EHRs. That was the proxy for whether it could be done today. So that was a different, that's an orthogonal, it was put in there because it was to assess today's work.

So in what we discussed today we could have categories. In fact, I still like the stages because it'd just give you sense of when it could be ready. But that's what that's for and so that could be a different dimension to this.

DR. BREGMAN: I liked Shannon's suggestion about getting rid of or combining availability and workflow, or getting rid of availability. And, you know, Rute's point about auditability is accurate. And, you know, certainly we're going to keep data standards and we're going to keep accuracy measure whether it, you know, accuracy is really, I guess, part and parcel of accurate source, so really the first one is about
accuracy.

MS. MARTINS: And I would say that availability is actually partially authoritative source and partially workflow fit. And so it's kind of assessing the two pieces at the same time.

DR. TANG: Well, try to dismiss availability because that was basically a timing. So pretend that wasn't there.

MS. MARTINS: Well, if availability wasn't there I would still think that workflow fit and authoritative source would still cover what we're trying to do against, and again I think that's important, against what are we scoring this? Is it the average EHR? Is it the certified EHR? Is it, what is it that we're scoring against?

DR. TINOCO: Just one more comment. What's helpful is, what's the process by which I actually determine if the answer in the EHR from an authoritative data source is a true state or non-true state? How do I do these
things? How do I ask someone else like a
vendor or a provider to answer these
questions? So as I'm trying to figure out,
well, how would I do it? And it's not that
clear.

DR. LIEBERMAN: Well, I mean I think
that's a good question. So the last time that
we did it is as a consensus process with an
expert panel. Now how it's going to be done
in the future, I think, is something that we
can make recommendations about.

DR. BUTT: So I think along those
same lines, I think that certainly the data
standards part of it is probably the easiest
one for expert panels to determine.

But I think that with the workflow
or availability, that's where you really need
a larger sample size, much larger sample size
whether it's done through the vendor or
directly. Because for it to be valid it can't
be an expert panel, I think.

And then I think the only other
issue is that if there's no weighting assigned to availability then how do we sort of determine the whole discussion we had whether something is available today, whether something will be available in 2014, whether something will be available as an aspirational goal in 2016 and beyond? How would we incorporate that into the scoring?

DR. BREGMAN: The answer to that is you create a feasibility score for now and you create a feasibility score for later.

(Off microphone comments)

DR. BUTT: So it would score high if it's available today and it would score low if it was available in the future? Is that kind of what you're saying?

DR. LIEBERMAN: Availability today and availability at, you know, some predetermined two years out, three years out, whatever that might be. So a measure developer may be very interested, may be thinking three years in advance. They don't
really care if it's available today.

(Simultaneous speaking)

DR. BUTT: So you would have a score
for today and then another score for future.

DR. BURSTIN: Wouldn't workflow
potentially change as well? I mean workflow's
not statically helping EHRs. Isn't that, I
mean it just seems odd to make it only
availability that's time-dependent.

DR. LIEBERMAN: Yes, I think you're
right. It could be data, the whole score
probably, yes.

DR. BREGMAN: It would be the whole
score for now and a whole score, and it's all
an educated guess. You would say, yes, we
think that it's this much more feasible if we
had a three-year timeline.

DR. LI: So I wonder, during the
actual scoring is there ever a scenario that
the data is, the availability in EHR the
answer is no, then if the answer is no, how to
score the rest of four criterias?
DR. BREGMAN: I can answer that. I think the essential answer is, everything is potentially available in an EHR, anything. I could write anything and I could codify it and it could be awkward but it could be anything. So, you know, anything is potentially available, so therefore it's just a matter of how difficult a workflow is it. Is it in anybody's workflow? Could you kludge it into somebody's workflow and get it and that would basically come out in the workflow score? So there really isn't anything that's not available in the EHR. You could put the weather in the EHR. All you have to do is find somebody to do it and a field to put it in.

DR. LIEBERMAN: Well, you should be able to get an interface at least to do the current weather.

DR. BREGMAN: No, you just need a window.
DR. BURSTIN: A great definition says, when a device or situation is of great complexity and either cannot be explained easily or leaves the respondent dumbfounded or perplexed --

MS. MARTINS: I guess I'm struggling with the availability in EHRs being a criteria that you would, and again we've talked about different scores. So then what I wonder is, if availability in EHRs is not a criteria at the level of which the other criteria are and simply the staged approach, so all the other criteria are assessed against a certain EHR stage and then there's going to be another overall score for all of these other criteria for another stage. So it's another dimension of evaluation rather than a criteria in itself, correct?

DR. BURSTIN: JD's question was more so, how can you rate any other criteria if it's not there, if it's zero? Workflow, it doesn't make sense. It's not there. But I
think that goes back to Howard's point. It's
never going to be rated as zero, presumably
it's a one, two or three. So it's gradations
of it, I guess.

DR. BUTT: But that's where that
gated concept might come in that you first go
through the first gate. If you can't get
through that until 2016, well, I guess you
still do it and say that this is a score for
that time period, yes.

DR. LIEBERMAN: So it seems to me
though the data availability in EHRs still
probably needs a score in that even, you know,
couldn't you look at something that's not
available today or that scores low today and
you think is going to score better in two
years but it still may not score a three, I
mean you still need some differentiation about
how available it will be at the next stage.

And that could be another kind of
expert opinion type of thing where your system
will not allow it to be captured but you don't
really feel like it's going to be that accessible at that point.

    MS. MEADOWS: Well, and that's a good point, but I think that brings us back to the whole cost versus benefit discussion too. And I don't really see that as a score here. So as Howard said, anything can be built. We said that earlier. Anything can be built in any EHR, but at what cost versus the value of that?

    DR. LIEBERMAN: But I think workflow is cost at this point. I mean I would equate those two, cost with workflow, really, at this point.

    DR. BREGMAN: Well, accuracy also feeds into cost. I could have a great workflow. It's very easy to do. It's very unlikely to be accurate and it's costly to get the right data. So it's not just, not quite workflow.

You know, my point about the weather was when you talk about feasibility and now
feasibility in the future, if I were to grade weather in this scale I would say low feasibility today and unlikely we're going to build a weather module in the next three years, and so low feasibility in three years from now.

And if it was another more clinically relevant example then I would say, yes, I could imagine that we probably are going to have some tool like this in three years and then that would have an increasing feasibility in the future.

MS. MARTINS: And also what's the difference between availability in EHR and is there an entry? If the data isn't there it's not going to be available and there's not going to be an entry. So again part of the availability is captured by the source and part of it is captured by the workflow.

And what I'm struggling with, and I'm sorry, it will be the very last time I say this, but is that availability in EHR seems to
me the set of EHR functionalities where answering this question are against.

So is the entry in the EHR from an authoritative data source against what, what EHR, what, right? And then the combination of all of those responses was all of those EHR installations would yield the score, the overall feasibility score which I would think is the availability in EHRs.

So maybe I’m confusing myself here, but it doesn’t make sense to me as a criteria in tandem with all the others.

DR. LIEBERMAN: You know, I think that availability in the EHRs to me is if you kind of strike out the currently, but do you expect the data to be available? So it’s subjective. It’s subjective.

It’s kind of looking at it and saying, yes, providers -- well, I don’t know if this is a good example -- but you can document a foot exam in structured data but nobody’s really doing it because it’s hard,
but maybe we think that it will be, so I would say available in the EHR, not really, workflow would be high.

But down the road, three years down the road if we have, for whatever reason, either a new workflow or better reason to do that maybe we can expect for some reason to have it more readily available so we could give it a higher score then. And that's not a great example but that's how I would think about it.

And it probably is, an expert opinion is, or perhaps it could be the currently available. You could look at EHR data and say, yes, it's actually available in most EHRs. It's being recorded as opposed to just can it be.

DR. TINOCO: Two comments. So I like that. I mean I think what we do need is a slideshow or benchmarks. So an idealized simulation environment that we can all come up with, for example, a perfect state system.
How does a given data element compare to that perfect state system? How does it compare to a system that's built on the rules put forth by the meaningful use certification objectives? That's another benchmark.

And another one would be, what does the EHR vendor say based on its user base? Another scale would be, what's the perfect health care organization that actually has been doing this for the past five years in their own QA program?

So the way to interpret these different results is by comparing them to different settings or different examples, and that gives us an idea of, well, when we say feasible, this is feasible. Well, what's the context of that assessment of feasibility?

The second one, availability, I mean I'm not sure if it's a criteria in and of itself for me, but I would have a time one versus time two. And then I'd have to ask people, well, what's the level of effort? What's the cost?
And I would operationalize that on a tool, a grid like this by getting, it's not a row in my mind, it's a column. Current state, future state, and then another column, how can you get me there and is it worth getting me there? And that's how I would plaster it onto this grid.

DR. BUTT: So the authoritative source obviously would depend upon the data element or the type of data element. So would you have to then, in order to operationalize this, define for each data element what the authoritative source or sources would be before you can assign a score?

DR. SIMS: I don't think so. I'll say no.

DR. BREGMAN: I think the proposal was to roll authoritative source into accuracy and we just call it accuracy. It includes that as a consideration. But that's the term we use.

DR. BUTT: How would you define that
then? For example, in that blood pressure example would an authoritative source have to be a doctor, a nurse, MA? Who would be the authoritative source?

DR. SIMS: I think that's a judgment call. I mean again I think we're focusing on, to me what we're trying to avoid is instances where in meaningful use stage when we had a quality measure where we had to assess whether a patient was sexually active or not. That's a bear of a thing that nobody actually collects at least at our house. Whether or not the blood pressure is accurate or not is a good problem to have because at least you have the data to chew on. Things like, when I'm thinking about feasibility, I mean I'm trying to keep the end game in mind here, and when we wanted to do the composite measure about smoking assessment and cessation, it was easy to do smoking assessment because it's already captured in social history.
But we had to create a kludge where we put in some pull-down menu in our template that most everybody in our house uses, but we don't actually use it because despite battering on them for weeks and weeks and weeks on how to do it, nobody does it because it doesn't fit into their workflow.

So I think that, you know, focusing on the kinds of issues that, like keeping that in mind, the accuracy of the blood pressure, yes, probably, you know, humans are rounding up. If you get data interfaced in it's actually going to not be 140, it's going to be 138 or whatever. That's good problems to have in that you look at the reliability of it.

A lot of this stuff we don't even have that option for, and I think that's the stuff we're trying to prevent and so we should more focus on that kind of stuff, I think, the accuracy of blood pressure and gender.

DR. BUTT: So actually I think it's not as much the accuracy of the data itself,
which is more in the reliability side, but it is more of the source of the data, right, that is implied in that first authoritative source.

So the question is that if it's not going to be defined and it's sort of this elusive thing, and most of them are, and I think this was addressed earlier that how is it going to then discriminate between a five and a one or a three and a one? Because if most of these are going to be, well, kind of, sort of, yes, whoever does it is okay, then why have it?

DR. LIEBERMAN: Well, I think that brings up a good question that Reva's going to ask me to ask which is, who's going to do this? So, you know, and that has a lot to go with, you know, how we define it.

Are we expecting this to be expert panel? Are we expecting it to be measure developers? Are we expecting it to be the NQF just in general?

DR. RADFORD: The providers are
going to do it. And I think that that's something that really isn't acknowledged that much. That since quality data is getting collected on us, we actually have a very vested interest in knowing what all of these concepts are.

And there's a lot of measure level, element level testing going on at the provider level that we know nothing about at the national level. And I've thought for a long time that we should have some sort of clearinghouse about these issues because the providers, believe me, are checking.

Now you have to take the provider information with a grain of salt because, you know, they want themselves to look good. And you have to sort these things out as to whether this is a real measurement issue or whether it's, you know, something else. But we don't do that.

DEBBIE: I think we're going to have a host of folks using this, to answer your
question, Michael. Besides the developers, you're also going to have the TEPs because we're going to be using this for de novo measure development and we're going to assess feasibility of what we're trying to create, can we capture this? So you're going to have the TEPs.

You're going to have the organizations that may take quality measures and massage them and call them something a little bit different for their own organizations. You're going to have other people that create, so measure stewards, they're going to be looking at this.

And so I think what, I have a number of different thoughts that I won't talk about now, but one thing that I think is important, whatever scale we propose is that you provide clear definitions for the terminology, because every time somebody says something they bring a little different twist to it. And so authoritative and accurate source, you know,
brings a whole lot of different connotations. So that's something I think that's important that we need to include, whatever we finally decide on.

MS. MAJOR: I don't mean to take the conversation in a different direction so I hope this question doesn't. It's just kind of an honest question, because I'm kind of trying to think this through down to what the end result is.

And what I'm sort of picturing is we've got this big old database that says, here are all the data elements that we care about and here are their scores on our three main criteria and our weights for each of those criteria.

And then as, you know, we develop a measure or think about measures at a technical expert panel level we kind of go in and take a look at this database and say, here's kind of what would be feasible and then kind of build measures based on that.
And then do we kind of take that result of, we've built a measure based off of our understanding of feasibility based on the scores and the weights, and then we say it's feasible, and that's it? Or is there another step that's going to go along with this?

And then we have to maintain that database, right, over time? Is that kind of how people are picturing this happening or am I missing a piece? Okay, I just want to make sure I was --

MS. CHRISTENSEN: So if I can just remind everybody, we listed a lot of stakeholders and I think all of those people play a part. If we let all the stakeholders play by themselves we'll get lots of different answers for the same thing which is not a desired outcome.

And then the other thing is that this stuff makes sense to everybody in this room, but from our work having actual provider organizations assess feasibility of real
measures most of them are not able to do this work, even folks that, you know, have had an EHR for a long time. This is very confusing for them and they don't necessarily all have the expertise that's necessary to do it. Folks like Shannon who have done it are a little bit different because they're well versed in the clinical aspect, in the IT aspect, in the national policy aspect, but that is not the case at most organizations that you've got folks that are able to play in all those fields. So I just think we might need to think very carefully about who's qualified to make the assessment.

MR. JENTZSCH: I think we did one of those assessments before. But if I was going to say how we would implement this as a consumer, I would hope that we'd get some kind of score from the developer, some kind of an overall score by data element.

I would also like to be able to see our vendor have their score based on that as
well. So it's not just who developed it but
the actual vendor that we work with. We would
probably take the same thing internally and
have somebody internally go through and do
their own scoring based on what they know, and
it would be a useful tool if we could get all
three of those things going.

We probably would not be as
interested how well other people are
implementing it because they don't implement
their product the same way we implement our
product, right, they don't have the same
workflows. So it's probably not as useful for
us.

DR. TANG: So I was going to try to try to
answer your question and Catherine's. I think
it was imagined, not that the individual
measure developers would calculate this but
you would get to score in this imaginary
system, that nirvana, and it just was 80 or
100 or 50, and you would know, in fact, that's
how those were scored in the original QDM.
And so that's all you would use, and then if you had 100, it would say those things that I said to Rute and she said would be nirvana.

Your question as far as who would do that, I mean I'm just trying to imagine one possibility is AMIA. It's sort of a clinician informaticist that would know how to score these things, and so potentially it's, well, NQF has to first get the money and then contract with like an AMIA or some group that goes and does this. It's sort of an expert panel, it's consensus development process for these scores.

But hopefully the end user, the developers would not have to get into the nitty-gritty and just would use this overall score to say, hey, of these things to choose from, hmm, this one looks better, and go with that.

And then it would be NQF in this case would be the maintainer, would make sure
that each new, and then clearly, new data elements you come up with would be submitted, get scored, and then it goes back out in the QDM and you'd have access to it in the, for example, the MAT.

DR. LI: I'm not sure my comment is directly relative to the L-score criteria, but in my mind I think these common data types are wrong for a long time. We just, you know, take another look from a quality scoring perspective. And also most of the common data types may already very well specify that by other, amino one, amino two, specifications.

So I think we need to find that they also leverage the precedent results. So for example, the CDA. There's a lot of data already exchanged in the CD format. We know these data are structured data with standard terminology. So for this data, I'm not sure we need to reevaluate or reassess the quality scores. Just a thought. It's more like a cross-checking the other standards.
MS. MARTINS: So two comments. One of them is, so it seems that we are talking about not only about a framework to assess feasibility but that all developers would work on the same set of data to assess feasibility, the same sample of data elements, the same sample of providers and vendors, and that strikes me as extremely biased.

Because if we have a single source for all the measure developers as opposed to conducting a feasibility assessment that actually focuses on the providers on which the measure is focused and has a potentially different market coverage, we could definitely get to different feasibility assessment results and they may be better or worse than the central database. And I'm not saying that the central database shouldn't exist, but is it the end-all of feasibility assessment?

And then to JD's point, and forgive me because I know not what I speak of, but I'm under the impression that CDA only requires
narrative descriptions of the data elements
that are being exchanged, and that just
because you're exchanging a CDA document it
doesn't mean that that information is
structured and coded.

DR. LI: The CDA is coded, so it
contains both narrative and coded data entry.

MS. MARTINS: But it has to contain
also a coded, or can you exchange a CDA with
just the narrative?

DR. LI: No. No, according to the
MU specification you must exchange both, the
narrative plus coded data.

DR. BURSTIN: It just seems to me
we're talking about two different things. No
one is saying this replaces the feasibility
testing you would do in the course of normal
measure development. That I think would still
happen. I think what we want to have is at
least some centralized way to at least say of
these data elements, can we pull in all the
right information from the different
stakeholders to say to inform your process.

Don't even go down the, like don't
even bother testing a measure that's got a 20
on this because it's just a recipe for
failure. Go back to your committee in the
way, at least we've heard, you know, Yale does
it, and say, you know, this is so unlikely to
feasible, do you really need this data element
or could we live without it? So making it
more of an iterative process.

I see it as sort of a way to feed in
information to the development process. But
you're still going to test your measure. No
one else is going to test your measure for
you. But I think we at least need some
centralized source of where we think the state
of the art is.

MS. MARTINS: But how is that
different from feasibility testing? I
understand that's different from pilot testing
the measures and doing validity and
reliability testing, but can you get to a
point where a developer will just rely on this database and say, you know what, anything that's outside of here is out of scope and that's a developer decision whether, you know, how do you want to dwell in less feasible measures or not.

But I think the process of collaborating directly with vendors and providers cannot be understated, and maybe that's in addition. And again I'm not arguing against this, I'm just saying that it may not be it for feasibility.

DR. LIEBERMAN: I think one of the key parts about this would be that just if you see a score of 40 it doesn't necessarily need to be a showstopper. You should look at that, right, and then think about it, and say, well, is this something that we need to address through Meaningful Use Stage 4, or is this something, you know, that maybe I take exception with the --

(Off microphone comments)
DR. LIEBERMAN: -- exactly, you know, whatever's down the road is what I was trying to say. Don't want to scare anybody.

But you can also, I don't know, we can have some sort of method of reevaluating something, because it could be that it's scored once and then we're finding that actually it's not correct.

So again I think that we shouldn't focus so much on a specific score but look at this again as a way of identifying issues with a measure and bringing up at that exact discussion you were talking about between people.

And I guess one other thing that you mentioned which I think is useful. So we talked about data elements being kind of the key component, but you keep coming back to context. And so it could be that for a specific measure all the data elements look pretty good but only in a certain context. So you would want to identify that and have this
in that current document.

You made a point where with the sexual activity, are you sexually active. You say that's not a reliable field in yours, but in an OB-GYN office it's probably very reliable. So it could be that, you know, and you could put that sort of comment in this analysis so, you know, it's not a good measure for primary care but it might be a very good measure for somewhere else where you might have different, you know, this is the data field that's causing it to score low but we think that, whatever.

MS. MEADOWS: I was going to comment to the same thing that you were talking about that Rute brought up. I think this data element kind of repository of feasibility of data elements is a place to start, but I think you really need to understand per measure the overall measure intent before you can really, truly evaluate the complete feasibility of that measure.
And I was talking earlier to JD about a conversation last week along those same lines as far as the data element catalog that we received for Meaningful Use Stage 2. Very helpful. We've all been heads-down as vendors making sure that we could collect those data elements.

But once we got the measure specifications and understood some of the measure intent and the logic behind it we realized that it was only part of the whole picture, and there were a couple of measures that were very problematic.

For example, I was using the example of exclusive breast feeding. That always comes up. Because sure, we looked at the data elements, said, oh yes, we collect whether the mom's are breast feeding. But when you looked at how you had to collect it and what was expected, the granularity of the data, right, nobody does that in a newborn nursery. It's just not collected today.
MR. JENTZSCH: You can go back to what I was saying before. If this is not done on a per measure basis it would be really of no value to us. It has to be on a per measure basis, not just globally this data element, for the exact same reason everybody's saying. It would be absolutely no value to us if it's not on a per measure basis.

I mean it would be good to say in general this particular data element would have this score, that's okay if we're looking at it. But we're not measure developers. We don't think in those terms. We have to actually implement the measure and execute on it. If you were to create a measure for us and we don't have that, it would be of no value to us.

MS. MARTINS: And that's why I'm struggling with the concept of a central database as, you know, it may be a good starting point but it may be of very, very little value depending on the measure that
you're evaluating.

DR. SIMS: Doesn't that get to the audience? I mean, you know, if you've already built a measure and you're looking for a scoring methodology about how feasible it is, I mean at that point you've already probably wasted what, 12, 18, 24 months convening an expert panel.

I mean I think we do need a roll up, but I think you should be thinking about, measure developers I would suggest should be thinking about this as they're developing measures and thinking about, okay, here's the evidence available to us. Here are the guidelines. Here's what our experts feel is important and here's the gap, the opportunity to improve care, can we do it?

I mean if you're not doing it during that expert panel process then I don't think that this work is that great, honestly. I mean I think that's the opportunity and that's where it should be injected in the process.
Am I wrong? I don't know.

DEBBIE: We definitely, I think I mentioned if we do incorporate this feasibility in the development process, absolutely, before we've completely defined the numerator when we're just talking about the concept, is it a concept that should be developed into a measure? And we have all that discussion, but then we look at exactly what we want to capture at a high level and do this discussion.

So again I think the importance of classifying the feasibility testing on the de novo measures versus retooled measures.

DR. SIMS: And there might be some value in creating, I mean I think you do need to roll it up but there would be value in evaluating existing measures, so CMS and other providers can think about, or payors and providers could think about whether it makes sense to incorporate.

But I think the real value will be
moving forward and creating measures that take better advantage of the EMR. Because the existing measures are mostly developed in the claims world so we know that they're not going to score very well on a feasibility scale.

DR. OVERHAGE: I'm sitting here struggling. I guess maybe I'm getting too far in the weeds, but I'm just trying to think about the, I understand that there situations in which there may be measure-specific issues for developers, but I'm also struggling with the scale and scope of the work that a measure developer would have to undertake to do that. I'm just trying to imagine going out and saying, well, gee, we're thinking about these 120 data elements and we've got to go out and somehow assess the feasibility/cost of gathering these. How could you even do that?

So I guess I keep coming back, and not to say it's the be-all and end-all but it seems like the collective effort somehow has got to go into this because it's going to be
very hard.

MS. MARTINS: And I think it includes this framework. I'm not sure if it's readily available information. It doesn't mean that we can't crowdsource it for a specific, but for the specific questions that we ask may not have a readily available answer. That may need to be provided in the context in which they're being asked.

And taking the example of a paper based world, we typically go, I think, to five sites to do what we call alpha testing actually, which is equivalent, I guess, to feasibility testing. And we try to get the gamut of providers that are going, so it's kind of, it's a guesstimate. It's not validity and reliability.

DR. OVERHAGE: Sure, and I guess that's where I'm going. And so if you want to assess, and probably I guess I'm getting tangled up in the every provider implements, every EMR, you know, I mean it's probably to
your point, uniquely, going to five isn't
going to help. I mean it's going to be very
hard to get a reasonable sample. You're
probably talking about 40 or 50, especially in
the ambulatory setting, maybe hundreds to
begin to get a handle that means anything.

MS. MARTINS: I agree with you and I
struggle with, and this is something that I
actually brought up earlier. I don't know if
you were here yet or not, is what we're
bringing up to feasibility given the impact
that after you do feasibility you put specs
out there and you want to test them, and so
you're kind of at the point where vendors are
going to be incorporating these data elements
in their EHRs and there better be a good
reason for it.

So I agree with you that feasibility
is probably going to be more expensive and
more burdensome. And I do think that there
needs to be a stop in terms of, you know, we
can't go to 100 organizations or maybe not
But I would argue that on the paper based side, five is also not very meaningful in terms of feasibility, but again it's a qualitative assessment that I think we're, are we going in the right direction overall?

The fact that each EHR implementation is a single EHR implementation certainly brings a lot of complexity to this, and I doubt it that even with that kind of sample we would get at all the EHRs. The final measure is not going to get at all EHRs. Some are going to have to make changes in order to accommodate it because we can't evaluate the whole population.

But I think that that's a good point in terms of where, how many organizations do need to be involved in this so that feasibility testing means something in the context of a particular measure?

DR. OVERHAGE: Maybe that's a key step here though is figuring that out, you
know, what is the scale that you have to get
to to get a reasonable idea? And that's
obviously a studiable or answerable question.

MS. MARTINS: And that's also a
cost-benefit analysis from the developer's
point of view, as it is for you to implement
these --

DR. BREGMAN: Rute, I'd like to ask
you, this committee, I believe, we have to
come up with a recommendation or conclusion,
what do you think that should be? At the end
of the day here, what should we be
recommending?

MS. MARTINS: You mean in terms of
the framework, in terms of the population?

DR. BREGMAN: In terms of whatever
the goal is of this what we're trying to
accomplish. What recommendation should we
produce at the end of this process?

MS. MARTINS: What I would like to
see is a framework that we can use as measure
developers to get out on the field and
evaluate given, you know, a feasibility on our end on the number of providers or vendors that we could include. And let's not forget that this will probably be voluntary based, so there's that aspect there. We can only use the responses that get to us.

And then produce an overall assessment that would congregate these views of the reached out providers, and whether that takes into account a central database that already includes valuable information, I think it could. And then when we have the final scoring we can make decisions on, is it worth it to proceed with developing this measure right now? Do we develop it right now or can we develop it right now and say --

DR. BREGMAN: If the product of this committee is what, is the assessment, right, what do you think the assessment should be?

MS. MARTINS: I think this assessment makes sense. I mean we've discussed it extensively in terms of what
should be columns and rows and all of that.

I think what we are talking about, and that's
a good question in terms of whether this
committee is going to decide on that is how
this should be applied.

DR. BREGMAN: Right. Well, I think
we're just trying to come up with a tool.

MS. MARTINS: That's fine with me.

DR. BREGMAN: And then we can figure
out how to, where it can go. I was going to
say that I don't think we just agreed to go
with this tool. I was kind of following what
Shannon said, to simplify it. I would like to
see it in three categories, accuracy,
standard, codify or, I don't know the term,
structured is the word I'm looking for.
Accuracy, structured, workflow. Those are the
three areas that we're evaluating, and then we
have a scale.

And we can weight them but, you
know, I think that basically those weights are
going to be pretty comparable. And structure
is essentially a boolean, right, or maybe it's a variant because there's a distribution of whether it could be structured or not but it's, you know, sort of a boolean. And then those would be the three realms that we're going to score on.

DEBBIE: And can we ask the question, the same question, current and then future.

DR. BREGMAN: Right, and then it could be based on whatever time scale we're looking at.

DEBBIE: Right.

DR. OVERHAGE: So I may be missing it, but how in the framework do you get at how hard it is to get it there? I mean you may be able to get a highly accurate structured result there, but is it okay if it costs $2 million to that?

DR. BREGMAN: Well, that's workflow. Workflow includes the cost of, workflow is essentially, in real life what is the cost to
an organization to get the data.

DR. OVERHAGE: Okay. And on accuracy, so you look at what's out, blood pressure is a great example. You look at what's out there in blood pressures in the EHR structured data fields today and the accuracy is horrid.

If your standard is what was the patient's blood pressure when they walked into be assessed, it's horrible across the board. So what do you do with that as a measure developer so that the, you know, the accuracy of that is ten percent?

DR. BREGMAN: Well, my response to that would be everything's relative, one. And secondly, I think that's really a question for the experts, the measure, the experts that are proposing the measures, they are the ones who know in general what is the accuracy of blood pressure taken at --

DR. OVERHAGE: But that's the point. It's not the accuracy of the blood pressure,
it's the accuracy of how it gets recorded in 
the system in real life in the workflow as a 
structured data element. And that's what 
we're hearing that the measure developers 
don't know.

DR. BREGMAN: Well, if that's the 
case then we would score it low based on our 
knowledge of that.

DR. OVERHAGE: Well, I guess where I 
was going is, so we look at that, so we do 
that, does that mean you, what do you do with 
the measure for which the data element, and I 
guess I'm just trying to understand how that 
gets --

DR. BREGMAN: What do we do or what 
does the measure developer do with our 
evaluation?

DR. BURSTIN: I think part of this 
gets to the timing issue of where we are in 
the process. So my sense of this is, my hope 
would be this whatever we want to call it, 
tool, et cetera, would have these data
crowdsource-collected across many EHRs, many implementers, have a central source, so that as you're sitting down with your TEP, you're sitting down with your workgroup, in the words of the other developers, you've got some basis on which to guide how feasible this measure will likely be.

Does that replace what you're talking about, Rute? No, of course you're still going to do your alpha and beta testing, whatever you call it. But at the beginning of the day, particularly for the sake of CMS or ONC or those that are funding it, they want to know that what they're actually putting forward there is going to likely result in a measure that's doable today or in the future.

So it still seems to me like it would be really useful. If I was a developer I would really like to be able to sit with my committee and say, I know you really think it's important to have that data element. Do you have any idea how hard it is and how
unlikely that's going to be in the next five
years? How much do you really need it?
Should we move to the next?

I mean I think this is what they're
trying to do is, how do you build this into
the development process so it's not something
that you're doing at the end of the day after
you've already got it built, but it's
literally helping you build your measures?

So I'd love to hear from the
developers. I mean that's, again I haven't
done that in years.

MS. CHRISTENSEN: Hey, it worked
that time. I agree. I think that would be
really useful. We definitely get a lot of
pushback from folks who are passionate about
a particular measure. And it's not a bad
measure, per se, it's just, you know, it's not
going to happen in the next five years so why
bother to spend your time on it? Now the
evidence could change in five years.

But the concern I have, and I know
I've said it before so I will take one of your lines and I won't say it again after this, I'm not sure that we can get good people to do this right now when we are offering to pay them.

So crowdsourcing, though a great idea, there's got to be an incentive. And I think if we walk out of here without understanding what the incentive is besides just we all love this stuff and want to do it, we're going to get a crowdsource with no information.

DR. BURSTIN: I guess I'm thinking more along the lines of, again I don't know if they would even fund this, but if there's a way to get, you know, the group of informaticists and informaticians, my AMIA friends tell me on the board, I'm sorry, informaticians and vendors and end users together, and it may just be 20 people, I'm not convinced it's that difficult to do this on an annual basis the way we're updating the
QDM.

You then, as part of this then you have an update of the data element feasibility scores that this get provided to all to use.

MS. CHRISTENSEN: So not really crowdsourcing then --

(Simultaneous speaking)

DR. BURSTIN: I didn't mean to be quite that ad hoc, yes. Sorry.

DR. LIEBERMAN: So would a system with three categories and three scores to each category create enough information that we can do what you're talking about doing? Would a score, so that would be, what a nine-point system? So if you see a measure, a data element that's less than seven is that a red flag and then you go in and look at it or what would the --

DR. SIMS: I think we have to develop that empirically. I think we have to look through a bunch of data elements and see how they're applied in existing measures, but
I think that's a rational way to proceed.

I would say as you form that 20-person committee, and I know that's aspirational but it would be heavy on vendors and providers. That clearly all stakeholders are invited and should participate, but that's where the rubber meets the road on feasibility.

And they your vendor products, you've got to get as many different vendors obviously.

MR. KRAVITZ: One thing that that would force, which I think would probably be a good thing, is that when the TEPs are working they would have to thinking in QDM.

So you have to get out of thinking in terms of English, you know, the doctor's assessment to the patient's mental state or something, and they would have to actually boil down the measures to, I'm looking for a depression assessment tool in order to do the assessments. And that would really push the
thinking way back up into the TEPs.

DEBBIE: And just one other point besides my plea for definitions in the document is, I think it's important that we clearly state that this is a tool to be used at the appropriate steps in the process of measure development.

Meaning that here's your first line of information with a score, we mean for you to discuss this further to reach out for a public comment to use it in the whole measure development cycle testing cycle, whatever cycles.

This is sort of, you're now at first base. You have some evaluation, and before you move forward consider these points. So I think that's important so that we don't have anybody say, oh, this has a score of six, we can't use it. We can't continue to discussion, or throw it out the door, or, you know, even at the lower end of the scale, this is meant for consideration, points to consider
and discuss.

DR. BURSTIN: The value of the element still has to be considered, and value of the element is nowhere here. That's the other half of it.

DR. BUTT: So I'm trying to sort of operationalize this. So we have three categories and then within each category we have a one, two, three scale, right? And so will each, all those scale items be clearly defined what they actually mean? And then will all of these be equally weighted or will there be a weighting attached?

DR. SIMS: I would say no weighting to keep it simple, personally.

DR. BUTT: Because if we keep the weighting equal I can see the structured data/coded standards as in a relative scale, easier of the three categories to get your arms around in terms of what you come with, workflow perhaps less so. But the accuracy, I'm still not sure I understand how will we
determine the accuracy of data in this context.

        DR. WINKLER: Might it help sort it out if you could offer examples of what one, two and three might look like for accuracy?

        DR. SIMS: I mean accuracy could vary at, you know, in my institution across, it could vary from provider to provider in the same clinic. I mean, you know, is one using a PA and one using an MA? So all of this is highly subjective.

        But I think, you know, it allows us to put some guardrails on the road to keep the car on the road as opposed to, you know, we're not the steering wheel, we're not the speedometer, we're just trying to keep the car going in the right direction in terms of measure development. So, you know, I think we can live with that subjectivity. I would postulate.

        DR. LIEBERMAN: I'd make one recommendation and maybe we can get a show of
hands or something. I would say that, I still think workflow should be weighted more heavily than the others.

And maybe it doesn't make it quite at simple, but when people talk about feasibility it seems to me that they're asking about how hard is it to get this information and how hard is really workflow, so I would suggest maybe doubling it and having everything else single-weighted or something of that nature.

DR. BREGMAN: I would second that. I also, you're probably going to shoot this one down as a group, but I would vote for a five-point scale just because you want to reflect a little bit of nuance and the distribution among, you know, there's so many factors that it would be nice to be able to say, you know, four is yes but five is really yes.

And it doesn't have to be we have to define all five points. You can just say one,
three and five have a definition and four is halfway between three and five. And for the record, Marc was nodding when I said five-point scale.

(Laughter)

DR. LIEBERMAN: Can we see a show of hands? Who agrees with doubling workflow?

(Off microphone comments)

DR. RADFORD: I'm not sure that we need to obsess about the weights of the different domains, because I think the different domains are going to be treated slightly differently by the measure developers.

So for example, if there's a workflow issue that's rated low, then depending on what it's all about someone may want to work on that and say, okay, we're going to work on what the workflow is so that it, you know, it's going to be easier. I think that's the point. And the other domain similarly. So I'm not sure we really need to
obsess about a weight.

DR. TINOCO: I agree. And as long as we have the option of seeing the individual subcriteria and the result of those and a current state versus the level we have to get to the future state, that's meaningful for me as opposed to rolling everything up into a single composite value that hides the nuances of each subcriterion.

DR. LIEBERMAN: So I would think that eventually we would, I mean we're going to have how many data elements in a measure? And does somebody want to throw something out? I don't know, is it --

(Off microphone comments)

DR. LIEBERMAN: -- lots and lots, right. So you can come up with an, maybe when you're looking at an overall measure you're going to do something like look at the overall score, I mean look at the average for all of the data elements and maybe you're looking at the outliers, you know, which ones are high,
which ones are low.

And I think what you're going to want to do is, I think you want the ones that have difficult workflows to stand out more than the ones that have difficult data standards. I mean that's my assumption, and again --

DR. SIMS: Well, I think you have to be careful because one bad data element can ruin, you know, one bad egg screws the whole thing up. So I think at the very least we need to publish the minimum score of the data elements, and then from there we can mean maybe a mode, whatever, that's fine. But the minimum value is an important factor, I think.

DR. LIEBERMAN: Right. But if you score perfect on accuracy and data standards and poorly on workflow it might not stand out is what I'm afraid of.

DR. BREGMAN: Just as a reality check, which measure has 200 data on it?

DEBBIE: You know, I was thinking of
the whole feasibility report that we had on 200 data elements.

DR. BREGMAN: Because even the SCIP measures have, I don't know, 30 or something like that.

(Off microphone comments)

MS. CHRISTENSEN: I think something just to throw out there, and it's probably obvious but sometimes I like to just say them out loud. Some data elements matter more than other data elements, so I think that that's something that as measure developers we have to take into account. If it's something that's essential to my denominator, you know, a two out of three may not be enough. That might be a no-go if I can't get the denominator.

But if it's something that's going to affect one percent of patients some of the time about why they're not going to meet a measure, maybe one out of three isn't that big of a deal.
DR. BURSTIN: And maybe you just won't include that data element in your measure.

MS. CHRISTENSEN: Yes, exactly.

DR. BURSTIN: Strip it down to what's really needed.

DR. TINOCO: So I hope this is, I love building tools. This is fun. But I sure hope that during this process we actually exercise the tool that we build and challenge ourselves to use the information that we generate, and make the decisions as if we were either a sponsor or payor that's going to implement this thing.

DR. BURSTIN: Actually in HITEP-I, since I may be only one of the people who was here for HITEP-I, we actually went through and we rated -- well, you were there too, Marc, weren't you? We went through and rated, what, 40 measures or something as a group.

So I think some of this might be, if we can come up with the anchors around these
one to five and some definitions here, we'll
go ahead and we'll just pull up some measures
and send it along. I think I heard some
support for five.

DR. LIEBERMAN: Maybe we should have
a quick vote on three versus five. Who's in
favor of five, a scale of one to five?

DR. TINOCO: I mean it will be on us
to actually define what each one means.

DR. LIEBERMAN: Howard, your
proposal for one to five got voted down while
you were out. I'm actually going to have to
go pretty soon, but some of the things that we
can do is work on the definitions of putting
some words around one, two and three for each
of three categories. I think we have, we've
got three categories there. We probably will
take a break in a few minutes. So that's one
thing to do.

The other thing is, you know we
talked briefly. Do we want to explore anymore
about how we want to aggregate the information
that we're collecting on all these various
data elements or is that something that can be
saved as later as well?

So we can look at kind of things
like average scores for the overall measure.
We can look at, you know, high score or low
score. What comes to me is, when we talk
about cost this will identify individual data
elements, but there is something to be said
for if you're, you know, if you have a lot of
medium cost data elements maybe it's not so
good either.

And so with the scoring system we've
put in place that might be hard information to
find. I'm not sure, because you might want to
invert the one through three, so one is good
and three is bad, and then you can add up all
of the workflow scores or something to get an
overall idea. But I don't know. I mean do we
want to spend some time on that or not?

DR. SIMS: I have some quick
thoughts. I mean I think as Keri alluded to,
if it's a denominator or a numerator element
then that's going to take precedence. It's
going to have a higher weight than if it's an
exclusion criteria which are more rarely used
and which, frankly, if you're using a
threshold based approach to measurement
accountability may not matter in the end
anyway.

So if we're going to go down that
road that would be the road that I would
propose. And I don't know that you can
discriminate with denominator and numerator
being more important. They're both pretty
important. But exclusions and exceptions you
might be able to navigate around a little bit.

DR. LIEBERMAN: Reva, were you going
to say something?

DR. WINKLER: Just something some of
you were talking about earlier, the idea that
a score for each data element somehow
aggregated and make some assessment about what
the aggregate looks like so that's moving
towards the score.

What other characteristics of feasibility of the score beyond the individual data elements are important? The actual measure should be part of a feasibility assessment that would be useful for developers, and then anyone else along the line.

DR. BUTT: So I think maybe not exactly answering your question, but I think that one of the things in addition to maybe the average score would be important to have the frequency of a certain category.

Because, you know, I'm still concerned that we'll get a lot of threes in accuracy because it's going to be so subjective that, you know, as was being mentioned that accuracy is so sometimes even provider dependent that how do you sort of generalize it in a scale that you're trying to put at this level with 20 people sitting around the room and trying to say whether this
element is going to be accurate or not? I'm just concerned.

So if the frequency of these elements through different either surveys or expert panels is showing very high or low then we need to revisit and see whether it's truly getting us what we need.

DR. LIEBERMAN: Shannon?

DR. SIMS: I think, Keri, didn't we joke about this on one call? That the, I might have put it out as, putatively, the Sims scale that the more data elements there are in a measure the less likely I am to implement it or the less likely.

So, you know, we might look at some metadata like the number of data elements. That is correlated with how many institutions like Mike and myself and other providers will do because it makes a huge build effort. I mean we can spend several FTEs for a month or two building one metric. I'm not kidding. So the sexual activity one, for example, took us
forever and actually we just abandoned it, to be frank.

So there might be those kinds of opportunities there, and the logic itself. So that probably the more data elements there are is not an independent variable from the complexity of the logic, but the complexity of the logic also, as a provider organization, not so much the computational aspect but our ability to explain it to our providers.

So this is why your population is only half of what you expect. It's because they're excluded for XYZ reasons is another factor, and feasibility, in a sense how likely we are to both implement it and then to use it for substantive quality improvement.

DR. OVERHAGE: That's one of those that, to me, falls into the category of, in God we trust and all others bring data.

I mean I think that's one that you've got to measure because the assumption about what it will be, you know, and so maybe
people like Kaiser or somebody may have 99, but I think we have some data on that when you're really going to be in trouble. Because it's so easy to assume that that obvious thing would, and gender is actually a great example. At Partners Healthcare we were doing a research study, and not to beat up on Partners because I don't think they're any worse than anybody else, 45 percent of the people didn't have race, okay. And something simple and obvious, not there, so I think you have to be really careful about the accuracy one.

MR. KRAVITZ: I just want to come down to the breakdown by populations that I think Shannon mentioned. I think it's really important to report the roll-ups by denominator, numerator exceptions and exclusions. And I think the number of data elements is an important measure to report both from the ability to explain it but also
the testing. If you look at the testability
of some of the measures that have a zillion
data inputs, it's clear that no one has ever
tested them computationally because to
actually test them in a software perspective
would be prohibitively expensive.

So we also have, just as an aside,
we have a research project to try and look at
the QDM expressed measures and computed
measure of complexity, your algorithm and
complexity of them to report that since that
also makes it harder to explain.

DR. BUTT: Another thing might be
for us to consider the average score
interpretation. So perhaps, you know, from
here to here it means this, from here to here
it means this. Because then the measure
developers actually can use that sort of to
determine that for a denominator that is a
critical data element they would only use a
score above a certain number versus perhaps
some exclusions they could select a slightly
lower score.

So we don't have to get into that part of it how they use it, because that will be up to them to determine how important it is in their measure to use a certain data element based on the score.

DR. LIEBERMAN: And then you can also see that if you start collecting this data you can start doing a lot of different analysis on it, so we could look and see which measures people are choosing to report for meaningful use, and then we could look at average scores for data elements on those measures and you can compare your new measure against those or, you know, it would open up a world of possibilities and lots of good work to be done.

Aldo?

DR. TINOCO: Forgive me, I'm going to try and sneak in four thoughts. With regards to Reva's question about the score of the measure itself, it's kind of hard to
disentangle that from the logic. I mean these
are computers. This is software.

And if we can generate values for
each data element then it's just a matter of
programming in what type of calculation we
want, even if it's a complex weighted
composite that we have planned for, you know,
future iterations of meaningful use. So
that's one thing about, I can't really answer
your question directly at this time.

Secondly, Shannon, the quantity of
data elements being rolled into this, should
I implement, should I not or the cost of
implementation, let's think about whether or
not those data elements already exist in our
systems -- I'm sorry, are already used for
existing quality measures.

So if I assemble components from
different Meaningful Use Stage 2 measures and
they're already out there in deploy and
successful, that's going to be like ten, for
argument sake, but they're all existing so I
wonder whether or not we can make sure that we know that there's a different level of effort for brand-new, brand-new versus reused data elements.

Third, we haven't really talked about data elements that are derived from other data elements. So there's some special cases here like we're talking about change over time measures and that is subtraction in many cases of two existing numeric values, but we have to keep those special cases in mind.

And lastly, we haven't really said much about attributes of data elements. So as measure developers using the QDM, we have this ability to say oh, I want a patient reported outcome result from use of a tool and I want to assert that the source information is the patient not the provider.

So as we flesh out these different rules of engagement with feasibility it's not just, sometimes a measure concept or component is not just a single data element, it's a post
coordinated collection of separate data elements.

DR. LIEBERMAN: Along those lines of this issue of using data elements that have been used in measures before, actually I think that's a really interesting idea. And it doesn't get at this issue of building for the future but it does give you, should we use that in our scoring system?

So if we've been collecting a data element, it's been used, it's been reported, the overall measure has been found to be accurate and useful, should you get more credit for choosing that data element than one that scores well but is untested?

DR. OVERHAGE: So I would suggest that that's independent. I think it's a huge value, right, that's our aspiration here is that I record the blood pressure once, use it for 900 measures and for patient care and for research, actually nine measures. So that's aspiration.
But to me that comes in sort of at the back end of the thinking. So A, presumably that will be a pretty accurate if it's been used it will score well as a measure. And then there is some level of, and as you said it's sort of a mixture of, yes, that's been used before. That's something that signals providers and EHR developers this is something that gets used a lot.

But I can also imagine sort of the wishlist need as you said for the future. Well, I don't see the measure there. It's going to be a tough one to collect, but let me at least tick the box and say to the world that this is something that we thought about and would like to have if we could, even though it scored too low that we couldn't, we'd like to have it, and that that's what helps drive the future development.

So capturing the fact that a measure developer thought about an element and abandoned it or whatever would be very
valuable, I think.

   DR. SIMS: I would agree with that.

There are a lot of data elements in use right now that aren't very feasible, so I think it needs to be an independent evaluation.

   MS. MARTINS: Just a quick comment based on Aldo's comment on what is a concept that we would feed here in what we're calling data elements.

   I think everything that whether that's a QDM element or an attribute of QDM element that has a value set attached to it would qualify as an individual concept to be evaluated as a unit of meaning, so a single atomic data element, if you will.

   And then there may be others that don't actually have value sets attached to it, so admission date, for instance, those are attributes of an encounter within the QDM. So again, to me, a QDM element is the highest level of specificity defined within it.

   DR. SIMS: Can we do it empirically?
I mean can we look at the measure authoring tool, pull out sort of most frequently used state, the combinations of QDM designations and sort of start with those?

Is that a rational starting point for the group assuming that that might be the most frequently used types of data elements moving forward? Because this is going to be a long list of potential data elements and that might help us as we work through the rubric that makes sure that we're identifying the needs of the measure development community.

MR. KRAVITZ: I would suggest we start with the data elements that are used in the Meaningful Use 2 measures as a sanity check.

DR. SIMS: But I mean the cat's kind of out of the bag on those, right, I mean those are writ. I mean I think --

MR. KRAVITZ: But we don't know whether they're feasible.
DR. SIMS: That's true. That's a good point.

DEBBIE: Right. They're not proven yet.

DR. SIMS: Okay, but then all right, so that's a good point.

DR. KENNEDY: But it may be a good data set to start with to assess feasibility.

DR. BURSTIN: I just wanted to go back to the, I'm the measurement geek, so I wanted to go back to the issue of data element versus measure, because I think we sort of glossed over it a tiny bit. So I guess one question is, if there's this level of rigor on the data element side, does anybody feel like you need to really do additional feasibility analyses for the overall measure?

Is that something we would require, or are we feeling pretty good that if you have the data elements and you're putting those forward and you think they're fairly reliable -- let me just finish the last part of this --
you think the rest of the concerns would likely come up because they are still going to have to be tested for reliability and validity.

So it's hard to imagine that a measure will turn out to be valid if it's not feasible. So I just wanted to try to disentangle that for a moment with you.

MS. CHRISTENSEN: I'm going to go back to Aldo's point about the data around the data element, because a lot of times that's where we're seeing people run into trouble, if they're not capturing all the information about the actual piece of information then you can't compare one piece of information to the next one or do the linkages between them that you need. So --

DR. BURSTIN: And how do you assess that? Is that the complexity of the logic? Is that the completeness of the logic? What is it?

MS. CHRISTENSEN: I don't know. I
mean I don't want to say that each of those
pieces of information that you gave is
actually a data element because it's not in
the sense that it's not a new data concept
it's a related bunch of data concepts, but it
almost is.

There is almost a feasibility for
each of those individual pieces of it and then
how they link together, and that's hard to get
at. Yes, it's the measure. Short answer.

DEBBIE: I would agree with Keri,
and I think although I know vendors can
program anything logic-wise, I think there's
a complexity related to implementation and
understanding, understanding to the providers
and sort of a whole picture there with the
complexity of the context and related to the
logic. So I think we should call out some
things to consider and for discussion and
that.

MS. MARTINS: And I would agree that
there is a dimension that is about the measure
that exceeds the data element by data element
evaluation. I think that data element by data
element evaluation -- and again to me I've
already explained what that means to me,
whether that's the pieces around a specific
measure or concept, I think it is -- I think
the aspects of eMeasure representation, so the
challenges of eMeasure representation which
are not necessarily here.

So for instance, a limitation in the
QDM. Whether we need to do that data element
by data element, I don't know, but we
definitely want to have some measure of those
limitations as well in order to create an
overall score of how we feel that the measure
is going to fare. I would say though that I
think it will be really hard to come up with
a scale for that.

DR. RADFORD: Yes, so I would agree
that we should at least say something about
the measure itself beyond the individual data
elements. I mean as you know I couldn't make
the phone call because of Sandy, and I had a remedial phone call after that. And I was very worried about this, actually, because, you know, we've had two EHR vendors. We just dumped one and had another, and both of them are telling us, gee, that's a great idea about measure construction and we'll get back to you on that, and they never do.

So, you know, this issue of how you take the data elements and aggregate them and put the logic behind them and, you know, give back either, in my view, decision support or a post care measure, it is a very real one and I think we need to acknowledge it.

DR. LI: Just one comment on the data element assessment. So we can suggest that we do the overall but to put more focus on the measure specific, that element assessment, because that basically fall into the scenario that Keri just mentioned.

For some complete, complex eMeasure you have all the data element defined
available but it's still hard to perform measure calculation. Why? Because overall these data elements are available but fall into that measure required a specific context. You may find out that some attributes required of that measure are not available for these data elements.

Another finding, we do the MU2 measure, the development and the testing, it's that sometimes for the very large, complex measure the logic, what's the most efficient way to represent a measure logic, make it easy calculated, easy understandable could be a issue, but that's kind of a separate issue from the data element assessment.

So there's always room to improve measure logic representation itself, but doing validity testing the logic compare reason on the later on phases.

DR. BUTT: So I think one of the things might be important to, you know, in this context again define what we mean by a
data element. Because, you know, the quality
data element has a certain definition and
often the granularity is different when you
come in at a data type level versus when you
put the attribute next to it. And even in
EHRs the data sometimes at the element level
is stored almost at the attribute level.

So the EHRs generally don't follow
the QDM model, if you will, of data, so I
think it might be very important to define
exactly what we mean by data element. Does it
sort of include the attribute itself, or how
do we define so we can bridge this gap if part
of the analysis will be based on QDM, and then
we're dealing with the real actual data.

MS. MARTINS: I think I've proposed
a definition for that, I think, from the
perspective of applying this framework to a
concept. Particularly because EHRs do not
necessarily follow the QDM model, it may not
make a lot of sense to evaluate an encounter
as a data element, but the admission date has
meaning within an EHR or doesn't. I mean this is part of why we're doing the assessment. But I think it goes all the way to the highest level of granularity with which the QDM element is defined, including any attributes and value sets associated with those attributes, and I would say that that's the most granular that you'll ever get. And if it's just an attribute without a value set then that's another level. That's another concept that we want to test against this framework that the element itself with no attributes and a value set that's another concept. All of these are unitary, atomic concepts that we need to test against this and this is my proposal.

DR. LIEBERMAN: That sounds like a lot more data elements though. I mean if we're talking about doing this as a, so I can see doing that individual measure to thinking it through that way, but if we're thinking about a repository of data elements that have
been evaluated and are available for kind of review I think that's a lot more difficult.

MS. MARTINS: I agree, but I would say that again because of this disconnect of how, and it's not necessarily a bad one, the QDM is an information model so it surrounds, it's about meaning. It's about the meaning of the information.

And the problem is that a QDM data type might not give you the meaning that you're looking for in a measure. So a QDM data type feasibility assessment may not mean anything as the concept you're looking for.

So for instance, an inpatient episode of care. That may have meaning for the measure, but if you are selecting patients based on the admission date and the admission date is an attribute and isn't in that pool of data elements or feasibility or concepts then you won't be able to do anything.

So I guess what I'm saying is that the most important concepts for a measure
might not necessarily be QDM data types
coupled with a value set.

MR. KRAVITZ: Can I ask just a
clarification question? So you're saying, is
your proposal that if I'm proposing a data
element, let's say inpatient encounter, so any
attributes of that data type that are required
by the measure should also be scored?

MS. MARTINS: Yes.

MR. KRAVITZ: Okay.

MS. MARTINS: Because again, the
importance of the concept and the measure does
not necessarily tie to just the QDM data type.

MR. KRAVITZ: The type as opposed to
the attributes. I agree with that. Because
a lot of the trouble we had during the rework
of the Stage 2 measures had to do with
attributes not the types.

DR. WINKLER: Marc's looking at me
because it looks like he's packing up his
stuff and getting ready to go. And we are
getting to the close of our agenda.
I think a couple things we need to decide is what we need to be able to move into starting to put this together into a document that makes some sense and be able to tell our story in a rational fashion.

I think it sounds like everybody's reasonably comfortable with the three-part framework that we have here using a three-part scale as certainly a starting point without any weights at least initially.

The thing I think that's missing that's critical is going to be what does a one, two and three mean under each of those categories? We don't have time to really do that now, but I would really like to have some volunteers propose some of those, if not I'll assign them. Because I really need your input to help us really understand what you're thinking about.

So what is a one, two and three for accuracy? You know, Aldo did a pretty good job of maybe representing what a one, two and
three under data standards might be, but maybe
it needs to be a little bit more robust or
fleshed out a little bit more, and similarly
for workflow fit.

So I'm right now soliciting, you
know, volunteers to try and, you know, pick
one and try and do it for one. It would be
great if everybody would kind of pick one.

(Off microphone comments)

DR. WINKLER: Well, it sounds like
the accuracy one is the one that people are
feeling the least comfort with, so who are the
real brave volunteers on that one? Go for it,
Marc.

Okay, that sounds great. If you
guys --

(Off microphone comments)

DR. WINKLER: And data standards?

Zahid was down there on data standards.

Martha, JD. Okay, that sounds great. You
know, that's really the sort of the missing
blank spots right now on this framework.
Essentially, I mean through the course you've made some recommendations that you may not have realized that that, but we can formulate them in terms of how this should be used, could be used, might be used, whatever terminology we ultimately agree with that we'll try and formulate.

But as after the last call, for a couple days there was this nice little email thing, exchange of ideas.

And I would really like to encourage you to think about some of these things and continue those sorts of discussions, because some good thoughts, especially as you ruminate today's discussion, you know, over the weekend or whatever else you may be doing the rest of the next couple days might give you some way of formulating it that may help communicate. Because that's the essence of the challenge for us is communicating all these ideas on a two-dimensional piece of paper, and so your help.
Also we started looking at the principles and guidance that we drafted, and we see that as an important part of the end report. If you haven't had a chance to look at those I'd really appreciate if you would, and then any feedback, certainly, if you think they're wrong, need to know.

If you think they are not quite on target and can be better formulated to get the concept across, need to know. If you think it's irrelevant or something like that, those sorts of things would be really, really helpful.

The next couple of weeks we're going to be trying to pull these pieces together. We do not envision this to be a lengthy tome. We really do see this as a relatively concise, focused document, but nonetheless it's a document about what you all think.

And so we want to be sure we've got your thinking accurately, and so we really want to continue getting your feedback and
getting your input in terms of how we're going
to put this together.

So essentially the document's going
to look like, you know, there'll be an intro
section. There'll be a section on guidance
and principles, section on recommendations,
section on the framework, and a section on
things that should happen afterwards. You
know, the parking lot issues, the
recommendations that come from of actions and
activities that should follow from these sets
of recommendations going forward.

We do have a very short timeline to
accomplish all this so I don't think we can do
much more than that, but actually I think
that's a fairly significant bit of work to
accomplish.

Marc, you look puzzled. Do you have
a question? Okay, just checking.

DR. SIMS: Can I request that these
groups give an example, just one example for
each of their categorical descriptions?
DR. WINKLER: I'd echo that. Because I think that again in explaining and describing the framework those examples will be very helpful in audiences getting it. So thank you, excellent suggestion.

Any other of those brilliant suggestions, please?

DR. RADFORD: If you could just send an email with this homework assignment in it, whatever you just said, so that we can kind of do that. Thanks.

DR. WINKLER: We can do that.

Debbie?

DEBBIE: I have a question. How will this feasibility testing be used as part of future measure endorsement or maintenance review?

DR. WINKLER: Well, I think that's one of the things we're hoping to be able to pull out of this. You know, we've kind of run out of a time to really talk about directly, but as you know feasibility is one of NQF's
evaluation criteria. And so just as we did
under scientific acceptability with
reliability and validity, they called out
specific applications of that for EHRs.

So I think what we envision is at
some point being able to call out things that
might be specific to EHRs pertaining to
feasibility, but I don't think we're there
yet.

DR. BURSTIN: For those of you that
have been around for awhile, after HITEP-I we
went ahead and put these criteria in
feasibility very prematurely. This seemed
really cool. This is feasibility, we put that
in there. It was premature.

So I think ultimately it would be
logical that these would reside under our
feasibility criteria. Did you assess? Was
this assessed? Can that be provided? But I
think that's something we'll continue to work
through with you as we get some definitions
and meat on the bones.
MS. MARTINS: Can I just make a comment on that? I think it has to do with the source or the information source for a measure. If we're talking about an eMeasure versus a paper based measure they're not the same measure.

They may have the same intent. They have different levels of feasibility, of validity and reliability, and I don't think they're, there's no understating this. You may have a very feasible paper based measure that is very reliable, and as you retool it, it suddenly becomes something else. It's a different animal and it needs to be treated as such.

DR. WINKLER: I think another thing Helen has brought up is feeling confined to respecify an existing measure on a real identical basis rather than pulling back and saying, what was that measure all about? What was it trying to measure? And then say, okay, how would we measure it in an EHR, not being
confined to how it was being done in the paper world because the worlds are very different.

I think that's another thing that we want to begin to encourage, and I think it speaks to the fact that they're not the same measure. They may be measuring the same idea but they are not going to be the same measure.

MS. MARTINS: Yes.

DR. BUTT: I think that also, maybe this is not the time for that discussion, but it also begs the question that from the NQF standpoint is every tool measure considered the same as the original endorsed measure, because currently it gets the same number. And I know the joint commission internally has already moved to assigning a different letter to distinguish it from the paper based measure, and the question there is that does it then have to go through endorsement again because it's a different number? At least internally it's a different number, but I think there are lots of angles to that
question in terms of going forward if we're not going to treat them the exact same.

DR. BURSTIN: This is part of what kicked off this whole project in the first place is we posted our guidance for what was required for eMeasure testing and very explicitly said what testing was required when you're back for maintenance the first time. And that's when people said, well, what about feasibility? So we're going to have to as we get through this process revisit that and figure out exactly what that means, because it was very clear.

And I'll tell you, the majority of those comments came from the provider side who said, I can't use these measures unless anybody's shown they're actually feasible. And especially if CMS has any desire to use them for accountability it's going to be really important we demonstrate that they work.

MS. MARTINS: And feasible is not
enough.

(Simultaneous speaking)

MS. MARTINS: -- measures are valid.

DR. WINKLER: One last thing we need
to do is be sure that, or at least ask any of
our audience members here if they'd like to
offer any comments.

And Operator, would you ask if
anyone on the line has any comments or
questions to offer?

OPERATOR: As a reminder, if you
would like to ask a question or make a
comment, please press star 1 on your telephone
keypad at this time.

MS. CRAWFORD: So just a quick
comment. I think it might be helpful, if
there's a technical capacity to do so after
these initial assessments of feasibility for
different categories are available, to give
measure developers an opportunity to provide
their results from actual feasibility testing
into a nice database or some sort of resource
or clearinghouse where people can see for
different settings, for different specialties
what are the actual feasibility results,
because I think that kind of information is
going to be crucial down the line.

DR. WINKLER: Last thoughts from the
audience?

Well, from everybody here at NQF,
thank you very, very much for helping us out
today. This is just the beginning. We're
going to be working together over the holidays
and into the new year as we try and pull all
this together.

We'll be reaching out to you all at
various points along the line as we run up
against questions or not quite sure which way
to go as we're drafting the report and trying
to figure out where we need to characterize
what it is you all really want to convey and
communicate with everyone.

But I think it's been a very
valuable discussion today. I can envision at
least, you know, how I'm going to start with it. I don't know at what point I'm going to run into major roadblocks and yell help, but I won't hesitate to do so.

So I've really enjoyed meeting all of you in person, and look forward to working with you as we go forward.

Helen, any last words from you before --

DR. BURSTIN: No. Thank you, thank you.

DR. WINKLER: Please travel safely. Take care.

(Whereupon, the foregoing matter went off the record at 3:45 p.m.)
<p>| Page 372 | Neal R. Gross &amp; Co., Inc. | 202-234-4433 |</p>
<table>
<thead>
<tr>
<th>Customizable</th>
<th>229:8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutoff</td>
<td>33:7</td>
</tr>
<tr>
<td>CV</td>
<td>5:5</td>
</tr>
<tr>
<td>Cycle</td>
<td>81:16 321:12 321:12</td>
</tr>
<tr>
<td>Cycles</td>
<td>321:13</td>
</tr>
<tr>
<td>C-O-N-T-E-N-T-S</td>
<td>3:1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D</th>
<th>47:16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>54:14 123:11</td>
</tr>
<tr>
<td>Dance</td>
<td>206:19 208:1</td>
</tr>
<tr>
<td>Dangerous</td>
<td>204:22</td>
</tr>
<tr>
<td>Dark</td>
<td>142:3</td>
</tr>
<tr>
<td>Darn</td>
<td>252:9</td>
</tr>
<tr>
<td>Dashboard</td>
<td>182:2</td>
</tr>
<tr>
<td>Data's</td>
<td>3:1</td>
</tr>
<tr>
<td>Data/Coded</td>
<td>322:18</td>
</tr>
<tr>
<td>Date</td>
<td>32:14 33:19 343:18 351:22 353:17,18</td>
</tr>
<tr>
<td>Dated</td>
<td>88:20</td>
</tr>
<tr>
<td>Dates</td>
<td>61:4 177:7,12 177:14</td>
</tr>
<tr>
<td>Dave</td>
<td>159:22</td>
</tr>
<tr>
<td>Days</td>
<td>357:9,17 29:7,21,22 83:5 106:5 108:18 289:3 305:13</td>
</tr>
<tr>
<td>Dead</td>
<td>240:5</td>
</tr>
<tr>
<td>Deaf</td>
<td>78:1</td>
</tr>
<tr>
<td>Deal</td>
<td>31:13 94:14 134:1,2 328:22</td>
</tr>
<tr>
<td>Dealing</td>
<td>31:13 351:15</td>
</tr>
<tr>
<td>Deal</td>
<td>152:17 231:9</td>
</tr>
<tr>
<td>Debate</td>
<td>248:14</td>
</tr>
</tbody>
</table>
low 105:12,12
160:11,14 175:13
175:18 176:2
178:13 181:22
189:13 222:18
248:21 250:2
252:9 254:16
265:16 269:1
274:14 278:15
280:2,5 301:12
315:7 325:16
327:1 331:6 334:5
342:17
lower 94:9 257:22
321:21 338:1
lowered 164:18
lowering 165:22
lowers 113:3
lowest 94:3 143:4
lumping 204:3
205:21
lunch 172:9 213:10
214:6 215:4
L-score 295:7
static 43:20
statically 275:7
stating 81:14
statistical 221:7
stay 60:9 189:3
247:2
stayed 32:20
steer 87:7 143:16
221:12
steering 323:15
step 39:12 79:2
195:20 209:20
291:6 309:22
stepping 103:19
steps 103:3 321:6
steer 87:7 143:16
stayed 32:20
stay 87:14 162:17
321:6
structured 35:19
structure 42:8
87:14 162:20
258:12 261:18,19
261:19 262:15
312:22
steering 323:15
style 133:13,14
206:1
subcommittee 205:14
subcriteria 326:4,9
subject 128:1 221:3
236:19
subjective 281:17
submit 154:4,13
subjectivity 323:19
submit 17:5 77:13
227:7
submitted 295:2
submit 154:4,13
sublevels 103:3
sub-criteria 321:6
structured 323:15
superstorm 4:8
support 18:5 19:3
19:11,12,14 84:17
127:6 192:1 196:2
199:7 330:4
349:12
supported 5:18
survive 61:5
89:17 120:8
145:18 248:5
sure 12:10 15:14
17:19 21:7 27:5
38:13 40:6 42:13
51:16 80:3,8
102:13 115:17
117:9 122:19
127:22 130:22
138:8 142:5
144:16 195:12
221:1,14 225:5
227:16 229:8
238:12 242:5
247:7 254:22
257:3 283:19
291:11 294:22
295:6,19 302:6,16
307:3,18 318:3
322:22 325:9,22
329:8 331:15
340:1 344:11
358:20 365:5
366:16
surprise 68:17
surprised 243:4
surprises 83:7
211:20
surrounding 201:12
surrounds 353:6
survey 23:13
111:17 123:20
150:4 161:22
191:16 192:3,4
197:5,14 204:9
206:13
surveys 99:19
161:20 242:17
334:4
survive 161:10
submit 83:12
systematic 13:21 26:13
53:8 63:3 77:17
79:10 116:12
118:18 122:3,3
139:14 144:2
153:18 156:21
163:4 167:18
174:6 191:19
219:15 230:2
232:22 237:9
238:18,18 239:6,7
239:20 241:16
251:2 260:7 261:3
263:7 278:21
282:22 283:2,3
293:20 315:2
319:10,15 331:13
341:9
systematic 21:1,19
158:5
systemic 21:1 56:3
systems 23:1,4,13
24:19 54:7 57:6
59:7,8,9,18,20
107:3 110:20
122:4 127:17,22
159:4,15 200:18
203:21 227:14

Neal R. Gross & Co., Inc.
202-234-4433
Page 407
16:13 17:2 24:21
90:7,7,13 98:13
34:12 37:9 61:14
102:11 113:6,10
73:15 192:10
113:12,16 116:12
360:11
122:5 123:17
theme 73:16
125:2 137:3
themes 22:21 75:16
139:18 142:2,11
129:18
144:17 145:7,14
thesaurus 193:14
152:3 157:1
193:20
160:19 161:4,6,11
they'd 365:6
161:13,16 162:10
thing 6:4 19:22
165:6 166:5,12,21
30:21 46:1 49:11
190:1 193:5,7,8
62:15 63:7 64:17
224:14 225:12
65:9 69:16 74:12
228:7 232:4 237:1
74:18 76:21 80:9
244:7 245:1,8,11
105:8 115:14
246:22 249:5
124:19 138:15
250:20 251:19
142:15 148:15
253:16 261:14
155:16 160:14
269:7,13,19 273:1
165:20 167:9,15
285:16 288:17
172:10 197:7
293:7 294:2,9,18
206:2 217:14
297:15 330:13
218:14 230:12
331:4 333:11
237:3 244:21
347:19 350:21
250:18 257:3
355:1 357:12
262:8 278:21
358:12 359:8
285:11 287:6
360:19 361:6
289:17 291:17,19 think 4:3,18 10:8
293:3 300:15
12:2 16:12 18:10
301:15 320:12,14
18:12 20:8 22:7
327:11 329:14
23:9 25:19 26:2,9
330:19,20 336:4
27:8 28:8,12 29:7
337:13 339:9
29:14,17,18,20
355:11 357:10
30:2,3,4 31:12,21
362:16 363:3
33:4 34:6 35:8,8
365:4
35:12,20 36:1
things 5:8 6:12
37:10,11,12 38:15
10:2 19:6 21:2
39:21,22 40:8,11
27:7 29:3,10 30:3
40:21 41:5 42:7,9
34:16 36:8 38:9
43:3,15,18,22
38:14 43:7 46:5
44:10,11 45:5,6
47:12 48:21 49:5
46:5,11,17 50:6
61:5 64:14 65:13
51:11 52:15,20
69:8 70:6 72:17
56:10,17 57:3
74:15,16 75:8
58:1,1,5,21 59:2
76:18 78:7 83:1
60:5,9 61:19 62:5
88:11,11 89:17
62:5,8,16 63:6,7

63:13,14,20 64:4
64:13,14,18,21,22
65:17 66:6,22
67:17 69:5,10,17
70:11,19,20 71:10
74:10 78:12,13,16
79:5 80:9 81:4,11
81:14 82:8 83:8
83:16 84:7,20
85:3 86:1 87:5,8
91:13,15 93:8
94:11,19,22 96:9
96:10,14,21,22
97:5,14,17 98:5,7
98:7,9,15,19,22
101:4 102:17,18
103:10 104:5,12
104:13,19,22
105:1,3,9,18,18
106:22,22 107:16
109:8 110:3,9,13
110:18 111:3,5,8
111:11,13 113:6,8
113:14,21 114:4,9
114:10,10,15
115:14,21 116:3
117:2,9,13,16,21
118:2 119:20
120:3,13,14,16,18
121:10,18 122:8
123:13,13 124:7
124:16,18,19,20
125:1,5,6,18
126:11 127:2,7
128:6 129:6,8,11
130:11 131:8
132:4,17 133:7,9
133:12 134:8
137:6,9,10,16
138:9 139:15,16
139:18,21,22
140:3,6 141:1,5
141:11,14,22
142:9,20,22 143:3
146:17,19 147:12
148:18 149:22
150:7,13,14,19,22

152:18,18 153:14
154:1,14 155:8
156:1,5,11,16,22
157:9,16,18,19,22
158:7,10,19 160:9
160:11,19 161:2
162:7 163:10,17
167:5,17 168:7,9
169:7 170:4,6,10
170:16,17 171:19
172:14 173:21
174:12,13,19
175:18,22 176:4,6
176:20,20 178:6,8
178:9,15 180:21
181:5,7,8 182:8
184:6 185:5,8,11
185:12 186:5,14
187:7,8,18 188:6
189:1,22 190:2,3
190:20 191:1,2,5
192:11,12,22
193:1 194:19
195:18 197:19,20
198:7 199:7,10
200:4,16 202:22
204:2,3,4,8,15,22
205:15 206:4,17
207:7 208:17,22
210:17 211:15,17
212:2,6,17,22
213:2,10,12,16,20
214:1,6 215:15
216:2,8 220:2,6
221:2,9 225:6
226:2,3,6,14,16
226:18,21,22
227:19 228:5,7
229:13 230:20
231:10,20 232:16
234:17 235:3
236:3 237:12
238:3 239:15
240:6 241:4,8
242:11,22 243:5
244:3,9,10,21
245:5,11 246:4

Neal R. Gross & Co., Inc.
202-234-4433

249:9,17 250:10
251:13 252:1,3
254:6,19 255:2,16
255:17 256:12
257:7 258:13,17
258:19,19 261:15
263:5,19 267:11
267:17,22 268:15
269:8,18 270:1,2
270:3 272:11,14
273:6,10,12,13,16
273:21,22 275:10
275:16 276:2
278:1,16 279:4,11
281:8,13 282:1,10
282:19 284:15,17
285:5,6 286:8,17
286:19,21 287:7
287:13 288:1,21
289:15,17 290:2,9
290:18 291:14
292:12,13,15
293:16 295:8,14
297:18,19 298:15
298:16 299:7,13
299:17 300:9,16
301:13,16,18
303:13 304:9,10
304:19,21 305:2
305:12,16,19,20
305:22 306:8
307:2,11 308:20
309:5,16 310:11
311:11,19,20
312:2,6,11,21
314:16 315:18
316:20 317:4,14
318:8 319:19,20
320:1,13 321:4,17
323:12,18 324:2
325:11,21 326:10
327:2,3,8,11,15
328:7,11 329:21
330:3,16 331:22
333:9,10 334:9
335:20 336:2,8,11
336:16,16,20


CERTIFICATE

This is to certify that the foregoing transcript

In the matter of: eMeasures Feasibility Assessment
Expert Panel Meeting

Before: NQG

Date: 12-07-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

______________________________
Court Reporter

NEAL R. GROSS
COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
WASHINGTON, D.C. 20005-3701