The Steering Committee met, in the Capital Room of the Venable LLP Conference Center, 575 7th Street, N.W., Washington, D.C., at 9:00 a.m., Tom Rosenthal and Bruce Steinwald, Co-Chairs, presiding. PRESENT: TOM ROSENTHAL, MD, Co-Chair BRUCE STEINWALD, MBA, Co-Chair PAUL BARNETT, PhD, VA Palo Alto Health Care System JACK BOWHAN, Wisconsin Collaborative JEPTHA CURTIS, MD, FACC, Yale University School of Medicine WILLIAM GOLDEN, MD, MACP, Arkansas Medicaid LISA GRABERT, MPH, American Hospital Association ETHAN HALM, MD, MPH, University of Texas Southwestern Medical Center (via phone) ANN HENDRICH, RN, MSN, FAAN, Ascension Health JACK NEEDLEMAN, PhD, FAAN, University of California, Los Angeles School of Public Health MARY KAY O'NEILL, MD, MBA, CIGNA HealthCare DAVID PENSON, MD, MPH, Vanderbilt University Medical Center DORIS PETER, PhD, Consumers Union
STEVE PHILLIPS, MPA, Ortho-McNeill-Janssen Pharmaceutical, Inc.
DAVID REDFEN, PhD, WellPoint
JEFFREY RICH, MD, Mid-Atlantic Cardiothoracic Surgeons Ltd.
WILLIAM RICH, MD, Northern Virginia Ophthalmology Associates
BARBARA RUDOLPH, PhD, MSSW, The Leapfrog Group
JOSEPH STEPHANSKY, PhD, Michigan Health and Hospital Association
DOLORES YANAGIHARA, MPH, Integrated Healthcare Association

NQF STAFF:
TAROON AMIN
HEIDI BOSSLEY, MSN, MBA
HELEN BURSTIN, MD, MPH
LAURALEI DORIAN
SARAH FANTA
ANN HAMMERSMITH
CAMILLE PRESBURY
LESLIE REEDER-THOMPSON
SALLY TURBYVILLE, MA, MS
ASHLIE WILBON, MPH, BSN
CARLOS ALZOLA, NQF Statistical Consultant

ALSO PRESENT:
BEN HAMLIN, NCQA
CHAD HEIM, HealthPartners
SUE KNUDSON, HealthPartners
TODD LEE, ABMS (via phone)

KEVIN STROUPE, ABMS (via phone)
ARJUN VENKATESH, Brigham and Women's Hospital
KEVIN WEISS, ABMS (via phone)
C-O-N-T-E-N-T-S

Call to Order 9
  Sally Turbyville, MA, MS
  Senior Director

Welcome 10
  Helen Burstin, MD, MPH 10
  Senior Vice President of Performance Measures

Bruce Steinwald, MBA 11
  Co-Chair

Introductions 12

Recap of Work to Date 16

Ashlie Wilbon, MPH, BSN 16, 55
  Senior Project Manager

Helen Burstin, MD, MPH 19
  Senior Vice President of Performance Measures

Sally Turbyville, MA, MS 31
  Senior Director

Disclosure of Interest 59
  Ann Hammersmith
  NQF General Counsel

Introduction of Others Present 66

Expectations and Process for Meeting 67
  Bruce Steinwald 67
  Co-Chair

Sally Turbyville, MA, MS 67
  Senior Director
C-O-N-T-E-N-T-S (CONTINUED)

Measure 1558: Relative Resource Use for People with Cardiovascular Conditions (NCQA)
  Ben Hamlin 69, 73
  NCQA

  Jeptha Curtis 69, 74

  Discussion 75
  Vote on Importance 90

Scientific Acceptability 91
  Jeptha Curtis 91
  Discussion 93
  Vote on Scientific Acceptability 140

Usability 146
  Jeptha Curtis 146
  Discussion 147
  Vote on Usability 161

Feasibility 161
  Jeptha Curtis 161
  Bill Hamlin 162

  Discussion 163

  Vote on Feasibility 182
Measure 1558 (Continued)

Recommendation for Measure Endorsement 183
Discussion 183
Vote on Measure Endorsement 186

Measure 1570: Acute Myocardial Infarction Episode-of-Care for 30 Days Following Onset (ABMS-REF)

Kevin Weiss 188
ABMS
and
Todd Lee
ABMS
Importance 192
Jeptha Curtis 192
Vote on Importance 195

Scientific Acceptability 196

Jeptha Curtis 196

Jeffrey Rich 202
Discussion 207, 218

Vote on Scientific Acceptability 276
Public Comment 216
C-O-N-T-E-N-T-S (CONTINUED)

Measure 1571: Acute Myocardial Infarction Episode-of-Care for Post-Acute Period (Days 31-365) (ABMS-REF)

Kevin Weiss
ABMS
and
Todd Lee
ABMS

Importance

Jeptha Curtis

Vote on Importance

Jeptha Curtis
Tom Rosenthal
Discussion

Vote on Scientific Acceptability

Measure 1572: Episode-of-Care for Management of Chronic Coronary Artery Disease (ABMS-REF)

Kevin Stroupe
ABMS

Todd Lee
ABMS

Importance

Jeptha Curtis

Vote on Importance
Measure 1572 (Continued)

Scientific Acceptability 324
   Jeptha Curtis 324
   Dolores Yanagihara 327
   Discussion 329

Vote on Scientific Acceptability 373

Measure 1604: Total Cost of Care PMPM 374
Index (HealthPartners)

Sue Knudson 374
HealthPartners and
Chad Heim
HealthPartners
Discussion 390

Vote on 1a 402

Vote on 1b 403

Vote on 1c 403

Vote on 1d 404

Vote on Importance 404
Scientific Acceptability 405
Discussion 405, 443
448, 462
480
C-O-N-T-E-N-T-S (CONTINUED)

Measure 1604 (Continued)

Vote on 2a1 442
Vote on 2a2 446
Vote on Overall Reliability 447
Vote on 2b1 462, 473

Vote on 2b2 475

Vote on 2b3 476

Vote on 2b4 477

Vote on 2b5 478
No vote on 2b6 (NA) 479
Vote on Overall Validity 479

Vote on 2c 486

Vote on Overall Scientific Acceptability 488

Public Comment 490
9:12 a.m.

MS. TURBYVILLE: So, good morning, everyone, and welcome, and a big thank you for finding time to come and meet with us today to talk about the resource use measures.

We really appreciate all the work that you have done thus far.

Sorry, I thought someone was snapping at me to call my attention.

(Laughter.)

I know we have a full agenda. So, before we get started, I wanted to ask Helen to provide some welcoming remarks and, then, your Co-Chairs. We will go ahead and go through the objectives for today and make sure we get any input from you on our objectives today, if needed. We will go ahead and get started as quickly as we can.

DR. BURSTIN: I will just add my welcome as well.

Helen Burstin from NQF.
I realize how much work is involved in reviewing these measures, having sat through a few of these TAP meetings. So, I just really wanted to say thank you. I guess we will have a chance to talk about some of the broader issues later.

MS. TURBYVILLE: Yes.

DR. BURSTIN: Okay. Great.

CO-CHAIR STEINWALD: Yes, has it been a full year since we met face-to-face?

MS. TURBYVILLE: It has.

CO-CHAIR STEINWALD: And how many hours of conference calls have we logged since then?

(Laughter.)

MS. TURBYVILLE: Must we count?

CO-CHAIR STEINWALD: No. No, I guess not.

So, from my perspective, it is really a pleasure to be meeting again face-to-face.

If you wouldn't mind, even though
we have done it before, could we go around the
room with people giving their names and
affiliations? I think we do have a few
newcomers, don't we?

MS. TURBYVILLE: Yes, and it may
be a good idea for efficiency sake to do that
with the disclosure of interest --

CO-CHAIR STEINWALD: Oh, okay.

MS. TURBYVILLE: -- if that's
okay. She's not here yet? Yes, so, then,
let's just go ahead and do the intros.

CO-CHAIR STEINWALD: So, go ahead?

Okay.

I'm Bruce Steinwald. I live right
here in Washington, D.C. For years, I worked
at the Government Accountability Office, but
now I am on my own.

CO-CHAIR ROSENTHAL: I'm Tom
Rosenthal. I'm the Chief Medical Officer at
UCLA in Los Angeles.

MEMBER BARNETT: I'm Paul Barnett.

I direct the Health Economics Resource Center
in the Department of Veterans Affairs in California.

MEMBER B. RICH: My name is Bill Rich. I'm local in D.C., and I'm the Medical Director of Health Policy for the American Academy of Ophthalmology.

MEMBER PETER: Hi. I'm Doris Peter. I work at Consumer Reports in Yonkers, New York.

MEMBER STEPHANSKY: Joe Stephansky. I'm with the Michigan Health and Hospital Association.

MEMBER RUDOLPH: Barb Rudolph, and I'm with the Leapfrog Group as Science Director.

MEMBER GRABERT: Lisa Grabert, here in D.C., with the American Hospital Association.

MEMBER YANAGIHARA: Hi. I'm Dolores Yanagihara with the Integrated Healthcare Association in California.

MEMBER O'NEILL: I'm Mary Kay
O’Neill. I'm the Chief Medical Officer for the Pacific Northwest in Seattle for CIGNA.

MEMBER GOLDEN: Yes, I'm Bill Golden, Medical Director for Arkansas Medicaid.

MEMBER NEEDLEMAN: Jack Needleman, Professor of Health Services at the UCLA School of Public Health.

MEMBER J. RICH: Jeff Rich. I'm a practicing cardiac surgeon at Sentara. Former life, I ran the Medicare Fee-for-Service Program for the Bush Administration, the last years of it, and am currently the President-Elect of the Society of Thoracic Surgeons.

MEMBER REDFEARN: I'm David Redfearn. I work with WellPoint, now based in Las Vegas.

MEMBER PENSON: I'm David Penson. I'm a urologist from Vanderbilt University. I also am the Vice Chair for Policy for the American Urologic Association.

MEMBER HENDRICH: Ann Hendrich,
Vice President of Clinical Excellence
Operations at Ascension Health.

MS. DORIAN: Good morning. I'm Lauralei Dorian, and I have recently started with NQF as a Project Manager. I will be working with the team on this project.

MS. FANTA: Hi. I'm Sarah Fanta, Project Analyst, NQF, and looking forward to working with you all.

MS. WILBON: Good morning, everyone.

I think everyone knows me by now because you've gotten at least a million emails from me.

But I'm Ashlie Wilbon. I'm the Senior Project Manager for this project.

It's good to see everyone, and thanks for coming.

MEMBER CURTIS: I'm Jeptha Curtis. I'm a cardiologist and health services research at Yale in the Center for Outcomes Research and Evaluation.
MR. AMIN: My name is Taroon Amin. I recently joined this team, about two months ago. I come to NQF from Brandeis where I was working on public sector episode-of-care work.

MS. TURBYVILLE: And I'm Sally Turbyville, Senior Director on this project.

Welcome.

MS. WILBON: So, while we are waiting for our General Counsel to arrive, we will go ahead and just into just to do a brief introduction, presentation for everyone this morning to get us on track for the next two days and make sure everyone is on the same page, and perhaps a few kind of overarching issues that we may run into over the next couple of days and allow people to ask questions and get that out of their system.

So, we have already introduced staff and done a roll call here. So, really, what we are here to do today is we want to make sure, again, that everyone understands the resource use measure evaluation criteria
and evaluation process. We realize that the
Steering Committee has been acting almost as
a TAP up until this point by doing the non-
condition-specific measures and evaluating the
subcriteria, the overall criteria, and making
recommendations.

So, as we start out this morning,
we are going to be kind of having you guys
shift gears into actually acting as a Steering
Committee and taking into consideration what
the TAP has already reviewed, and, then,
making your overall criteria ratings and then
recommending the measures.

And we do hope that, by the end of
this meeting or throughout the meeting, that
you guys would be able to provide us some
feedback on how the process went. This is a
little bit new for us, this particular process
with these measures. So, we are open to any
feedback on how you think that we could make
the process more efficient and helping move a
little bit smoother as we move forward.
So, these are the measures that we are going to be looking over today. We have got five cardiovascular measures, three diabetes, and two non-condition-specific measures. It is a really full agenda.

We may be looking at potentially tabling one of these measures, 1591, based on the TAP review wasn't quite complete. So, we will see how that goes as we get through the day. It is one of the last measures of the day. So, depending on how things are going, it may not be an issue. So, we will kind of keep you guys updated as we go through the day.

So, obviously, the purpose of this project is to endorse cost and resource use measures as a building block towards measuring efficiency.

And back a year ago, we discussed how we would like to define resource use for this project. What we came up with, just to jog everyone's memory, is that resource use
measures are broadly-applicable measures that compare health services in terms of units or dollars and can be applied to a population or event broadly defined to include diagnoses, procedures, et cetera. They count the frequency of defined health system resources. Some may further apply a dollar amount, allowable charges, et cetera, standardized prices, to each unit of a resource.

So, I think I was going to hand it over to Helen at this point, just to kind of talk a little bit about how this project fits in with some of the other work that NQF has done through our efficiency framework.

DR. BURSTIN: Just briefly, an issue that keeps coming up is this issue of endorsing resource use measures and how does that fit within the framework of NQF's being all about endorsing quality measures. So, I thought it would be helpful to just recap the work that was done a couple of years ago now for the NQF-endorsed measurement framework.
Actually, Bill Golden was one of the members of that Committee.

And just again to reemphasize, this was the ultimate finding of the Committee: efficiency measurement is multidimensional. No news to anyone here. It specifically said measurement within these constructs should not be pursued individually or in isolation, but, rather, as a subcomponent of a larger set of measures needed to adequately assess sufficiency overall. And they specifically listed definitions for quality, cost, efficiency, and value of care.

So, again, I know this keeps coming up. We very much view the need to endorse these measures as applicable, if they make it through your process, as the building blocks to let us start those subcomponents, as they talked about, for us to begin building measures of efficiency. This has come up a lot, particularly around the usability criterion, as some of the TAPs have struggled
with that, and I assume probably as you talk
about that today.

So, again, understand we are
looking at it in the context of how it applies
as a subcomponent of a broader efficiency
measure. We would not advocate using it
completely on its own divorced from quality.

And on the next slide, this was the
other work of the Committee, which was really
just trying to begin thinking in an episode
framework, as you guys were doing today.

This slide is a little different in
that it specifically has the overlay of the
National Priorities that emerged from the
National Quality Strategy the Secretary
promulgated recently.

So, again, as we start thinking
about these episodes, we are also trying to
think about our ultimate measurement framework
for NQF, as we want to be able to move toward
the longitudinal assessments of quality for
high-impact conditions or multiple conditions,
which is a very common scenario for those of
us who see primary care patients at least,
allayed with having those cross-cutting
measures that allow us to look at those high-
profile, cross-cutting areas.

The other thing, just briefly, is
that, as we talk about usability, we have been
making some significant progress in really
thinking through what usability means in this
sort of emerging era for NQF-endorsed
measures. We have a Usability Task Force that
Chris Queram from Wisconsin is going to chair
for us on July 27th to really take a critical
look at that criterion.

But, at least for now, we really do
view it as measures useful for a broad range
of public accountability functions, not just
public reporting. Public reporting is where
we really want to go, obviously, for as many
of these measures as possible, but also
recognizing there are other important
accountability functions, pay-for-performance,
certification, accreditation, that are also
important usability functions for these
measures as well as quality improvement.

With that, I will turn it back to
you, Ashlie.

MS. WILBON: Thank you, Helen.

So, again, the next couple of
slides I'm going to breeze through, but just
about the consensus development process. We
are, obviously, in the standards review step.
Then, once we have a set for cycle one, which
we are hoping to do a report based on the non-
condition-specific measures and the
cardiovascular measures, and send those
through the process as a pack or as a group.
And, then, as we finish the second-cycle
measures for pulmonary and bone joints and
cancer, that those will go through in a second
report. So, again, the process here.

We wanted to put this upfront in
the beginning, so that we are acknowledging
for everyone some of the challenges that we
anticipate, and we have actually encountered, after now having two TAP meetings, we have meet with the Cardiovascular/Diabetes TAP, and yesterday Dr. Penson chaired the Cancer TAP, which went very well.

So, we are encountering some of these challenges along the way, and we are trying to address them as we go, but kind of just pointing out that, obviously -- and you guys have seen these before -- that these are the first resource use measures that we have ever evaluated.

Particularly with the first cycle, we have experienced some time constraints. The timeline was very tricky. We have done our best to try to move the measures through it, and I think what we are finding is that, for the second cycle, for the Pulmonary, Cancer, and Bone Joint TAP, that we are finding things to be a lot easier if we have more time. We have spaced things apart a little bit different. It is making things a
lot easier.

So, I think you guys will also begin to feel that things will get easier as we have gotten better; we have gotten more efficient with our process as well. So, bear with us. We realize it is going to be a little bit bumpy, but just be patient with us.

Again, the size and length of these measure specifications, particularly for some of the developers, they do get very long. The complexities of the measures and, then, again, applying the slightly modified criteria to the different measures.

So, this slide just illustrates what I was talking about before of how we have grouped the measures -- and you have seen this before -- for the two cycles. I am not going to spend time on this, but we are on time, as of now, moving through the timeline for both of the cycles and hope to have a set or a group of endorsed measures, or whatever makes it through, by January for the first cycle.
So, activities to date: I did want to let everyone know the results of the vote for the HealthPartners measure that everyone voted on. We had 17 of the Steering Committee vote, and for the 1598, total resource use measure from HealthPartners, it was recommended for endorsement 11 to 6.

We did get about halfway through the evaluation of the Ingenix 1599, the ETG-based, non-condition-specific measure, for the review of the importance and scientific acceptability. We will pick up reviewing the remainder of the measure on day two. We kind of wanted to get you guys into Steering Committee mode, reviewing some of the stuff the TAP has already done to start out with, and, then, we will circle back to what you guys have already started.

So, there are 32 measures for this project. I have just kind of listed them out here and kind of where we are in the process.

There is also a table in your
folder that I had emailed out before, but it
is a measure status table, just so you guys
can kind of keep track of how the measures are
moving through the process. It kind of looks
like this. It has got some grayed-out rows,
and it just kind of illustrates the condition
of the measure category, the measure name, the
developer, where it is in the TAP review
process, and when we expect the Steering
Committee to review it.

So, it gives you an idea of what we
are hoping to kind of move through and when,
so kind of what the workload is going to be
for the next couple of months. So, hopefully,
that is helpful to you guys, and if you have
any questions, let me know if you have a
question.

MEMBER BARNETT: Just one question.

So, that says that 1599, the Ingenix non-
condition is complete? But it, actually, is
still pending, right?

MS. WILBON: Right, right. We will
have to update that. Thank you.

MEMBER BARNETT: I was just worried I missed something.

(Laughter.)

MS. WILBON: Oh, yes. No, no, you're right. Thank you.

I think that was a little optimism on our part, that it would actually be done by the time we finished that conference call, but it wasn't. So, that's okay.

(Laughter.)

So, just a quick recap for you guys, as you get into evaluating the measures as a Steering Committee of what the TAP has been instructed to do for their responsibilities in terms of evaluating the measures.

So, we have asked the TAPs to evaluate the measures against the evaluation subcriteria, identify strengths and weaknesses of the measures, particularly focusing on the clinical aspects and applications of the
measures. We also have seated methodologists and other people of a technical nature to kind of really do the deep dive, particularly into the scientific acceptability of the measure. So, hopefully, you will see that reflected in some of the feedback that you get from the reviews that have already been done.

And again, the role of the Steering Committee is to review the TAP input and evaluation ratings, identify and discuss any TAP areas of concern. There may be areas of concern that you have that the TAP didn't identify. So, obviously, we want you to highlight those as well.

And, then, we would ask you, based on the ratings, to rate the overall criteria. So, for instance, importance has four subcriteria, which the TAP has already rated, and then we will be asking you to give an overall rating for importance of yes or no, if the measure passed, and so forth, for each the remaining criteria. And, then, finally, make
a recommendation for endorsement.

So, I am going to actually hand it over to Sally at this point to kind of talk through the evaluation process.

Yes, Bill?

MEMBER B. RICH: Ashlie, one question --

MS. WILBON: Sure.

MEMBER B. RICH: -- just a procedural one. Did Dr. Curtis' Committee have the report of Carlos, you know, the technical analysis at their TAP meeting?

MS. WILBON: They at Carlos at the meeting, but I don't believe -- Carlos, he's there. I think because he came on after we had already started the evaluation process, that we probably had distributed them maybe midway. And, then, he came to the meeting. He presented verbally at the meeting, but --

MEMBER B. RICH: So, the report that we reviewed, they did not have the advantage of reviewing before the meeting
then?

   MS. WILBON: Before? I don't think so.

   MS. TURBYVILLE: And yours have been updated based on input that we may have gotten after the TAP meeting as well. So, hopefully, those are the most recent evaluations of all the information we have.

   MEMBER B. RICH: That would explain some discrepancy. Thank you.

   MS. WILBON: Yes. And, also, from the thumb drives that we gave everyone, his reviews of all the measures are in a folder for consultant review. So, if you want to look in there, those are the most up-to-date as well.

   Go ahead.

   MS. TURBYVILLE: So, we just want to recap quickly some of the principles of the resource use measures that all of you outlined for us. And they are here for you to look at. I won't read through all of them. They are in
the report that we worked on with all of you.

But there were 11 of them, and I think that they clearly set the groundwork prior to us, then, requesting for measures to be submitted, including making sure that we were open to all types of resource use measures from a population, episode and procedures, and make sure that we are trying to consistently send the signal that we realize and acknowledge that these are measures of resource use. Our hope, as Helen said, is that we are getting ourselves, we are building blocks to get to value and efficiency.

Go ahead to the next slide, Ashlie.

So, this just continues through these principles.

And next slide. That's fine.

So, then, as we think about endorsing the measures that we are doing today, we do have the four criteria. We worked with all of you to update it as
necessary in order to allow them to adequately evaluate measures of resource use, but we still have the importance to measure and report the scientific acceptability of the measurement properties, how usable or useful is the measure, and their feasibility.

And later on in the process, if we do find that the Steering Committee is recommending measures that are similar, we will work with you to provide some justification to understand why NQF would be putting forward two similar measures. But we will wait and see what happens before we do that. We don't jump the gun at this point.

So, what we are going to be asking of all of you today is to evaluate and rate the measures based on the overall evaluation criteria. The TAPs have already gone through the sub- and the sub-sub-criteria, and you will be using that as input points.

But we will really just be asking you: was the measure important? Are the
scientific acceptability criteria met, et cetera? So, we won't be asking you to rate all the sub-criteria, as you did for the population-based measures.

So, the first one, is the measure important to report and measure? And it is really about the focus area of the measure. This is prior to getting into the very details of how the measure is constructed. It is, is this area in which the measure is examining, for example, episodes of care and cardiac heart failure, is that important to measure resource use there? It is important to measure the resource use in a population-based measure?

And we ask all of you to vote on that first. Because if a measure is found by the Steering Committee to not be important, then we don't go through the rest of the criteria. This is part of the hierarchy that has been talked about before. So, it needs to be important in order for it to be worthwhile
for us to go through the rest of the criteria.

    For the scientific acceptability,

as all of you know, the focus is on the
reliability, the ability for the measure to be
reproduced, based on where it is being
proposed for endorsement, and the validity,
how well is the measure measuring what we
think it is intended to, or the developers
tell us it is intended to?

    And, then, we also ask you to think
about disparities. The TAPs have had some
interesting conversations about disparities
and how does this weigh into resource use. Is
there enough in the literature right now to
think about stratification by socioeconomic
status, or even if the data are consistently
available? And we would certainly benefit
from further conversation from the Steering
Committee to provide guidance.

    And I think Jeptha can probably
articulate very well with that Cardio and
Diabetes discussed in terms of disparities.
And, then, David, who just chaired the Cancer TAP -- I'm kind of springing this on him -- might briefly share what that TAP talked about as well. But we are looking forward to your input on this.

So, these are the sub-subcriteria of 2a, which is reliability. So, we did have, for the example, the TAP, as all of you did for the population-based measures, think about each of these very detailed points, which, then, feed into whether or not the measure is considered reliable.

And the same with validity, which has six sub-subcriteria. So, it is really a deep dive into the measures, along with the benefit of the consultant review in these areas.

So, briefly, so you have some context for what the ratings meaning -- and we did go over this with the TAPs as well -- when we are talking about a high rating for reliability, the threshold is that the measure
developer has demonstrated that both the data
elements and the measure score demonstrate
that they are reproducible and consistent.

And, then, the same for validity.

It is a high bar, and it is really looking
that the measure developers are demonstrating
that the data elements that are used to
support the measure, as well as the measure
score that comes out after the measure is run,

demonstrate validity. We also ask that they
have considered threats to validity and have
been transparent about what those are and
addressed, when appropriate.

Ashlie, next slide.

So, moderate, you can see when we
think about reliability and validity, is you
have this "or" option. So, they might
demonstrate that the data elements are
reliable or that the measure score is
reliable. And, then, the same validity, that
they can focus on the data elements
demonstrating properties of validity or the
data score themselves.

Next slide, Ashlie.

And, then, low is really the low bar where the measures are not demonstrating reliability or validity on either of them. And, then, there is the possibility for insufficient evidence, and this would be when the testing protocol or methods applied do not support any examination of whether the measures are reliable or valid. I will say that in the testing report NQF did state that face validity is the minimum threshold for demonstrating validity. So, that might give someone a moderate -- you know, it has to be a systematic true face validity, a systematic review of the measure demonstrating face validity.

And this crosswalk, so to say, or matrix is very helpful, I found it. It kind of demonstrates the mix of how you think about how high reliability and, then, you might have a moderate validity, and how that would
determine whether it is passing the scientific acceptability of the measurement properties.

And so, I don't know if there are any questions about this table. And it does come from the testing report where NQF convened a Testing Task Force that really thought very in-depth about these types of issues for scientific acceptability and how developers would demonstrate that, both to give developers guidance as well as Steering Committee guidance in thinking this through.

Okay. And, then, again, the disparities that we talked about. And clearly, as you know, for the quality measures, we don't want disparities to be risk-adjusted away. Often, we want them to be exposed, so that there can be action taken on them. And clearly, we do know that there's probably an evidence of disparities in resource use. The question is, what does that mean for measure reporting and stratification?

Yes?
MEMBER GOLDEN: When you are looking at the reliability and the validity and all these measures, we were talking earlier that some of these measures end up with substantial exclusions or case removals. So, you might have a reliable and a valid measure after you've gotten rid of all the exclusions.

How does that all factor in? Or how are you playing with that?

MS. TURBYVILLE: That's a great question. And one of the things that has come up in the TAP discussions -- and, Jeptha or David, please feel free to jump in -- is that if there are too many exclusions potentially made, that maybe the intended target audience is too narrow. So, what is really being measured? Or perhaps it comes up in a sample size issue. Now are the samples too small?

So, I don't know if you have any comment.

MEMBER GOLDEN: Like I said, that
raises a question of generalizability, I guess. And so, is that part of the assessment?

MS. TURBYVILLE: I don't think so because it would be generalizable to that narrow population. But it gets to, I think, whether or not it is measuring what is intended to be measured.

And, please, as clinicians, feel free to --

MEMBER PENSON: Yes. So, this came up in the Cancer meeting. I think that there were a number of measures where, once you started applying the exclusion criteria, your sample size got very low. And the TAP really started to feel as though, well, maybe this isn't really applicable to all patients with this disease.

And the scores were affected, actually, in the validity scores. That is where the TAP sort of ended up putting that. Because it basically said, well, is this
valid? Do this measure what we think it measures in the population that they have defined? And the answer was, no, it doesn't pass the smell test, the face validity test.

So, I think the TAPs, at least the Cancer TAP took that into account.

MEMBER GOLDEN: So, as we look through this, then, the notion of validity would be to the general population with that disease, rather than the operation of the measure, as defined, when you get rid of all the exclusions. I mean that is a technical --

MEMBER PENSON: Yes. And, Jeptha, I am curious to hear what your TAP felt. But I think, in the end, the TAP sort of, this is a moving target. People are sort of making it up as they go along, for lack of a better way to put it.

You know, there was no easy place to put that. It wasn't in the usability piece. I think the usability piece is what the TAP wrestles with the most at this point,
frankly. Because even if you get a meaningful number, no one knew how to interpret that.

But, that being said, when you are talking about generalizability, everyone sort of said that is a validity issue. I mean there is the statistical and mathematical validity, but there is also that sort of, you know, criteria on face validity. I mean, does this make sense to you as a provider? And I think that is where you are going to see that effect.

I don't know if that happened in the --

MEMBER CURTIS: I think we took a slightly different tact with our TAP. But I think we really considered those exclusion criterias and the generalizability of the resulting measure in the scientific acceptability. I think that is where we saw the predominance of those comments.

I think we considered it in validity testing inasmuch as most of the time
we were assessing face validity. But I don't think we really made a clear distinction as to where that generalizability criteria would be. And so, I think there are elements of it within scientific acceptability as well as within the ability to do testing.

I don't know if that --

MR. AMIN: The only thing else I would add, I mean, from both of the TAPs, I think what we are seeing is it actually came up in two places.

In 2A1, which we will go into, there was a discussion around whether the measure was well-defined, which is really where it looked like the CV/Diabetes TAP went, and 2B3, where the exclusions were supported by clinical evidence, was really where Cancer evaluated them. So, really, it came up in both places. So, I think that is why you are seeing it having come up in both places.

MEMBER NEEDLEMAN: I am trying to think about the potential use of these
measures and how the exclusions relate to them. That should affect, potentially, the way we think about the exclusions and where the end gets driven to.

The goal of the exclusion is to create a cleaner comparison. So, I can compare Provider A to Provider B and not worry about idiosyncratic cases that may be in their panel.

It drives down the end, which makes the precision of the estimates less useful. It also has the risk of excluding cases where there are resources, obviously, being used. So, the clean comparison, we have to ask whether the resources used in the excluded cases are likely to be correlated with the resources used in the cases that are left in. That makes the comparison valid.

The other issue is, for a provider looking at their ranking, looking at their data, looking at the drilldown in the data for the patients that are included, are the things
that they would do to change their resource
use based upon what they see in the data for
the patients that are in the measure
consistent with what they would do for the
patients who are excluded? That is to say, is
not only the resource use correlated, but are
the actions that the provider would take
correlated within their larger panel?

And if we are uncomfortable with
that, then the exclusions are not doing their
job. They are allowing a cleaner comparison,
but they are not allowing us to draw broader
conclusions about resource use for this
provider for the whole panel with this. And
it doesn't give them all the guidance they
need to change the resources.

CO-CHAIR STEINWALD: Bill and I
have consulted and think that many of these
issues we are discussing right now would
probably be maybe better discussed in the
context of a particular measure. So, we are
thinking maybe we should move on through the
agenda and, then, address these issues as we
do.

CO-CHAIR ROSENTHAL:  Bill, do you
want to have one comment on this?

MEMBER B. RICH:  Yes.

CO-CHAIR ROSENTHAL:  Because,
otherwise, we will get to this when we get to
the individual measure.

MEMBER B. RICH:  And I will
withdraw my discussion.  Then, just a question
that I raised at the end of our Steering
Committee, Bruce.  To really look at this, if
you are not going to exclude things, you have
to stratify them.

And do we have any inclination that
these people developing measures in the
commercial world are going to start collecting
data on ethnicity and race, as mandated in the
ACA?  And that was unresolved in our Steering
Committee call.  That is one way where you
don't have a lot of exclusions and decrease in
your end, but you can't stratify if you are
not collecting the data.

So, I was wondering, does anyone have that answer, especially on the commercial side, because Medicare already collects that data?

CO-CHAIR ROSENTHAL: I think the answer is we don't know.

MS. TURBYVILLE: And Jeptha and David are absolutely right; this is my recollection, that what was found in scientific acceptability does affect how, then, the TAP thought about the usability. And I am sure it will also affect all of you today, as you think about your ratings on usability.

Usability does want to assess how meaningful and understandable the measures are for public reporting, accountability, and quality improvement, and transparency, and the ability for people to understand what is being measured. So, that is, clearly, going to be affected by how you assess the scientific
acceptability of a measure.

Ashlie?

MS. WILBON: So, some context for usability that we want to provide to all of you today because NQF continues to learn more about how to frame usability. And in particular, acknowledging that this effort is the first time that NQF has collected resource use measures for the CDP process.

And what I am about to say applies to the first time that we do the same for a quality measure as well, that we realize that some of the measures that are collected have been tested in discrete databases and haven't been nationally implemented. That's okay. That might affect how you vote, you know, high, low, or medium, on some of these ability criteria, but we acknowledge that is often going to be the case. When we do an endorsement process for maintenance, at that time we would expect the measure developers would be providing more information on what
has happened in the subsequent three years.

And so, when we think about public reporting, we are asking the developers to demonstrate that the results are meaningful and understandable to the intended audiences, and that they are useful both for public accountability and informing performance improvement.

And this is consistent with NQF policy, again, for all measures, quality as well. This is not a special change for the resource use measures. And we acknowledge, also, that these measures are building blocks for efficiency or value.

And to give you an idea of what we think about when we are talking about accountability and public reporting, you see benchmarking all the way on the left. When we are talking about benchmarking, we are not talking about the benchmarks that are produced in resource use measures.

We are talking about quality
improvement, those internal quality improvement projects that various systems or organizations undergo, whether it is process oversight or actual quality measures. Those type of measures that perhaps are only suitable for internal improvement efforts are not what NQF looks to endorse because they don't really necessitate an endorsement for national implementation. They are not for comparisons across organizations.

However, all the things as you move towards the right there, certification, accreditation, et cetera, those are the types of measures where we are talking about accountability, that we are looking to endorse and are requesting and, through that process, are ensuring that there is transparency in what is being measured by those measures.

And, then, thinking about feasibility -- and this will come up as we ask the Co-Chairs to really lead these conversations for the resource use measures --
we weren't 100 percent certain if we would get administrative-only measures submitted for this project. That is what we anticipated. You know, perhaps some of them have been able to figure out clinically-enriched or other types of integration of data.

But, indeed, all the measures that have come through are measures based on administrative data. So, for a and for b, when we are talking about, are the data elements routinely generated, they are generated by claims data. Certainly, if anyone wants to discuss that, it will be open to the Steering Committee.

And, then, also, are they electronically available? Administrative data are electronically available. So, 4a and 4b for this particular effort are a little bit more straightforward, at least across all the measures.

Then, certainly, we want your input on the errors or unintended consequences and
assessing, has the measure developer thought
about ways and implemented ways to minimize
that or monitor how these measures, once
implemented or while implemented, are creating
unintended consequences or they identify
errors?

And, then, data collection, are the
data that need to be used to support the
measure available? And can the measure
operationally be implemented?

So, just as a reminder, this call
is open to the public. We already have the
lines open to the public. And, then, we will
pause here and there, and we will signal the
Co-Chairs here to make sure that we allow both
the public and the audience that we have here
with us physically to ask any questions or
provide input to the Steering Committee and
open the lines as well.

We have the measure developers here
today. Just to kind of give you an idea of
how this is going to work today or how we are
proposing it, if you look -- and all of you received this, the table that has the various assignments for each of you -- what we are going to have happen today is we are going to ask the measure developers to introduce the measure that you are about to review, provide you the description, et cetera.

Then, we are going to hand it over to the TAP Co-Chair, which for today will be Jeptha because we are going to be doing the cardio measures here today. And, then, Jamie, who was the other Co-Chair for the CVDM TAP will be leading the diabetes measures. And Jeptha is going to introduce the TAP discussions to all of you and provide you some context of what the TAP discussed and how they rated the measures.

And, then, we will go to the Steering Committee assigned reviewers and move through importance, et cetera. So, we will have the Co-Chairs lead the importance discussions. We realize that there might be
opportunity for us to gain some efficiency on
the importance area and, then, from there, ask
the Steering Committee who was assigned for
the scientific acceptability, usability, et
cetera, criteria, to lead off the discussion.

Now what we want is everyone to
participate in these discussions, provide your
input. So, having a lead reviewer is not
meant to limit the discussion. It is just to
kick it off, ask questions. They did a deep
dive, et cetera. We are hoping that helps
facilitate the conversations here today.

I just did that. I am going to ask
Ashlie to do the electronic voting. So, we
will be voting today, and you have the
clickers in front of you. And Ashlie is going
to describe that process.

MS. WILBON: Okay. So, this is
something new that we have been using in the
last couple of months. So, rather than
everyone raising their hands and us counting
hands for votes, we have started using an
electronic voting tool.

So, we have a laptop over here with a sensor on it. So, as we move through the process and we are ready to vote on a particular measure, we will have each of you enter your vote.

And let's see here. So, on your keypad, let's see, I guess it is not on here. But on each slide that we pull up for the voting, it will say, if you hit one, that means high, if you hit two, that means moderate, if you hit three, that means low. So, we will prompt you and walk you through it when we get to that point.

But we are just going to have you all point to this laptop over here since the sensor is in this direction. And we will know at the point when everyone has voted, and the results will be projected up on the screen, so you can see the distribution of who said high, medium, low or yes/no. And, then, we will read that out loud for everyone and for the
people on the phone and in the room, and then
move forward.

So, there will be instruction as we
go. So, don't fret. But that is what those
remotes are for you. And if you have any
issues along the way, let us know.

There is also a one-pager in your
folder with some instructions on how to vote,
and if you want to change your vote, what
buttons to hit and all that stuff.

So, let us know if you have any
questions, and we will recap before we vote
again.

And just very quickly, too, Ann is
here to lead us through the disclosure of
interest. So, I will just wrap up here and
ask for any last, final questions.

So, we will do any developer
followup and forward it to the Committee as
needed for review today. There are developers
here. So, hopefully, they can provide some of
the information that you might need here in
1 person.

   We are expecting some followup conference calls and save-the-dates to be emailed this week, based on the survey that we emailed, the availability survey that we emailed out. We are hoping that we may be able to get through some of the remaining measures in the next about three conference calls and the in-person meeting.

Then, again, the next in-person meeting is August 30-31st, and we are looking to have that meeting be focused on just the pulmonary measures. So, the Cancer TAP has already gone. So, we will try to do those over a conference call between now and August. The Bone/Joint TAP will be going next week, will be meeting next week. So, we are hoping we will be able to, hopefully, address those in a conference call. And, then, we will try, and we, also, actually need to wrap up the cardiovascular/diabetes measures, which the TAP meets again on July 14th.
So, we are trying to kind of work everything in and get as much done as we can, so that the next in-person meeting we are not left with too much stuff left over, and we can kind of wrap things up at that point.

So, again, thank you for all of your time, and bear with us through this process.

Does anyone have any questions before we move forward?

(No response.)

Okay. I will hand it over to Ann, who is here now. Thanks.

MS. HAMMERSMITH: Good morning, everyone.

I am Ann Hammersmith and NQF's General Counsel. I am here with you just for a few minutes, so that we can do the disclosure-of-interest portion of the meeting.

If you recall several weeks or even months ago, you should have received a form for us where we asked you some specific
questions about your activities and your affiliations. You completed that and returned it to us. We reviewed them carefully.

What we like to do in an open meeting is have you disclose any interests that you believe are relevant. Just because you disclose something does not mean you have a conflict. The idea here is to be open and transparent. So, you don't need to be concerned, if you do, indeed, have something to disclose, that you are in some way conflicted.

I just want to remind you that we do not expect you to summarize your CVs, which I am sure are quite lengthy in all cases. We do ask you to disclose things that you think are relevant to your service on this Committee.

We are specifically interested in your disclosure of grants, research support, consulting relationships, or speaking relationships that you may have that may be
relevant to the subject matter before the Committee.

We also want to remind you that you sit on this Committee as individuals, not as a representative of the organization which with you are affiliated, including any organization that may have nominated you for service. Sometimes people forget that. And we want to remind you that it is very important to keep in mind that you serve as an individual. You are here because you are an expert and we value your individual insights.

So, with that, I am going to ask you to go around the table, identify yourself, tell us where you work, and, then, if you have anything to disclose.

So, I would like to start with Dr. Rosenthal.

CO-CHAIR ROSENTHAL: Our hospital has a small consulting arrangement with Ingenix.

CO-CHAIR STEINWALD: Bruce
Steinwald. I have nothing to disclose.

MEMBER CURTIS: Jeptha Curtis. We have contracts with CMS for development of quality outcomes measures.

MEMBER HENDRICH: Ann Hendrich. I'm serving as principal investigator on an R18 AHRQ grant for reforming medical liability and patient safety. I also manage the Premiere contract for Ascension Health.

MEMBER PENSON: David Penson. I am the PI for one of the AHRQ Choice Awards in prostate cancer, and one of the aims does deal with quality-of-care measures. Also, in my role with AUA, as Vice Chair for Health Policy, I am a paid consultant to the Board of Directors.

MEMBER REDFEARN: I am David Redfearn. I work for WellPoint. I have nothing to disclose.

MEMBER J. RICH: Jeff Rich. As the President-Elect of the SGS -- the SGS is obviously a quality measure developer -- but
I have nothing to disclose here.

MEMBER NEEDLEMAN: Jack Needleman from UCLA. Nothing to disclose.

MEMBER GOLDEN: I'm Bill Golden. As Medical Director of Medicaid, we are working with Blue Cross, who is using Ingenix for looking at data. I'm also on the Executive Committee of the PCPI.

MEMBER O'NEILL: Mary Kay O'Neill, Chief Medical Officer for the Pacific Northwest for CIGNA. Nothing else to disclose.

MEMBER YANAGIHARA: Hi. I'm Dolores Yanagihara with the Integrative Healthcare Association, and I have nothing to disclose.


MEMBER RUDOLPH: Barb Rudolph. I'm employed by the University of Wisconsin, Madison, Center for Health Systems Research
and Analysis, as a senior scientists, and I have contracts with the Leapfrog group and also with the National Association of Health Data Organizations. I have nothing to disclose.

MEMBER STEPHANSKY: Joe Stephansky. I'm with the Michigan Health and Hospital Association. I have nothing to disclose.

MEMBER PETER: Hi. I'm Doris Peter from Consumer Reports. We license data and publicly report data from some of the organizations that have submitted measures, like NCQA and groups like that.

MEMBER B. RICH: My name is Bill Rich. I get a stipend from the American Academy of Ophthalmology as Medical Director of Health Policy. We develop measures. I sit on most of the alphabet soup quality organizations.

I have noted the fact, and I forgot to put this in mine, I was added to a Cost-of-Care Workgroup about a month ago, and I have
been one call. But I have no financial conflicts to disclose.

MEMBER BARNETT: Paul Barnett. I work for the U.S. Department of Veterans Affairs. I have nothing to disclose.

MS. HAMMERSMITH: Okay. Thank you. Are there any Committee members on the phone, Sally? Are there any Committee members on the phone?

MEMBER HALM: Ethan Halm. I work at the University of Texas Southwestern in Dallas, and have no disclosures.

MS. HAMMERSMITH: Okay. Thank you. Is there anyone else on the phone who is a Committee member?

(No response.)

Okay. Thank you for those disclosures.

Do you have any questions of each other or anything that you would like to discuss with each other regarding these disclosures?
(No response.)

Okay. Thank you. Have a good meeting.

MS. TURBYVILLE: And before we start, I do want to remind everyone, if you have forgotten it, and acknowledge that I did work at NCQA during the development of the resource use measures. So, when NCQA measures come up, I'll just be very quiet. The staff have led the review of those measures. The only thing I did was make sure their submissions were complete. And so, I just want to remind everyone of that relationship that was in the past existing.

CO-CHAIR STEINWALD: Before we begin, could we have the people at the back of the room identify themselves? Carlos, I think you raised your hand, but could you acknowledge that you are, indeed, Carlos?

MR. ALZOLA: I'm Carlos Alzola. I am an independent statistical consultant, and I was hired to review these measures.
MS. KNUDSON: Good morning.

I am Sue Knudson with HealthPartners.

MR. HEIM: I'm Chad Heim with HealthPartners as well.

MR. HAMLIN: I'm Ben Hamlin with NCQA.

DR. VENKATESH: Arjun Venkatesh from Brigham and Women's and Mass General.

CO-CHAIR STEINWALD: The agenda says "Expectations and Process for the Meeting". My expectation is that we should go forward.

(Laughter.)

Tom?

MS. TURBYVILLE: Yes, just to briefly explain what you are looking at, and I should only have to do it once, these are the compilation, both from in-person meeting and any followup votes or ratings that we got from the TAP members.

You will see the name of the
measure at the top. Then, you will see it distributes by the subcriteria; high, medium, low, and if there is an NA or insufficient. You can see on this one, there are nine, I think it's nine. And it's highs, and then the orange is the low, and the green is the medium. And it is consistent across.

So, we just pulled these up for you to have as a reference, but you will have the feedback from Jeptha, and we will move through them for the measures for you.

CO-CHAIR STEINWALD: Jeptha, you're up.

MEMBER CURTIS: I think we are going to start with having the measure developer provide their overview --

CO-CHAIR STEINWALD: Oh, okay.

MEMBER CURTIS: -- and, then, go from there.

CO-CHAIR STEINWALD: Okay.

CO-CHAIR ROSENTHAL: So, I think, NCQA, we are doing No. 1558.
According to the schedule, we are due to take a break at 11:00. So, just as a time check on us trying to get through this in an hour, it will sort of test our metal in doing it in this fashion and not being a committee-of-the-whole.

So, I think you're on.

MR. HAMLIN: Thank you very much. Can you hear me?

So, NCQA has currently five condition-specific total annual population-based measures that are reported at the health plan level. Cardiovascular conditions is one. These are risk-adjusted measures of utilization using, for the clinical side, for identifying the eligible population using primarily identification criteria that are defined, that correlate with our HEDIS measures. So, that's the two-minute overview.

MEMBER CURTIS: As everyone knows, the Diabetes TAP, the Diabetes/Cardiovascular TAP had its work cut out for them, reviewing
I think a total of, well, supposed to be reviewing 14. We have whittled it down some, as measures peeled off.

We chose to start off with this measure because the NCQA measures in general were, I think, more straightforward than some of the other developers' measures. But, that being said, for 1558, it was reviewed not in person, but a subsequent phone call in which only, I think, five of the 12 members were able to attend.

So, because of that and because of its overlap with the other condition-specific measure of diabetes, I will be sort of creating a conglomerate of the comments across both measures where I think there is applicability.

So, walking through importance, I think this is probably true for just about most of the measures that we are going to review today, in that there was really not a whole lot of disagreement about the importance
of the measures. And this one specifically, obviously, chronic cardiovascular disease is a high resource intensity and highly-morbid and mortal condition.

And so, the thought was there is suitable proof of variation in resource use in this condition, such that accurately measuring it and characterizing it would be an important activity.

With this, I think they have the individual comments in the packets. Okay.

So, there was one concern about this specific measure here that we are evaluating, which is cardiovascular condition, in how it is defined. And one of the members thought that it was slightly misleading because, on the one hand, it is cardiovascular conditions. On the other hand, how you are diagnosed with cardiovascular disease can vary widely, depending on which codes. So, I think the logical extremes of that were a patient with an MI was included in this as well as a
patient who had a carotid ultrasound and a
diagnosis consistent with cerebrovascular
disease, based on an asymptomatic carotid
ultrasound.

And obviously, the prognosis and
the associated resources used would be
expected to vary widely.

That being said, I think across the
TAP there was agreement that this was an
important measure, and it could combine or
consist of this wide variety of conditions.

So, leave that up for importance.

CO-CHAIR ROSENTHAL: Jeptha, would
you mind -- I know we had the two-minute
version from NCQA -- but would you mind just
quickly summarizing what it is that is being
measured, in what populations, and who it is
attributed to, just so it is clear?

MEMBER CURTIS: Right. Well, we
will get into a lot of that in the scientific
acceptability. But, to expand a little bit,
maybe, actually, the developer could expand a
little bit beyond the two minutes because you will probably do a better job than I would.

MR. HAMLIN: Okay. So, for the cardiovascular measure, primarily we are looking at their procedures or diagnosis of what we term ischemic vascular disease. There's a series of diagnosis codes over both the measurement year and the year prior. So, it is effectively a two-year identification of people with cardiovascular conditions.

Once they are in the measure denominator, if you will, that population is risk-adjusted and divided up into looking at their total utilization across a series of service categories for the measurement year alone. So, while it is a two-year denominator, we are only looking at resource use for the measurement year, which for us is a calendar year.

The primary procedures that we look at for identification are AMI, CABG, and PCI. The list of diagnoses for ischemic vascular
disease is fairly extensive, and I can certainly provide that list, if the Committee members are interested. But it is usually using ICD-9 diagnosis codes in the current structure.

CO-CHAIR ROSENTHAL: And the attribution is to --

MR. HAMLIN: To health plans.

CO-CHAIR ROSENTHAL: Health plans?

Okay.

MR. HAMLIN: It is a health plan population.

CO-CHAIR ROSENTHAL: A health plan population.

MEMBER CURTIS: And to expand on that, I think that was one of the major points of why this was more easily acceptable, is that there was no attempt to attribute to an individual physician. And they demonstrated that there was a minimal sample size of 400 patients, which they had arrived at through serial or sequential bootstrap analyses,
suggesting that they were getting relatively
stable estimates at that level.

CO-CHAIR ROSENTHAL: Well, we will
come back to the scientific things. I'm
sorry, I just thought it was useful to be sure
that everybody knew what the measure was and
what it tracks to. And I think what is open
for discussion, then, is the importance
question.

Bill and, then, Bill.

MEMBER B. RICH: It is the standard
question that you raise, Tom. The last time
we looked at a population-based measure for a
health plan, the discussion, then, devolved to
this actually it could be applied down to an
individual level and it had been. I believe
that was the Ingenix measure last time.

Has the measure developer made
clear that this is for a health plan
population-based measure?

MR. HAMLIN: Yes.

MEMBER B. RICH: And it will not be
used at the individual provider level?

MR. HAMLIN: We currently only use
this measure as a health-plan-level measure.
I am aware of several testing in some
physician groups. However, NCQA, currently,
at this time only uses this measure as a
health plan population-based-level measure.

MEMBER GOLDEN: Yes, I am looking
at the summary form for 1558. And the first
sentence is the summary: "This measure is
based on standard prices and includes all
costs for treating people with cardiovascular
conditions, whether they are related to the
condition or not."

So, help me understand what that
means. Does that mean -- I mean you say all
costs related for cardiovascular. Is that
only with the codes for cardiovascular
conditions or is that any disease they have
during that period?

CO-CHAIR ROSENTHAL: If I could
just for one second, Sally has reminded me we
are getting into the scientific part of the thing. And that was my fault. We really just need to vote, I think, or discuss the importance quickly, and, then, we can get to the scientific discussion.

MEMBER GOLDEN: Okay. I was just trying to understand what it was measuring.

CO-CHAIR ROSENTHAL: Well, I understand, but we will get to that in the scientific part.

Does anybody want to discuss the importance aspect of the measure? That is to say, cardiovascular disease and its importance.

MS. TURBYVILLE: Right. So, as a reminder, the way the importance looks at it, is the focus of this measurement area important to measure? So, is it important to look at resource use in a chronic cardiovascular area?

And, then, as far as the nuances of how the measure is constructed, and how that
is applied, that goes into the scientific acceptability. And typically, that conversation takes a lot longer because it gets to these nuances.

Now is the measure adequately addressing that important area to focus on? So, it is more the measurement area of focus, does it make sense? And, then, when we talk about the nuances of the measure, that goes into the scientific acceptability, the usability, and the feasibility.

So, is this area, chronic heart failure, an important, in your perspective, area to measure resource use?

CO-CHAIR ROSENTHAL: So, that is open for discussion.

MEMBER NEEDLEMAN: Just to reinforce what Bill was saying, this is not a measure of resource use for cardiovascular condition. It is a measure of resource use for people who have cardiovascular conditions. They get a cold and go to the doctor. Their
dollars are included here. They sprain their leg; they sprain their ankle. They go to an orthopod. Their dollars for that are included here.

So, I think in terms of importance, yes, it is important to know what cardiovascular disease costs. Is it important to know what the total resources are for this defined population? Is there enough information there to differentiate and make it useful?

CO-CHAIR ROSENTHAL: Well, as a point of order, it has certainly been suggested that the mechanism by which we would address these things was to deal with importance, vote; science, vote; usability, vote; feasibility, vote.

I am getting a suggestion that perhaps on this one we can't even consider the importance without understanding what it is in more detail. So, I think we would be happy to entertain either way to do it. If, in fact,
you would prefer to defer a discussion about
importance or a vote on discussion of
importance until we have had the scientific
discussion, we would be okay with that? Or do
we have to follow, do we have to vote on
importance?

DR. BURSTIN: Importance is a must-pass criterion. So, we wouldn't even move on
to the other criteria unless you guys think
this measure is important. And again,
important means that it is an important focus
area.

CO-CHAIR ROSENTHAL: Yes, but I
think what I am hearing from both Bill and
Jack is that --

MEMBER CURTIS: But that is the
specifics of how you are actually measuring
it. All we are talking about now, is
cardiovascular disease important and is there
variation in the use of resources in
cardiovascular disease? To me, that's done,
right?
MEMBER NEEDLEMAN: Cardiovascular disease is important. Is it important to have a measure which includes colds, sprained ankles --

MEMBER CURTIS: That's the specifics of the measure. I mean I really think you --

MEMBER NEEDLEMAN: -- for that population.

MEMBER CURTIS: If we are going to get very detailed in this, if we can't move beyond importance in five minutes, then we are never going to get through today. I mean, trust me, I've been down this road.

(Laughter.)

We've got to keep moving.

MEMBER B. RICH: I would move that my heart is important.

(Laughter.)

CO-CHAIR ROSENTHAL: Okay. So, does anybody want to discuss importance outside of the context of the science of this?
I am hearing what you guys are saying loud and clear. But if the question is posed as it is, which it isn't the case of whether this measure is important, it is the question of is this subject matter important, does somebody want to discuss that point or call the question?

Paul?

MEMBER BARNETT: I think it is too bad, well, I think that we are going to find out by the end of the two days that we are going to endorse all of these as being important, and that we ought to just skip that and just take it as a given.

(Laughter.)

So, I move the question on this one, and all of them, in fact.

CO-CHAIR ROSENTHAL: Okay. All right. So, we can move -- thank you, Jeff. So, the question has been called. There is further discussion.

So, Ashlie, do you want to describe
how we are going to vote?

            MS. WILBON: Yes. I think we have
switched the screen in front of you, so you
guys can only see -- sorry -- the Co-Chairs.
If you look on the righthand screen, that is
the voting slide. And when we hit Start,
there will be a timer that will start at 60
seconds. So, everyone will have one minute to
enter their vote.

            If you hit one, it is, yes, you
think it is important or two means, no, you
don't think it is important. Point towards
Sarah. And I believe you have to hit Send
after you hit your number. And that's it.

            So, if everyone is ready, we will
go ahead and start the timer. Ready? All
right, let's go.

            (Whereupon, a vote was taken.)

            DR. BURSTIN: Keep in mind, point
towards Sarah, not the screen. People
routinely point in a strange direction.

            So far, nobody has responded. Is
anybody pushing? This isn't working.

(Pause.)

CO-CHAIR ROSENTHAL: Would you allow us to raise our hands?

DR. BURSTIN: Yes, please.

CO-CHAIR ROSENTHAL: I do think No. 4 is relevant, though, to the question posed. I mean there are four subcategories, and I understand we are not voting on subcategories. But subcategory 4 says that the resource use service category is consistent or representative, and that gets to, I think, exactly the point that several of you were making.

But I do think we are calling the question in aggregate, correct?

DR. BURSTIN: Yes.

CO-CHAIR ROSENTHAL: And so, I think we will have to do this by a show of hands.

DR. BURSTIN: Yes.

MS. WILBON: And take into
consideration that the TAP, you know, they have already done a deep dive on each of these, which is why we projected the results here.

CO-CHAIR ROSENTHAL: Well, how do we interpret the TAP results? The first two categories are all blue, and as you get over to the fourth category, obviously, the TAP had some of the same questions that we did. How should we interpret their overall score? As a thumbs-up?

MEMBER CURTIS: I think that the medium here -- Sally, correct me if I'm wrong -- but, as I recall, that has more to do with the types of resources that are being measured. So, you're right, it does stretch, overlap. Well, then, I take it back.

CO-CHAIR ROSENTHAL: But, generally, if the TAP -- how would we interpret this? This was an affirmative --

MEMBER CURTIS: The TAP wasn't addressing that when they answered the
question. They weren't addressing this all
resource use --

CO-CHAIR ROSENTHAL: But this is an
affirmative vote from the TAP?

MEMBER CURTIS: Yes.

CO-CHAIR ROSENTHAL: Okay. All
right. So, the question is called. We'll do
it by a show of hands.

All who want to vote in favor of
this being an important measure? Overall. We
are not voting on them individually.

So, all in favor --

MEMBER GOLDEN: Point of order.

CO-CHAIR ROSENTHAL: Okay. All
right. Yes?

MEMBER GOLDEN: All right. There
is a difference in question here. You have
asked two different questions.

Is the subject matter important,
yes or no, is one question. Is the measure
important, yes or no, is a different question.

DR. BURSTIN: The subcriteria under
importance to measure and report are listed there. The TAPs have done a deep dive for you on every single subcriteria and rated each subcriteria.

Our view of the Steering Committee is you are going up a level. You have the information from the TAP. You now need to look overall and make an assessment overall of does it meet importance to measure and report. And keeping mind, you won't even discuss the measure further if you don't think it meets importance to measure and report.

CO-CHAIR ROSENTHAL: Do you want to weigh-in one more --

MEMBER GOLDEN: But the slide and the question you are asking us to vote on was different than the discussion we had a little while ago.

MS. WILBON: It is actually about the focus area, not the measure.

MS. TURBYVILLE: That's how we framed it for this.
MEMBER GOLDEN: That's not what the slide says.

MS. TURBYVILLE: So, the subcriteria actually mapped to how the measure developer responded to the submission form. It is a very detailed review.

So, did they feel that the measure developer clearly identified what the purpose of this measurement area is? It doesn't get into the details of the measure. It is like four sentences. Does it align with, you know -- so, are they saying on measuring cardio, but, then, in the purpose they said we want to see what diabetes looks like.

So, it is still at this, is it an important area to measure? The subcriteria are a deep dive, but we framed it the same for the TAPs, that we are talking about the focus area, is it important to measure? But, then, when they are looking at it, they are also looking at the submission to make sure it is adequately mapping to what that importance
area is.

So, there's a lot of moving pieces that go on when in the TAP they are looking at it, which include: is this submission complete? Is the purpose clearly stated as far as we are trying to measure cardiovascular or chronic disease?

It is the detailed underpinnings of the measure. They are not looking at whether or not it is actually meeting its purpose at this point. That is in scientific acceptability.

CO-CHAIR ROSENTHAL: Well, if we vote no on the importance, we don't even get to discuss the scientific acceptability. And I think there's some virtue in discussing these various scientific issues because it will lead us to some avenues in some of the others that I think will be useful.

So, we can carry on the conversation, but, as Jeptha says, if we spend an hour talking about the importance on the
first one, it is going to be a very long two
days.

I would like to suggest that we go
ahead with the vote. And you can vote your
conscience, but I think the way Sally has
described it is the way that we should be
thinking, then, about the importance of the
measure.

All in favor of importance raise
your hand.

(Show of hands.)

Ashlie count.

Do you have the count?

All opposed?

(Show of hands.)

I see one opposed.

Abstain?

(Show of hands.)

One abstention.

Duly noted.

All right. Now we can discuss the
scientific acceptability. Now the fun begins.
So, Jeptha, do you want to give the TAP report on the scientific acceptability?
And, then we will get into, obviously, the various issues.

MEMBER CURTIS: You know, I am trying to think about how to summarize because I have a good memory, but it is hard to keep 14 separate measures in my head.

I think, broadly speaking, this is a measure that tries to capture patient cardiovascular disease using the specific codes that the measure developer referred to. There are some specific exclusion criteria that are worth considering which generally adhere to, I guess, the HEDIS measures of exclusion of end-stage renal disease patients and HIV patients, other patients in whom it would be expected that cost would not necessarily -- well, an attempt to make a more standardized population.

I think that was one of the big components that we discussed at the TAP.
There were some people who felt very comfortable with that. There were people who didn't feel as comfortable with that.

Other points of scientific acceptability that warranted discussion: the major one, and I think maybe we should just stop after I talk about it, is this is all resource use in an identified group. That is the single biggest assumption that this makes, is that that is a valid way of assessing resource use.

There was a great deal of comfort in this approach in comparison to the alternative approaches that we evaluated with other measures, in that it really didn't attempt to parse out, well, this office visit was associated with cardiovascular disease and this one was not. And the sense was that that was a much simpler way to go, but it does carry with it some consequences that we discussed, but I think, on balance, felt was a valid approach.
Yes, Bill?

MEMBER B. RICH: I think if, indeed, this is a population-based measure, I think that is true. But if we go back to the last population-based measure that we looked at, I think Jack raised the issue of someone with a malignant melanoma with cardiovascular disease. That would be funneled in here.

As long as it is clear, and we make clear, that this is a population-based measure and not for attribution to an individual level, but we have heard the last time that, yes, it is being used, a population-based measure for individual. So, someone is going to get stuck with it. Some cardiologist or internist is going to stuck with that malignant melanoma.

And we have had a hint here from the developer that said, well, we have some reports that it is used. But I think we have to make very clear, to emphasize what Jack and Bill pointed out, that this is only a
population-based measure. We can't control
how people use it, but that is how we have
gotten into trouble before, when we have a
circular definition of a measure and, then, we
take it down to a different level than it was
intended or designed.

MEMBER CURTIS: And to specify,
with the TAP, though, we did ask that
clarification be made, and that we are
explicitly endorsing a measure that would be
used per their application at the service
level and not be attributed to individual --

MEMBER B. RICH: And that should be
part of our minutes of this discussion.

DR. BURSTIN: And just to be clear,
the measures are endorsed for specific levels
of analysis. The measure has only been
submitted at the health plan level. I think
what we just heard Ben saying is that there is
some testing going on at the physician group
level. It is not endorsed at that level.

Should they come back at a later date with
testing at that level, we would consider that.
At this point, that is the only thing before
you, is what has been tested and submitted.

Do I have that right, Ben?

MR. HAMLIN: That is absolutely
correct. And even in the testing at the
physician group level, it is still a
population-based measure for your physician
care team. It is not an individual physician-
level. These cannot be used, be attributed to
an individual physician.

CO-CHAIR ROSENTHAL: Well, I just
want to be clear, at least in the piece of
paper that I am still referring to, unless it
hasn't been updated on page 25, the level
analysis says clinician group practice health
plan integrated delivery system, national and
regional. So, this is talking, at least the
piece of paper says group practice clinicians.

MEMBER CURTIS: There are
inconsistencies within the application.

Another place it says this is at the payer
level. That is why we asked for that clarification.

CO-CHAIR ROSENTHAL: Well, I'm with Bill. So, the piece of paper isn't correct? That's fine. Okay.

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: Other discussion?

MEMBER B. RICH: I assume that adjustment, that correction will be made in the submittal. Thank you.

MEMBER REDFLEARN: It seems to me that this measure complements the more specific episode-based methodology in which you have an episode of CVS or heart disease and you relate costs specifically back to that definition of that episode. That is a different, much more focused view of it.

So, I wonder, do we take into consideration the fact that we have kind of complementary measures that look at the same kind of condition from two different ways, one
globally, all services that are involved, and
another that is very focused on a definition
of that particular disease state?

CO-CHAIR ROSENTHAL: Bill? Now you
can weigh in on this.

(Laughter.)

Now it's okay.

MEMBER GOLDEN: Well, that's okay.

I have a question for the developer. Since
you deal with different plans, does this
measure perform differently if it is a
Medicaid HMO versus a commercial HMO?

MR. HAMLIN: Yes, I mean we only
calculate and compare these measures Medicaid
to Medicaid, commercial to commercial, and
Medicare to Medicare only. For this
particular measure, there is a broad
distribution of resource utilization within
all three of those different categories So,
we don't compare a Medicare plan to a Medicaid
plan or a Medicaid plan to a commercial plan.

MEMBER GOLDEN: So, are you saying
that one other limitation of this measure is
that it has to be compared across
socioeconomic population groups? You cannot
use it as a generic? You have to first define
your socioeconomic group before you can define
how you can compare data?

MR. HAMLIN: Well, we don't include
socioeconomic status as part of the measure
strata because the data is not available in
the --

MEMBER GOLDEN: But I am assuming
that a Medicaid group is different than a
commercial group.

MR. HAMLIN: Which is why they are
only currently reported by Medicaid only, a
Medicaid plan is only compared to a Medicaid
plan at the current time. The methodology
will allow a calculation of combined plans, if
that is what you would intend to do. But,
right now, we are only holding it, because of
the differences in the --

MEMBER GOLDEN: But that is a
limitation. But you are saying there is a
limitation in how you can compare activities,
depending on what the populations are in that
group?

MR. HAMLIN: I mean we have held, again, to our -- you know, we are only
comparing like plans in these population levels.

CO-CHAIR ROSENTHAL: And I missed that in reading through the thing. Is that
specified in the material, that, in fact, the comparator groups are only --

MR. HAMLIN: Right.

CO-CHAIR ROSENTHAL: -- like paired --

MR. HAMLIN: So, commercial,

Medicare, Medicaid, HMO, PPO are all only
compared to like plans for purposes of
reporting this information.

MEMBER NEEDLEMAN: Okay. I need

some help from the clinicians in the room.

Clearly, we've got a population with a disease
we care about and it is an expensive one to
treat. I am trying to understand the
rationale for looking at the total resource
use in this population, what we learn from
that, why that is important to look at for a
subpopulation.

So, I can see a number of possible
reasons. One is that cardiovascular disease
kind of colors whatever is being done to a
patient, regardless of what else they are in
the room for. Or the cardiovascular disease
dominates their payment, so the resource use
here is principally about cardiovascular
disease.

So, I want to understand to what
extent we think that is going on, that this is
dominantly a measure of cardiovascular disease
use, the extent to which we think having heart
disease colors the way clinicians deal with
other kinds of illnesses the patient is
bringing into the office or the hospital or
the emergency department.
And, then, with regard to that, we have got other serious conditions that these patients may also have. How well is the risk adjustment, when you looked at the risk-adjustment methodology, how well did that do in taking into account there are other conditions that will also color the way treatment decisions are made and resources are used, when a patient comes in for an unrelated condition?

MEMBER CURTIS: So, I think that, ideally -- and this is sort of trying to reflect what was the discussion in the TAP -- I think if there were a reasonable and validated alternative, such that we could break out only the cardiovascular-disease-related costs, that would be better. But, on reviewing at least three different methodologies, two of which tried to do that and one of which did not, I think the thought from the TAP was that this was a stronger methodology with increased noise, but, also,
less risk of making incorrect inferences.

MR. HAMLIN: And I think one of the things that would be important to understand for these measures is the results are only reported with their quality results. So, what we are looking at is effectively the value of healthcare provided. So, the resources used for a defined population that correlates with their quality score, and these two items are reported together. So, it is looking at the utilization of this population over a year and the quality that is achieved for that same population. So, it is how they achieved that quality score effectively by looking at these different utilization categories.

And that is the approach that we have taken. So, we are only reporting these results for RCA with the corresponding quality measures that are derived from HEDIS.

MEMBER HALM: This is Ethan.

The thing that I found confusing was sort of lumping the apples-and-oranges
decision. So, I can understand the patients with acute coronary syndromes or MIs or bypass surgery or stenting, that that is one group. But some of these codes include peripheral vascular disease, just as a diagnosis, or you mentioned someone who gets a neck ultrasound and gets described as having asymptomatic carotid disease.

When you are saying in the context of other quality measures, you know, the quality measures for treating MI are different than the quality measures that don't exist for treating asymptomatic carotid disease or other potentially sort of incidentally-related vascular disease in the body. I struggle with that a little bit.

MEMBER B. RICH: I think it is two. One is, if I am working for WellPoint, I like this in helping to figure out my premiums for groups. Also, it is valid for a health services resource where you want to look at associations, for instance, cardiovascular or
diabetes and diabetic retinopathy, and things like that.

But, again, if you look at the CMS's ATC criteria, they have specifically moved from a population down to a resource group, but what we are seeing is a dangerous trend as long as this isn't used at the individual physician or group level. And unfortunately, we have heard that it might be.

So, I don't mind this measure as long as it stays as a population-based measure. I think it has merit for a health service resource and for plans for premiums, but not for attribution to a doctor or group.

MEMBER CURTIS: So, again, I think they have made the decision. We can argue about whether or not it is the right decision. I think that will be reflected in the voting.

Some of the other thoughts from the TAP that we have recorded on these measures: there is an exclusion of age greater than 75; patients over the age of 75 are excluded from
the measure, which was not, I think, particularly well-justified in the application. Or I can't remember exactly the justification. There was some concern about why that was done and whether or not it was appropriate.

The other issues to be aware of is that to be included in the measure does require a continuous enrollment for, I believe, two years. There is an identification year and then there is the actual measurement year. And so, that does limit. That is one of the sources, the biggest sources of exclusion criteria within the population. It gets you down to a much smaller number.

And I think the other major thing that we considered was in the risk adjustment. So, it does use HCCs for risk adjustment, which the TAP felt fairly comfortable with as a validated methodology.

One point that was worth
considering is that -- and maybe the measure developer can follow up on this -- the risk adjustment takes into account resource use within the measurement year for risk adjustment. So, it is not in the year prior exclusively. It is taking into account the resource use within the year in providing those results.

And that is very different than I think the approach that is generally taken for quality metrics. And so, on a personal level, that made me feel uncomfortable with the risk adjustment. I don't know if that is a valid approach in resource use. And maybe that is something worth discussing.

If you can to follow up on that,

that would be --

MR. HAMLIN: Yes. So, I think it is an important distinction to understand how people are assigned to the particular HCC categories. And that is using the entire two-year algorithm timeframe for that. So, again,
looking for diagnosis of IVD over the year and
the year prior. We are only, however,
measuring the resource use in a single year,
which is the measurement year.

So, we are using effectively a two-
year algorithm of multiple diagnoses and
encounters, and so on and so forth, to get
people into the appropriate cohort for risk
adjustment. However, we are only tracking
their actual resource use during the one-year
timeframe.

CO-CHAIR ROSENTHAL: Can I ask a
followon question in relationship to that
point? Some of these index events have a very
high initial cost and, then, it spreads out
over time. Does the fact that you take this
two-year window, if we are now in 2012 and the
event was in January 2011, and compared to
somebody whose event was in January 2012, the
one who is in the year of attribution is going
to have a very high triggering cost where the
one that happened the year previously is going
to have that washed out. Is that accounted
for in the methodology?

First of all, is it a correct
assumption that I am making that there is this
high index cost, which I think there is. But
do you have a method for accounting for that?

MR. HAMLIN: The high index costs
generally are around some of the procedures.
So, CABG, obviously, is a very high index
cost. However, that is only used as an
identification, and it is only CABGs performed
in the year prior.

So, the actual measurement year is
not looking at CABGs because that only gets
you in the criteria if you have had one in the
year prior to the measurement year. So, what
we are looking at is the cost associated with
someone identified as cardiovascular disease
because they have had a CABG the year prior.
We are only looking at their encounters, and
their followup visits effectively or any
other -- obviously, if they had a second CABG
in the measurement, that would be a second spike. But the AMI, CABG, and PCI events, to get into the denominator, are only the year prior. It is only January through November of the year prior. Ischemic vascular disease diagnoses are the year prior and the measurement year. So, again, I think balancing out some of that --

CO-CHAIR ROSENTHAL: And, Jeptha, were your questions answered about the risk adjustment?

MEMBER CURTIS: I guess it is a larger question for the group as maybe opposed to the developer. Is that reasonable to adjust for things that are happening during the measurement year?

MEMBER HALM: Before we get to that -- this is Ethan -- I was also puzzled by this. This seems like to me anti-bundle or anti-episode-of-care approach. You have got people with MIs, you know, stents, acute coronary syndromes, bypass surgery. That is
where all the cardiovascular costs are. And
so, you are identifying those people, but,
then, you are saying you are not looking at
the year in which all of the money is being
spent to treat their cardiovascular disease.
You are seeing what happens the year after
that. If found that very puzzling.

MR. HAMLIN: So, I mean, again,
when we define our eligible population to try
to track resource use for a predefined chronic
condition, we really stuck with the HEDIS
criteria. So, this eligible population is
what we used to identify cardiovascular
conditions in the HEDIS quality measure
population.

I would agree that these sentinel
events, if you will, a CABG to get someone in
this population, is a rather high-cost
condition. But, again, we are looking at
overall utilization for an identified chronic
condition. And so, I think by avoiding a lot
of sentinel events that might, in a small
population, that might spike versus sort of a broader cardiovascular at-risk population would provide a little bit more balance. I mean, obviously, there are some high-cost events that do occur during the measurement year, of course, for this population. But, again, sort of in the overall large population-based approach, we feel this is the best way to try to track utilization and map that to the quality scores.

CO-CHAIR ROSENTHAL: This certainly will get into the utilization questions, but let's try keeping it in the scientific realm. But several of the clinician types have opined that they would be very uncomfortable with this being a physician- or a group-level measurement, for a variety of reasons. But let me pose a question back to several of the health plan folks that are here. Is it your sense that in your health plan either (a) the risk-adjusting
methodology is adequate to wash out the
potential spike of some, coming back to the
melanoma thing, that your health plan doesn't
have 27 melanomas in it, and if it does, it is
accounted for by the risk-adjusting
methodology. So that, when NCQA says that
your health plan gets the same 100 percent of
the HEDIS measurements that everybody else
does, but your cost is 50 percent higher than
Blue Cross of Maine, is that, given the
methodology that you have seen, going to hold
water? This gets to the face validity of the
thing.

So, the clinicians have weighed-in
and said, face validity, probably not so at
the physician level. We are just all
assuming, well, no problem at the health plan
level. How about some health plan folks
giving us your sense of, does it play out with
face validity at the health plan level, is I
could be so bold as to sort of ask you that.
Because, otherwise, the group is now beyond
its potential, its ability to weigh-in on the
question of face validity.

MEMBER O'NEILL: So, you are
asking, if two health plans were evaluated
based on this metric in terms of the cost for
my health plan to take care of this population
of patients versus the cost of another health
plan to take care of this group of patients?

You know, we haven't really looked
at things that way very much. Maybe there has
been more in like the managed-care Medicaid
populations or things like that. There have
been more comparative data. But there are so
many variables in how we cover things in
benefit design, co-pay, contracted rates.

I mean, first of all, this whole
idea of a standardized payment doesn't make
any sense to us whatsoever. So, I am a long
ways away from understanding how we would use
this in that fashion.

CO-CHAIR ROSENTHAL: I appreciate
that is going to get to the usability
question.

MEMBER O'NEILL: Yes.

CO-CHAIR ROSENTHAL: And that's why I struggled raising it at this point.

MEMBER O'NEILL: Yes.

CO-CHAIR ROSENTHAL: Except it seems to me it has a lot to do with the scientific validity. Because if you believe that it is scientifically-valid --

MEMBER O'NEILL: Then we can use it.

CO-CHAIR ROSENTHAL: -- then you would use it, I would presume.

MEMBER O'NEILL: Right.

CO-CHAIR ROSENTHAL: So, I think that the question does devolve back to, do you believe that it would have face validity for your population, if you were comparing your health plan to the health plan in northern Maine --

MEMBER O'NEILL: Right.

CO-CHAIR ROSENTHAL: -- or in
southern Florida, et cetera, these various
issues about having the whole cost of care
with the risk-adjusting methodology that they
have proposed?

MEMBER REDFEARN: My concern is
about I don't have a heck of a lot of faith in
the risk-adjustment methodology. And I am not
picking on this measure. I think I wouldn't
have any faith in any of them to adjust away
a lot of this kind of variability.

My concern about the measure is,
what do you do when all of the cost
variability is associated with characteristics
that are not directly related to the
underlying condition, either the accidental
stuff or the stuff that is kind of peripheral?
And you can't adjust that away. I think that
is an interpretation issue that you end up
with.

I mean, what do you make of those
differences? You call it a cardiovascular
measure, but all the variability are things
that are very distant from cardiovascular
disease. What do you make of that? And that
is my concern.

CO-CHAIR ROSENTHAL: That question
was sort of answered in a way, or at least
addressed, in a sense of the grouping, that
this isn't really a cardiovascular condition.
We have treated it like that and it got sent
to this TAP, but it really is a population
measure.

And I think to view it any other,
I resonated with the people that said, okay,
if we think of it as a population measure,
maybe. Because it, clearly, in my head isn't
a disease-specific measure because a vast
preponderance of the variability is going to
be related to a variety of other things.

MEMBER REDFEARN: But it is labeled
that way.

CO-CHAIR ROSENTHAL: Does anybody
else from another health plan want to weigh-in
on this? Or is Mary Kay going to be the
spokesperson?

And I wasn't quite sure what your answer was at the end of the day.

MEMBER O'NEILL: We haven't looked at subpopulations, our relative efficiency of management of subpopulations compared to other plants. I have never seen it sliced and diced in that fashion.

Now, if we are going into exchanges and stuff like that, this may be our new world. But we haven't done this historically. So, I am trying to figure out how this would work. I mean we are looking at, I mean the complexity of our world in terms of how we are doing things and who controls what variables is very high.

For example, in my company 85 percent of our customers, the individuals that we are the carrier for, are covered by self-insured plans. And so, that means the finances, the benefit structure, all kinds of things are the decisions of the employer.
So, when you say, does CIGNA do "X", I say, well, it depends. You know, so we have 15 percent fully insured, and we are working in every state jurisdiction for that group of people. We have unique contracts in every single market. So, it gets very difficult to say, you know, what we are doing.

We also have two major levels of medical management that are products that we sell. We have wellness. I mean the complexity of our world, to say that we can tell you how we manage a cohort of patients in the entirety of the 13 million lives that we have in this country is just not standard. It depends on benefit design, not even just the financial aspect of benefit design.

And we also have companies that tell us, "Don't call our folks, even if you know something is going south, because they don't want phone calls." And those are our clients.

So, that's my answer, is that in my
world we have a whole set of complexity that
is equivalent to the rest of the sets of
complexity that everybody else is working on.

CO-CHAIR ROSENTHAL: But if this
measure were adopted, would you view it as
scientifically-accurate, acceptability? Well,
accurate.

We will wait until we vote. You
know, we don't have to do the thing.

MEMBER O'NEILL: Okay.

CO-CHAIR ROSENTHAL: Jeptha, you
were trying to get a word in edgewise?

MEMBER CURTIS: I mean I think it
is interesting how the conversation is
evolving. It is a little different than the
way that the TAP evolved.

I think that there was concern in
the TAP that this all resource use part was
introducing noise, but I don't think you can
say that it is not a cardiovascular measure.
The things that are driving costs to a large
stress testing post-MI or post-PCI.

Sort of there's a lot of elements that are going on at the clinical level that are going to be driving costs that are directly related to cardiovascular. I think there is noise. There is going to be the melanomas. There are going to be the outliers.

But I actually appreciated the way they kind of got around that by, again, some statistical analyses to figure out what was the minimal population number at which you started to get a stable result. And I think that, actually, was very reassuring in the sense that, at around 400 or 600, whatever the specific level was, it didn't matter what group of patients you were identifying, within a plan you started to get a signal, right? And I think that that was a key thing for me in terms of feeling more comfortable with the signal that you are getting out of that is actually representing something more than the
noise, and probably honing-in on something that is cardiac, in my opinion.

MEMBER HALM: This is Ethan.

Another conversation is that what you are actually measuring there is just sort of baseline care for non-cardiac things in older, in adults up to 75.

I mean imagine if you were doing crisis resources measures for the bone grid now, and we are talking about management of patients with hip fracture or a bad knee or hip arthritis, and we are going to use the same identification. And we are going to say, well, in the year that you did or didn't get your knee repair or your hip repair, we are not going to include those costs. We are going to look at the costs in the year after or we are only interested in the costs in the year after your transplant.

It seems to me that we are missing the vast majority of the action in the variations and how aggressively people use
resources or not to manage, you know, to find sentinel incidents.

MR. HAMLIN: So, I think one of the critical things to understand about this measure is it is not a single result. The measures are reported out by service category very specifically. And for this measure as well, there are a series of frequency of services that are reported alongside the measure. So, we are capturing some of the procedures, endarterectomy screening, carotids, along those kind of lines.

But we have very detailed information that is reported out. So, inpatient and outpatient surgery and procedures are separate service categories that are reported out for each plan that meet the criteria for this measure. Inpatient and outpatient --

MEMBER HALM: In year two.

MR. HAMLIN: For the measurement year, yes. For the calendar year that we are
calling the measurement year, which is --

MEMBER HALM: But even that is sort of the horse-out-of-the-barn year.

MR. HAMLIN: Well, that is the year in which we are comparing one plan's utilization to another's effectively, by each of these service categories. So that, the measure breaks down utilization. So, even if your total resource use result looks high, you can then dive down into the specific service categories and understand particularly what is driving that by looking at the individual service categories.

MEMBER HALM: Yes, I don't want to dominate this. I guess what I am suggesting is I think you actually are losing the vast majority of the variations by focusing on the year after all the action or not.

MEMBER HENDRICH: Well, I mean we are tracking, like I said, the service frequency for high-frequency procedures in this population that would probably capture
some of that. Are they getting some repeat procedures? But, again, this is a multi-year population that we are looking at with chronic conditions. And so, they do tend to be managed on a regular basis one year and two.

So, it is a measure we don't lose a lot of plans because of a lack of a 400 population or less. It is one measure we actually have very little problem with continuous enrollment with the size of the population.

MEMBER HALM: And I guess this could be empirically answered, and maybe the developers have done this. But if you looked at the degree of variation as a spread in the year one utilization compared to the year two utilization, which you are defining as the measurement year, it would give you your answer as to whether or not you are really missing a lot or not.

MR. HAMLIN: Well, we do compare the results year to year, but, again, we are
using the year prior procedures, CABG, PCI, and AMI, as identifiers to get people into the population for the risk-adjustment approach.

The resource utilization we do actually track and compare year to year to year. I mean these measures have been in use and reported for about five years now. And so, we do an annual analysis to look at the changes in utilization between plans, between products, year to year to year, the number of plans that report the information. I mean I think we have provided you with that, last year's analysis report that was released in January.

MEMBER HALM: Okay.

MEMBER CURTIS: Let me just follow up, though.

CO-CHAIR ROSENTHAL: Go ahead, Jeptha.

MEMBER CURTIS: Ethan, what you are proposing is really more of an acute episode-based measure, right? In which case the
fundamental assumption that they are combining all these different conditions would be fatally flawed.

So, I think it is something of a misnomer, I guess, in the title, and maybe a clarification of the title would be that you are trying to get a chronic cardiovascular population. And I think if you take that as your point of reference, then these decisions make a lot more sense as opposed to we are missing everything that happened when they had their MI or they had their index CABG.

CO-CHAIR ROSENTHAL: Paul, you have been very patient.

MEMBER BARNETT: Yes, just people keep raising that example of the person with a melanoma. So, reading this, I think they would have excluded anyone with active cancer in the measurement year or HIV or organ transplantation. They wouldn't be included in the measure. So, that particular example is not right.
MEMBER RUDOLPH: Yes. If you think about from, for example, the employer's perspective, you can't really control when that first AMI hits, but what you would like to be able to control are the costs associate with the care, the long-term care after that of that employee.

So, this measure makes sense because, at least in my experiences, patients who have one of these serious cardiac events, the care is really managed by the cardiologist, almost even into the primary care arena. So that it makes sense to have this focus on the cardiovascular kind of conditions and incorporate all the other care that is associated with it.

CO-CHAIR ROSENTHAL: Bill?

MEMBER B. RICH: But, again, we are looking then at an attribution issue, Barbara, because I agree with Jeff that it is very reassuring. If you look at an "N" of 400, you know, it looks fairly stable. The typical
internist has 2,000 patients. About 30
percent are cardiovascular disease. You can
the math.

So, it depends on the size of the
group. So, that would be your goal as an
employer, but you can't --

MEMBER RUDOLPH: No, actually, I
was talking about how the plan -- so, in my
determination, if I am an employer and I am
looking at choices between plans, this is
exactly the kind of information I would want
to know. Are they managing chronic
populations well in terms of resource use and
quality? Obviously, the charts that they
included in the document showing both of those
and where those plans fell on that plot would
be of high interest to me.

MEMBER B. RICH: And again, as long
as it stays as a population or a plan thing,
then it is okay, but --

MEMBER RUDOLPH: But that's what
this measure is. It is a plan population.
CO-CHAIR ROSENTHAL: I am going to do just a little process check with the group. It is five of 11:00. We were due to spend one hour on this measure. We are 55 minutes into it. It is a spirited conversation, and I am sort of checking with our bosses. Can we let this go a little bit longer in the interest of sort of hammering this out? Or are we beginning to get repetitive and that we should maybe move on to use and feasibility?

MS. TURBYVILLE: I think that is your call.

(Laughter.)

If you would reach out to your Committee members and see if there is anything new to add, clearly --

CO-CHAIR ROSENTHAL: All right. And one other thing I do want to do, we have a statistical analysis of the thing, and I think it is worth hearing from that independently. So, maybe, unless there is an objection, Bill, maybe sort of last comment.
And, then, we will ask for the statistician --

MEMBER GOLDEN: Okay. I am trying
to get us up to 30,000 feet from wherever we
are right now, but we are at a low altitude.

(Laughter.)

When all is said and done -- and I
will go back to this slide No. 40 you showed
earlier, which was the arrows about
accountability and transparency. I could live
with this measure, but it is on the left side
of the slide, not on the right side of the
slide.

So, I guess, as we go through this
exercise, how far to the right side of the
slide do we have to be to endorse a measure?
That is sort of a 30,000-foot -- that might
save a lot of time and energy because we won't
have to find a lot of things. If we realize
that this thing is not going to get too far
across the middle of the slide, that may not
be sufficient.

CO-CHAIR ROSENTHAL: Well, that's
philosophical, and I guess everybody will have
to put that in their conscience.

MEMBER GOLDEN: No, I mean --

CO-CHAIR ROSENTHAL: All right.

Carlos, do you want to give us your quick
assessment of this?

MR. ALZOLA: Yes. I tend to be on
the same side as Jeff in that we are looking
at a measure that is aimed at cardiovascular
patients. It is probably true that there is
a lot more noise that you would see if you
just restricted yourself to the
cardiovascular-related costs.

But that has issues in itself
because how you attribute those costs to a
cardiovascular episode, it has some issues.
So, they do this approach; thus, make the
measure a little more clean in that respect.

And from the point of view of the
health plan, which is what they are interested
in to know, what is the cost of treating these
kind of patients? Whether the costs are
cardiovascular-related or not, it doesn't matter that much because we are going to have to pay for it anyway.

       So, the other issue is, does that noise, additional noise, really impact the measure that much? And again, the sample size requirements that they have shown show that the standardization really stabilized after 400 patients.

       CO-CHAIR ROSENTHAL: So, there are no statistical red flags --

       MR. ALZOLA: I don't see any statistical issues.

       CO-CHAIR ROSENTHAL: -- that you see at all?

       And can I just clarify one thing? These were all done with standardized prices on the various units of things, so when you are comparing one part of the country to another.

       Jack?

       MEMBER NEEDLEMAN: Yes, I just have
a question for the developer for clarification around the standardized pricing.

Can you speak a little bit about how the standardized pricing algorithm is applied to inpatient care?

MR. HAMLIN: Yes. Ingenix is the company that helps us with our standardized pricing approach. Our approach is based on the National Medicare Fee Schedule, but we do make certain adjustments based in the codes that are included in our standardized pricing tables.

So, for example, we make several relative adjustments based on inpatient and outpatient. So, for example, on the inpatient side, actually, if you look at the procedural codes, the price is actually lower because on the outpatient side we include a facility charge in that because that is the way it shows up in the claims.

So, again, the way the units of service are defined is down at the coding
level. And each code is priced using, you know, you apply a price to that code based on what is available in the standardized pricing table.

We price about, right now, about 80 to 82 percent of the services, and the approach is detailed in vast detail in the documents that were provided. Effectively, what you do is you scan for all services rendered and, then, you map each of those codes to a standardized pricing table and, then, use those to inform -- multiplied by the units of service and, then, use those, apply those to each of the applicable standard service categories.

MEMBER NEEDLEMAN: So, just to clarify, when you are looking at an inpatient bill, you literally take all the charges that are on the inpatient bill and you standardize price them?

MR. HAMLIN: And you price the ones that, if you are on the standardized pricing
tables, yes.

MEMBER NEEDLEMAN: So, if there is longer length of stay, that is going to be taken into account.

MR. HAMLIN: Yes.

MEMBER NEEDLEMAN: If they make more use of radiology in this hospital, that is going to be taken into account?

MR. HAMLIN: Yes. For inpatient, we currently report out on days, discharges, and average length of stay for each of the individual inpatient service categories.

CO-CHAIR ROSENTHAL: All right. I would like to suggest that, unless there is some compelling unanswered question regarding the scientific validity, that it perhaps is time to put hands up or click the clickers.

Paul, is this urgent?

MEMBER BARNETT: Yes, I think so. So, there is just one thing that occurs to me that makes me a little bit uncomfortable about this. It is the idea that the people qualify
to be in this group based on a procedure being
done. Did they get revascularized in the
prior year? So, I am sort of backing up what
Ethan is saying. That is the real issue.

So, if some provider does lots of
PCIs because they have a very low threshold,
then they end up with a group that is very
much healthier than some other provider that
may have a more conservative management
strategy. And maybe the case mix controls for
that, but that worries me a little bit.

Really a lot, yes, it worries me a lot.

MR. HAMLIN: So, you would see
those results appear in this measure, both
under either the inpatient or outpatient
surgery and procedures, but, also, PCI is one
of the frequency-of-services procedures that
appear reported out in this measure. So, you
would be able to drill down and find out if
those PCIs were, in fact, driving the result.

MEMBER BARNETT: But the way they
get into the cohort is by having had a PCI in
the prior year, right? And so, it is the question of who's in this chronic disease group, whether it is people who have had an AMI, CABG, or PCI.

So, if a provider has a very low threshold for doing revascularization, they are going to get a lot of people who otherwise are kind of healthy into their cohort. They are going to look like they have low costs in the measurement year. And actually, they are the high-cost provider.

CO-CHAIR ROSENTHAL: All right. Again, unless there is a burning question, and again, I think our charge here is thumbs-up or thumbs-down, correct? I mean we are not taking each of the six scientific submeasures and voting on them independently.

And just to quickly review, there was a point of order in terms of -- okay.

MEMBER B. RICH: Before we vote, could we have another test on the electronic voting thing? I think that would be --
MS. FANTA: I think we have it working, but that is a good idea. Let's try it now briefly. We've got it sitting up on an elevator a little bit to kind of give everyone better access to the sensor here. So, let's go ahead.

MEMBER B. RICH: One other point, we have a count of how many clickers are out there?

MS. WILBON: There are 16, 17 actually now that Jack is here.

MEMBER B. RICH: Do we have to hit Send or no? Can we try it without?

MS. FANTA: I think if you revote, you have to hit Send.

MS. WILBON: It won't hurt if you hit Send. So, we always just say hit Send, but it won't hurt if you -- we're testing. Go ahead.

(Whereupon, the voting system was tested.)

CO-CHAIR ROSENTHAL: This seems to
work.

Now this is the TAP summary scores, is that correct, Sally?

MS. TURBYVILLE: Yes.

CO-CHAIR ROSENTHAL: And just remind us, blue meant --

MS. TURBYVILLE: High.

CO-CHAIR ROSENTHAL: -- high; green is --

MS. TURBYVILLE: Green is moderate.

CO-CHAIR ROSENTHAL: And purple is?

MS. TURBYVILLE: Insufficient. So, the blue is high, the green is moderate, orange is low, purple is insufficient, and a light blue, which none of these have right now, would be not applicable.

CO-CHAIR ROSENTHAL: Okay. And the TAP had four votes on the scientific validity?

MS. TURBYVILLE: For this measure.

MEMBER CURTIS: Let me just say, though, that, overall, this is similar to the range that we had for the 1557, which is the
diabetes. But I think, again, a lot of the assumptions got more significant review.

CO-CHAIR ROSENTHAL: All right.

Got it.

And, then, the qualifications here and the data is that this applies only at the health plan level? Okay.

Then, I think it is time to vote.

One means yes and two means no.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: There's 17 of us now or 18 with --

MS. WILBON: There's 17 with Jack.

CO-CHAIR ROSENTHAL: With Jack, okay.

Ashlie, do you want to announce the vote?

It appears 13 yes and 4 no.

All right. So, this measure passes the Steering Committee's scientific review.

And I would suggest that we take about a 10-minute break, and we will come back
and do usability and feasibility.

So, for the folks on the phone, we will be back at about 11:15 Eastern time.

(Whereupon, the foregoing matter went off the record at 11:08 a.m. and went back on the record at 11:26 a.m.)

CO-CHAIR ROSENTHAL: All right. I think we will reconvene.

So, the vote, we have done now scientific acceptability. Now we have usability and feasibility to get to before lunch, if I am looking at the schedule correct. Actually, we have got a long way to go before lunch. Oh, my God.

(Laughter.)

Okay. Get another piece of fruit, everybody.

I am going to suggest, I would like to take one minute and just ask the group or posit to the group that the group might work better if we, in fact, saw the votes.

Because, then, the votes would line up in our
heads with the discussion. In the absence of that, I don't know who voted after hearing somebody speak in a certain degree of positivity or negative, then how that individual actually ended up voting. And consequently, I haven't learned anything from the vote other than that vote probably did seem to generally reflect the sense of the group.

But does anybody object to doing hand votes and the idea of this being, quote, "anonymous"? I don't think the intention was to make this anonymous. I think the intention was just to make it go faster. And as we have seen, it didn't make it go faster.

But if we could take one minute on this subject and then we can decide? I think the one minute extra that it would take to hand count -- we don't learn anything. I mean I didn't learn anything from that vote.

A couple of people said to me they voted no on this and they were the no voters,
and it was actually interesting to me who
voted no and why. I think that would help us
learn for the next ones.

I think, again, there is an issue
of, are we going to learn as a group and start
to trust each other as a group in the
discussions we make, so that the subsequent
votes or discussions don't end up having to
take two hours each on exactly the same issues
every time?

But I am okay with doing it any way
the group wants. I am just positing that we
might learn something if we actually had hand
votes.

MEMBER NEEDLEMAN: Well, if we are
going to see votes of 17 for -- nobody around
here seems particularly shy.

(Laughter.)

But one does want to create a safe
space for a minority vote. And if you can
figure out how to do that, Tom, go public with
the voting.
MS. WILBON: Can I make a suggestion?

CO-CHAIR ROSENTHAL: David? Yes, I'm sorry. Oh, Ashlie? Yes, David and then Ashlie.

MEMBER PENSON: So, I feel strongly one way or the other. One thing that I did learn yesterday, and I think it is a little bit more this is yes/no, whereas, yesterday we had four levels of grading. As the Chair, I would look at the votes, and if I didn't understand why, for example, we would have a very positive vote and someone would vote no, I would basically say, "Listen," and I said right upfront, "I don't want anyone to change their vote, but we need to have a comments as to why people voted that way."

Sort of like anytime you have a study section and someone votes outside the range, you just need to, you know -- so, I don't feel strongly one way or the other, but I think that if the vote doesn't reflect the
discussion, it would be helpful for people to at least, I won't say fess up, but just justify why they voted that way for the public record.

CO-CHAIR ROSENTHAL: It is difficult to do that without knowing who voted how.

MEMBER PENSON: No, all I did yesterday was say, you know, this doesn't really reflect what we talked about. So, would whoever voted low or insufficient just do me a favor and make some comments as to why? And people were always very straightforward with it. I certainly would be in that setting, too.

CO-CHAIR ROSENTHAL: Dolores, one more comment and then we will move on.

MEMBER YANAGIHARA: Yes, I was actually just thinking that it would be helpful to know why people voted no. I mean, if it is very clear, I mean if it is kind of evenly-split, it would maybe not make sense.
But when there's only a few no votes, just to hear the rationale would be helpful. So, you could still do the electronic voting. And, then, if there is just a few that are different than the rest, kind of what their rationale was would be helpful for me.

CO-CHAIR ROSENTHAL: All right. Well, we will try this on the usability thing and see how it goes. Okay. That is helpful. Thanks.

All right. So, let's see, the developer and then Jeptha on usability, or is it just Jeptha at this point? Okay.

MEMBER CURTIS: So, I think for usability, again, combining the reflections of the two NCQA measures that we reviewed, there really were very few concerns about the usability. And this is something that has been pilot-tested or in use for five years. They have actually done focus groups within their customer base, and they have gotten generally positive feedback. I don't think
they had specific feedback to this measure in particular, but broadly across the resource use measures that they have done. So, there wasn't a whole lot of discussion about the usability or concern about it.

CO-CHAIR ROSENTHAL: Any discussion, then, on this? Concerns? Discussion? Dolores?

MEMBER YANAGIHARA: I will just add that there is a lot of interest in also trying to figure out how to make this relevant for public reporting for consumers. And California HealthCare Foundation is actually funding some work around that to try to make it meaningful for consumers as well.

CO-CHAIR STEINWALD: Could I ask the user, I'm sorry, the developer, who are the major users? Could you characterize them for us briefly?

MR. HAMLIN: So, yes. The users we found so far are, obviously, the health plans themselves, but, also, many of the employer
groups and the business groups are very interested in the results from this information because it helps them inform their purchasing decisions for the next year.

And much of the push for us to continue to develop this approach and publicly report the results is from the employer side and the purchaser side.

MEMBER CURTIS: Just to follow up, one of the concerns that was expressed was that the overall relative research use measure was considered not terribly usable or interpretable? But, actually, the breakdown within the individual service categories was thought to be quite potentially useful.

CO-CHAIR ROSENTHAL: Going once, going twice, last comment, maybe.

MEMBER NEEDLEMAN: Yes, just I heard Mary Kay express some real skepticism about the usability of this at the plan level. So, I would like to hear more about how it is being used in practice? Also, is this part of
the required measure set or is this voluntary
by the plans?

MR. HAMLIN: It's all of our
submissions are voluntary from the plans.
These measures are currently not part of the
accreditation scoring for health plan
accreditation. We do have a large number of
plans, over 600 plans, that do report the
results to NCQA.

Most of the work we are doing right
now in interpretability is looking at the
individual results, and we have targeted areas
of education where we work specifically with
the plans to help them, one, they can actually
plug in their actual real prices into this
structure and, then, go back and there are
specific ways that you can look for
opportunities to improve these results using
that approach.

On the employer and customer and
stakeholder side, we have really tried to help
them understand what these results mean with
regard to the fact that they are standardized, and it is really sort of a snapshot of utilization for a predefined population; how these compare with the quality results and how to sort of interpret those scatterplot graphs that you have seen.

We are working, obviously, further now on the policy side, where the feds are sort of becoming more and more interested in cost-of-care measures, spinning that sort of in their perspective and helping them understand the complex methodology in sort of laymen's terms, if you will.

So, we have sort of a multi-armed approach to target the specific audiences about where it is useful, and we offer a lot of outside support through webinars, education series. There's conferences that we hold, and we present that to help each of the individual stakeholder groups understand, interpret, and make useful these results.

And that goes for all the measures,
because the measure methodology is fairly consistent across all five measures. It is just a different chronic condition. So that, again, the approach has been more, generally, how do you use relative resource use results that are produced by NCQA?

CO-CHAIR ROSENTHAL: Jack, does that answer your question?

He is asking if Mary Kay will opine on the usability.

MEMBER O'NEILL: Well, I am just trying to imagine the comparability of different entities in these measures, you know, because I think that there is so much variation in what kind of populations that different plans cover, whether it is a regional carrier or a national carrier, whether the population covered has a large distribution across the country, and what different patterns are.

I was trying to understand how some of these measures might take into account some
efficiencies that we get through certain
contracting strategies, bundled payments, and
different kinds of capitation, as opposed to
breaking every service out and looking at cost
per service line, when those costs don't
actually accrue to the payer or the purchaser.

And so, maybe I should understand
this better at this point. But I know that we
use NCQA measures. I know we are, as a
company, usually first in line for anything
NCQA does, that we have been in the quality
compass I think as long as any plan.

And so, it is not that my company
is opposed to anything that is going on here.
It is just how they are applied and whether it
is going to give people meaningful comparative
information, and it will drive accurate
decisionmaking on this larger scale.

So, those are my reservations.

CO-CHAIR ROSENTHAL: Well, that is
the usability question.

MEMBER O'NEILL: Yes.
CO-CHAIR ROSENTHAL: But it does sound like it is being widely used.

The question I have is -- and maybe it was in the materials, so I apologize if I missed it -- but what percentage of the plans of the various ones, how many did you say are using this in the voluntary mode that you have?

MR. HAMLIN: Well, right, we have about a little over 1100 plans that report HEDIS quality measures.

CO-CHAIR ROSENTHAL: Right.

MR. HAMLIN: And of those, roughly, I think 800 now are reporting the relative resource use results across the board.

CO-CHAIR ROSENTHAL: Okay. Well, that is really impressive.

What percentage are statistical outliers on this measure, either above or below?

MR. HAMLIN: For this measure, less than 1 percent. At this point, I think it is
less than even a half percent for 2010.

We just received the 2010 data last week. And so, I may have more results, but there is a very low proportion of outliers for this particular measure.

CO-CHAIR ROSENTHAL: That says to me that it is accurate.

(Laughter.)

MEMBER B. RICH: What do you mean by an outlier?

CO-CHAIR ROSENTHAL: So, I am assuming you have got statistical bands around what says that one thing is actually statistically different than another on this observed -- but you explain it.

MR. HAMLIN: Right. So, we eliminate plans from the public reporting through several methods, but right now we use the .33 to 3.0 as our cutoff points to define outliers for the plan results, generally, because we have found that the plans that fall outside of that range probably is not
necessarily an outlier in the resource use, but perhaps in some of the reporting or calculation methodology. We wanted to make sure that those are the accurate results.

Like I said, that is less than a half percent right now of plans. I am not even sure there were any for cardiovascular conditions measured this last year, this last round, which was 2009 data. They are reporting it in 2010.

MEMBER NEEDLEMAN: So, let me ask a different question. How much variance are you seeing at the plan level in this measure?

MR. HAMLIN: Well, if you look at the scatterplot that was provided in the materials, you can see there is a lot of variability both in the relative resource use and the corresponding quality scores. So, it is a very evenly-distributed scatterplot when you are looking at the results. Plans are achieving different levels of quality with very different levels of utilization across...
the board for this particular measure.

MEMBER O'NEILL: I just wanted to
point out, in case people don't realize this,
but I think CIGNA would be counted as probably
about 80 plans.

MR. HAMLIN: Yes.

MEMBER O'NEILL: I mean because we
have an HMO and PPO plan.

MR. HAMLIN: Right. We identify
plans by sub-ID. We don't count CIGNA as one
plan.

MEMBER O'NEILL: Yes. Right.

MR. HAMLIN: Yes.

CO-CHAIR ROSENTHAL: All right.

Did you have a comment here?

MS. FANTA: Yes. I just have a
question about how the results have changed
over the last five years that you have been
using it.

MR. HAMLIN: Well, obviously, the
number of outliers has significantly been
reduced. You know, the reason we waited four
years before we actually publicly reported any
of the results is because we continued to test
the reliability and the validity of these
results year over year over year.

One of the annual analyses we do is
we look at the number of new plans that are
reporting for the first time versus the number
of plans that have reported year over year.
And we look at the differences in those
results.

And right now, we are at a point
where there is very little difference in those
results. Earlier on, there was much greater
variability in the reports, partly because of
the utilization patterns, but, also, partly
because they are very complex measures with
lots of moving parts and data points. And
there were just some calculation errors, and
we were working time and time again to go back
to the plans and help them with their
calculation and find out what the reasoning
behind those outlier plans were.
MEMBER PETER: Well, I guess I am looking more in terms of whether plans are using it to improve resource use and quality.

MR. HAMLIN: Like I said earlier, we work directly with a number of plans to help them identify opportunities or ways they can calculate opportunities to improve. Again, we have generalized a lot of that knowledge and tried to publish that now, so that it is sort of available to everybody.

We are hearing from the employer and purchaser groups that the plans are bringing this information to them and showing them now some of these results. I don't have anything published to show which plans specifically are doing that, but certainly we have received feedback from multiple communities saying they are looking at this at least, if not trying to show their improvement, if you will.

MEMBER YANAGIHARA: I will just add that the health plans in California, the HMO
plans are very interested in actually moving measurement down to the physician group level. They delegate care for the population to physician groups. And so, they are very interested in that. We are doing some testing around that.

And we are finding the same kind of variability at the group level. We are finding that the majority of groups do have reportable results, you know, meeting that minimum denominator. And so, there is just a lot of interest in trying to figure out, okay, so how are the groups doing and, then, looking within the group, where is the variability?

So, I think that there is a potential to move it, where there is responsibility for caring for a population of people, to that next level.

CO-CHAIR ROSENTHAL: But, just to be clear, this group, at least in the scientific endorsement, and I am sure also in the usability endorsement, is endorsing this,
if we endorse it, at the health plan level.

Other comments on usability?

(No response.)

If not, I think it is time to vote on this. And again, I think this is binary, Helen?

DR. BURSTIN: No.

CO-CHAIR ROSENTHAL: No? This one?

How are we doing this one? Instruct us.

MS. WILBON: So, only importance and scientific acceptability are yes/no. Usability and feasibility are still rated on a high, moderate, low scale. So, you can now use one, two, or three. Or, if you think what you have learned is insufficient, which I don't think that --

CO-CHAIR ROSENTHAL: Okay. So, one is one is high, two is moderate, three is low, and four is insufficient. So, that is the voting, and I guess the consensus is we are going to speed this up.

So, let's vote.
(Whereupon, a vote was taken.)

(Six, high; nine, medium; two, low.)

CO-CHAIR ROSENTHAL: We're so good. See, it only took six seconds. You guys were right who said it was fast.

(Laughter.)

I bet you we would have seen a scatterplot like this if we had had scientific where we could have voted the same way. But that's okay.

With that, I will move to feasibility. Jeptha, you're on again.

MEMBER CURTIS: Right. So, I think for feasibility, again, this mainly has to do with whether or not it can be calculated, I believe.

In honesty, I don't think the TAP spent so much time on this. The bulk of our time was spent on scientific acceptability. But no barriers to feasibility were identified, as this is electronically-
specified and has a track record of five years of being calculated.

CO-CHAIR ROSENTHAL: I guess our NCQA friends would tell us how feasible it is, and it sounds like it is imminently feasible. Would you want to make a 10-second comment? And, then, I will call on Bill, and we will move on.

MR. HAMLIN: I mean recognizing that these are inherently complex measures with many, many data points, you know, again, our experience over time is that, in working with the reporters, the plans directly, they have increasingly become feasible. And we are continuing to try to make them more and more so.

CO-CHAIR ROSENTHAL: Do you have instances where there is not good encounter data and you have pretty good evidence that there is not good encounter data?

MR. HAMLIN: No, because we don't collect member-level data. We only collect in
the aggregate. So, we only get what the plan tells us.

CO-CHAIR ROSENTHAL: Well, okay, that's my point.

MEMBER B. RICH: I was wondering if we could separate this and tease out the third one about unintended consequences, just to emphasize -- and, obviously, the voting at the TAP reflects that, too, that there was some disagreement there.

Part 3 of the feasibility, can there be unintended consequences? And again, it gets back to our need to emphasize that this is a plan-based measure. We hear that people are trying to use this at an individual level.

And so, I would move that we remove item 3 and vote on 4a, b, and d as a block.

Is the vote --

CO-CHAIR ROSENTHAL: We are on the feasibility now. Is unintended consequences one of the feasibility ones?
MEMBER B. RICH: Yes, 4c.

CO-CHAIR ROSENTHAL: So, you would like to pull that out?

MEMBER B. RICH: Yes.

CO-CHAIR ROSENTHAL: There is a point of order. It sounds like it is still high, medium, and low on all elements combined. But if that one is a prevailing one for you, then that could form the basis of your vote.

Dolores, did you have a question?

And, then, Jack.

MEMBER YANAGIHARA: Yes, I have a question about, I think it is the fourth point.

Ben, could you speak to whether a plan can actually calculate this on their own or not? I mean my understanding is that they have to submit a whole bunch of data elements to somebody, an aggregator of some sort, to be able to do all the comparisons and everything. Is that correct?
MR. HAMLIN: Right. Somebody at NCQA. So, all the plans submit their observed data to NCQA, and we actually calculate the benchmark expectants for each individual plan. And each plan gets an individual benchmark calculated for each service category for that plan.

All of that data has to go through the entire audit process, which is why it is very complex. They are very complex measures that require a multi-step process to be submitted. So, that is where we get the validity of the data.

But, again, a plan plugging in their own actual costs will get more to the real dollar effect. And they can use the calculated benchmarks that we provide them as a relative comparison tool, but, really, when you start plugging in member-level data and actual cost into these, they are actually looking mostly at some of the specific service categories to try to identify opportunities to
improve. It is not to compare themselves to another plan.

So, while those calculated benchmarks will help them understand what they look like compared to the same plan in the same population, relatively speaking, it is more of sort of a reference point when you start plugging in your own numbers. And plans are actually using their own numbers. We have heard that many plans are actually plugging in their own numbers and seeing how they comparing using those.

CO-CHAIR ROSENTHAL: So, it is not a black box?

MR. HAMLIN: It's not, no.

CO-CHAIR ROSENTHAL: Okay. Jack?

MEMBER NEEDLEMAN: I need a clarification. Mary Kay talked about groups that are basically accepting capitation, where provision of the encounter-level data is a courtesy. And many plans carve out their pharmacy benefits and many plans carve out
their mental health or behavioral health benefits.

And in terms of getting total cost, it is critical that, particularly given your pricing mechanism, you have to get all the encounter data back from all the places that it has been carved out to.

And this is a common issue across all the charge-based measures that we are looking at. What are your plans telling you about how successful they are in fully capturing their carved-out charges, you know, use experience at a level that it is fully captured in the costs that are being calculated?

MR. HAMLIN: So, our perspective is that the plans are responsible, obviously, for correlating pulling all of this information together.

We have attempted to do some research in the last year, and we are trying to expand on that this year, in looking at the
differences in plan reporting between different benefit designs. And we have seen some relationships that have started to form there.

The confidence in those relationships is still not where we would like it to be. So, we are diving back into it again to try, with more plans reporting and more of this information now available, and people are more comfortable with the approaches, to essentially redoing that analysis to understand how benefit design might affect results. I mean that information is not yet available.

CO-CHAIR ROSENTHAL: Yes, and, unfortunately, Jack, that is a scientific question.

MEMBER NEEDLEMAN: No, no, I consider it a feasibility question.

CO-CHAIR ROSENTHAL: Well, it would have also been a scientific one. Because if Plan A has all the mental health carved out
and those data aren't in the datasets, you would skew results when, again, you are comparing CIGNA of the Northwest with Blue Cross of the Southeast.

MEMBER NEEDLEMAN: Right, but on 1604 I was on the feasibility thing. So, where you sit determines where you stand.

CO-CHAIR ROSENTHAL: Fair enough.

MEMBER NEEDLEMAN: But, in principle, if we had all these costs, we could do it. That, to me, is the scientific question. In practice, can we get all the billings? That is a feasibility question. And what I heard, not to be too crass, is you don't know.

MR. HAMLIN: Well, we do know. I mean our auditors are the ones who are responsible for validating the plan data and reporting back to us if there are significant gaps in the data before they report it to NCQA.

I don't know particularly from the
performance measurement department because it is a whole separate process. It is involving reporting and data collection. So, I could probably find out, but I don't know.

CO-CHAIR ROSENTHAL: Mary Kay?

MEMBER O'NEILL: Things like bundled payment and DRG are not, they are not benefit-design-related. They are contracting. So, that would not be teased out by looking at different benefit design.

CO-CHAIR ROSENTHAL: I think Dr. Needleman's concern is that, depending on plan design, you have stuff that is carved out. Or, in California where it is capitated, and some of the groups are less enthusiastic about sending the encounter data in because it is just a cost to send the data back to the health plan, when you are fully at risk. And so, whether or not those kinds of things could skew a comparison is both a feasibility and scientific thing.

But, I mean, I think the major
point of this we blew past without raising it
when it might have affected the thing.
Because the plans are sending the stuff, and
his point is they get all the data. It is
perfectly feasible for them, then, to
calculate --

MR. HAMLIN: And there is
definitely an incentive for them to get that
data because it only helps their results.

CO-CHAIR ROSENTHAL: And there is
an incentive for them to get it in, if they
have it.

MR. HAMLIN: If they can, if they
have it.

CO-CHAIR ROSENTHAL: If they have
it.

Jeptha, I wonder, if you wouldn't
mind, there were three votes from the TAP on
this susceptibility to inaccuracies. Would
you maybe share your group's thinking, whether
it was the same as the question that Bill
posed or whether there was some other thing
that caused there to be sort of those three "ifier" votes, if you can remember?

MEMBER CURTIS: You know, I don't think I can recall specifically. I think we really only focused on that when there were lows.

CO-CHAIR ROSENTHAL: Okay.

MEMBER CURTIS: We tried to break out like why were they low. I think in this case -- and these were three moderates as opposed to lows -- it wouldn't have hit that threshold.

CO-CHAIR ROSENTHAL: Okay. All right.

MEMBER CURTIS: But I think there was, again, always concern about any of these, that there's susceptibility to unintended consequences. But this is actually less so than --

CO-CHAIR ROSENTHAL: Some of the others that we are going to see.

MEMBER CURTIS: -- some of the
others.

CO-CHAIR ROSENTHAL: Yes. Okay.

All right. So, this is a pretty good vote from your -- Bill and, then, Jack, and then I think we ought to call the question on this one.

MEMBER B. RICH: Just a point of order. Are we forbidden to divide a question? I mean it is a normal parliamentary procedure.

DR. BURSTIN: It's not that you're forbidden. It is just that the way we have established the process, the subcriteria are more so in the domain of the TAPs, who do that work for you, bring it to you for your review, so you can make the overall assessment of the criteria. So, it would be taking a deep dive.

I think, again, keep in mind, everything that we have put out will, of course, be in the reports. So, in this section there will obviously be a discussion on this measure, that there was some concern expressed by the Steering Committee regarding
potential unintended consequences.

CO-CHAIR ROSENTHAL: And I suppose, to respond to the point of order, you could make a motion to the effect of you would like to have the thing called out. We could vote on whether to call it out and do it like that.

I mean the risk, if we do it here, likely everybody is going to have one thing they might want to pull out on one, and, then, again, it will be very hard to work our way through all of this.

MEMBER B. RICH: The reason I raised it --

CO-CHAIR ROSENTHAL: Yes, okay, go ahead.

MEMBER B. RICH: The reason I raised that is not only for this issue, but later on, when we are actually functioning as a TAP, are we going to consider these as a whole or individually?

CO-CHAIR ROSENTHAL: Well, no.

When we have served as a committee-of-the-
whole, we did vote on all of them
individually. If you recall on the Ingenix
ones and the others, we pulled every single
one out.

      Paul?

      MEMBER BARNETT: I would just
resist deviating too much from the process
because I think we are never going to get
through all of this.

      CO-CHAIR ROSENTHAL: I am not
hearing much enthusiasm, but if you would like
to make the motion, we could vote on it.

      MEMBER B. RICH: No, I was just, a
point of order --

      CO-CHAIR ROSENTHAL: Okay.

      MEMBER B. RICH: -- wanting to know
if we were proscribed from it.

      CO-CHAIR ROSENTHAL: No, I think
the answer is no, but I am not hearing wild
enthusiasm for it, either.

      So, Jack, last comment on this,
and, then, let's --
MEMBER NEEDLEMAN: Yes, this issue about the carve-outs, and I appreciate that you have got folks who are auditing the data and supposedly telling you if it sufficient or not, but this issue of the carve-outs came out on the phone call we had last week when we were talking to the Ingenix people about their whole resource use. And they are one of the folks that pull all this stuff together.

If I am remembering that conversation correctly, they talked about imputing pharmacy costs for some of the groups that couldn't produce the full billing. And that makes all of these charge-based measures, that is a big red flag for me in terms of getting accurate, full resource use measures from the data sources that supposedly are providing this out of administrative data.

I am going to vote insufficient on this one because I haven't got a firm answer that the data is really there.

MR. HAMLIN: Well, just to be
clear, there is no imputation for any of the
assignment of cost for these measures, for our
measures.

MEMBER NEEDLEMAN: No, but what
Ingenix said is there are a number of groups
that they were working with on their full
resource use measure that couldn't provide
them with the carved-out pharmacy costs, and
they were imputing it.

So, I don't know what your plans
are doing if they can't get it.

MR. HAMLIN: They are not allowed
to impute it. So, they are either reporting
it as non-reportable for the pharmacy
components or they are somehow tainting the
data and integrating it into their systems.

MEMBER NEEDLEMAN: Is it possible
to find out how many are reporting it as not
available?

MR. HAMLIN: I mean it is possible,
not in the next couple of days, though. I
would have to go back to our Audit Department
and make requests.

MEMBER REDFEARN: I think what Ingenix does is they stratify. They do what we do in California for profiling because we have a lot of members that don't have pharmacy benefits, and we just stratify. We have a dimension with pharmacy and without pharmacy, and we calculate it separately. I think that is what Ingenix was suggesting in the discussion last week. That is my recollection.

CO-CHAIR ROSENTHAL: Well, if I could, just to keep us on the point, it may be extremely, extremely relevant in the Ingenix case. It may be less relevant here; I don't know. Some of this may wash out.

But, certainly, we ought to keep to discussing this NCQA one on its own bottom and not get distracted necessarily, unless we can't -- but that is a very good point.

MEMBER NEEDLEMAN: Yes. No, I mean the assertion is that all the pharmacy costs
are in the measure of resources that are being counted here.

CO-CHAIR ROSENTHAL: Right. And they are clearly not. What the implication is, then, going to be is that there will be a relative inaccuracy in comparing Plan A with Plan B, one that has pharmacy in and one that doesn't.

MR. HAMLIN: And the pharmacy is a separate component of these measures.

CO-CHAIR ROSENTHAL: Huh?

MR. HAMLIN: The pharmacy component is completely separate from these measures. It is not part of the total medical rollup that we include in those scatterplots. So, it is total medical against quality and pharmacy against quality. So, they are actually held completely separate in these results.

CO-CHAIR ROSENTHAL: Okay, which is another way of stratifying it --

MR. HAMLIN: Right. Exactly.

CO-CHAIR ROSENTHAL: -- with and
without pharmacy benefits. Okay.

So, that is the other thing that is
different about this and the Ingenix measure.
The Ingenix measure ended up, as I recall,
with sort of one number, totally rolled up,
and this is not one number totally rolled up.
There's, as you have now described it, fairly
complex reporting out of the various things
with a variety of stratifications.

MR. HAMLIN: I mean there are
several high-level rollups that we use for
public reporting in those scatterplots. But,
again, each one is then subdivided into these
specific service categories. So, it is both.

CO-CHAIR ROSENTHAL: David, the
last comment on this.

MEMBER REDFEARN: It just seems to
me that this issue of availability of data is
going to apply to every measure. So, I don't
think there is anything unique about this
issue for this particular measure.

And basically, I would say it has
got high because they are all subject to the
same problem. You always know that you have
this data issue. So, I wouldn't use it
against this measure.

CO-CHAIR ROSENTHAL: Yes, I might
not use it against this measure, either, but
I am going to be inclined to use it against
some other measures, because I do think that
the level of aggregation and who it is being
reported on makes a big difference.

I think it is probably the case
that this washes out across 800 health plans.
It probably doesn't wash out comparing Medical
Group A or Doctor A to Doctor B.

So, I think the issue is going to
be there on all of them, but it may quite vary
as to its applicability or whether it renders
a particular measure not usable. That is my
opinion on the thing.

I would suggest, then, unless there
is some other burning comment on this point,
that it is time to vote. And this one, again,
is high, medium, low, insufficient. Right,
Ashlie? Am I right on that one now? So, one,
two, three, and four?

MS. WILBON: Right, uh-huh.

(Whereupon, a vote was taken.)

So, for people on the phone, the
vote is seven high, six moderate, three low,
and one insufficient.

Who voted insufficient?

(Laughter.)

We knew. We knew. You announced
it. So, of course, we knew.

Okay. So, are we done, then, with
this? Oh, there is an overall vote? Okay.

That's right. Of course.

MS. WILBON: Whether or not the
measure should be recommended for endorsement.

CO-CHAIR ROSENTHAL: Okay. Thank
you.

So, do we need any further
discussion on the overall thing? We have
discussed each of these sub-elements. I am
open to some discussion, but I think Bill was
first and, then, Bill was second.

MEMBER B. RICH: One of the things
you asked us to do, if there was a real
disparity, to discuss it. I was going to vote
high, high, high, except for 4c, unintended
consequences.

We have heard that it is being used
and tested at a group in California and
others. I think this obviates the episode,
some of the tenets of an effective episode
group or of attribution and statistical size
of sample.

And so, I voted high. But since I
was unable to express that with the division,
that is why I voted low.

CO-CHAIR ROSENTHAL: Other
discussion? Bill, I'm sorry, you're next.

MEMBER GOLDEN: When we say we are
endorsing this, we are endorsing as described
or as delimited and for what purpose?

CO-CHAIR ROSENTHAL: I believe that
that is exactly the question.

But, Helen, do you want to clarify that?

DR. BURSTIN: NQF does not specifically delineate which purpose. There is a new Measures Application Partnership that has been brought up to specifically try to make some of those calls.

At this point, you are recommending for endorsement, and you are still fairly early in the process, if you remember that flowchart. At this point, your recommendations go out to the public for comment. So, it is still a long way before it is endorsed. So, you are recommending for endorsement as appropriate for public accountability and quality improvement.

CO-CHAIR ROSENTHAL: But the use is --

DR. BURSTIN: At the level of analysis --

CO-CHAIR ROSENTHAL: Yes, is that
what you meant by the uses?

    DR. BURSTIN: Yes.

    MEMBER GOLDEN: So, this would be endorsing it for public accountability and quality improvement, for both functions?

    CO-CHAIR ROSENTHAL: Yes, we have clarified that --

    MEMBER GOLDEN: Okay.

    CO-CHAIR ROSENTHAL: -- I would say now ad nauseam.

    (Laughter.)

    If you vote yes, you are voting both for quality improvement and public -- for me, the test is, do I want to see these results on the front page of The New York Times?

    You snicker, but, I mean, that is the way it is. Now it doesn't appear that NCQA has used these that way, but if it were endorsed by this group, they would be free to take all these results and put them in The New York Times.
MR. HAMLIN: We don't have control over the use of our measures once they --

CO-CHAIR ROSENTHAL: Got it.

MR. HAMLIN: We restrict how we use them in our programs, but --

CO-CHAIR ROSENTHAL: And that is what is being said when the answer is put the way it is put. We are not in control of the uses once we have endorsed them. But we clearly are endorsing them for both purposes with a yes vote.

Now this is a yes and no now, right?

DR. BURSTIN: Yes.

CO-CHAIR ROSENTHAL: So, it is yes, no, and abstain. Ah, you can abstain.

So, if there is no further discussion on this, are we ready to vote on this? Okay.

(Whereupon, a vote was taken.)

All right, so for the people on the phone -- oh, how are we getting the votes of
the people on the phone? Oh, he is emailing it, and you are calculating that in?

MS. WILBON: No, we will add it.

CO-CHAIR ROSENTHAL: You will add it into the final tabulation? One person on the phone. Okay.

So, the vote on this is 13 yes, 3 no, and 1 abstention.

So, by this vote, the measure would be recommended from this group to the Board for endorsement -- to the public for comment and, then, after that to the Board. Thank you. Too many steps.

All right. I think we are concluded on this measure, and I believe this is the only NCQA measure that we have to consider now this morning for today.

MR. HAMLIN: For today. Tomorrow, it is all over again.

CO-CHAIR ROSENTHAL: So, we will see you again tomorrow.

(Laughter.)
Now we will be really ready to roll tomorrow.

Okay. So, I think, with that, do we have somebody now from ABMS to discuss 1570?

DR. WEISS: Kevin Weiss and Todd Lee here.

CO-CHAIR ROSENTHAL: I'm sorry. I apologize, we are a little bit behind, but we appreciate you being on the phone.

So, for everybody in the room, this is Measure 1570, acute myocardial infarction episode-of-care for 30 days following the event.

So, who's on the phone?

DR. WEISS: Kevin Weiss and Todd Lee.

CO-CHAIR ROSENTHAL: Terrific. If you would give us a brief synopsis? And, then, Jeptha will give us the TAP review.

DR. WEISS: Great. So, good. I guess it is getting to be good afternoon to...
We are presenting this measure, which comes from a series of measures from the Robert Wood Johnson Foundation-funded project. I mention that because some of the things that you will be seeing in this measure will be consistent across a number of our measures, such as our risk adjustment and our costing methodology, and may prove to be efficient for you as you deliberate later.

In developing this measure and the other measures, we were looking to identify a very specific set of resource use that reflected an episode-of-care that was directly tied to a clinical episode. It was following very closely the concept outlined by NQF by the Episode-of-Care Workgroup, which released a report, and I had the privilege of co-chairing.

This particular measure, actually, was set to meet a diagram that was actually put forward in that episode-of-care framework,
which is to parse out parts of care that relate to critically-important episodes that could be measured in terms of these issues and must be matched with quality measures, because the resource use by itself was thought to be inadequate.

In developing this measure, it reflected the efforts of a multidisciplinary workgroup of clinicians, mostly physicians, but a broad multidisciplinary group. And we asked them to try to work towards attribution, if possible, to the most granular place, down to an individual physician, if possible. And you will see that this was added to the hospital level.

The period that was chosen was one that is very familiar in clinical literature in terms of hospital AMI episode. And that is the 30-day window. The intent was to identify individuals who had AMI and follow them through the hospital and post-hospital care through 30 days.
Because of the nature of how care is delivered, and so much of the care for individuals in this environment where it is not one where an individual will direct their care, they will often be taken by ambulance to a hospital, assigned a physician in the emergency room, assigned whoever is on call for the cath lab, assigned a hospitalist, and so on and so forth. It didn't feel like there was a way to look at anything other than a system-based approach, an integrated attribution look.

And that is why you will see the hospital level, even though the immediate post-discharge may be to a physician. Much of that was predetermined at the time of discharge for some of the resource requirements that would be in terms of tests that were ordered in that time period.

So, that is how this measure is framed. Maybe it is very good to stop there and just let it go from there, and we will be
here to answer questions.

CO-CHAIR ROSENTHAL: Great. Thank you very much.

DR. WEISS: Let me just put a final note on this for the project because you will see this in other measures as well, and it is probably good at the beginning to give you that sense.

The purpose of this, the RWJ project, was to develop measures. We are very different than NCQA in the sense that these measures are newly-developed. They haven't been field-tested widely. We are undergoing a field testing right now in a couple of environments.

But you will see pretty clearly that these measures have not been fully tested in various communities as yet. And that will be something that will be consistent not just for this measure, but other measures that we have submitted.

CO-CHAIR ROSENTHAL: Well, thank
you, and we will get to that question, I think, when we get to the scientific part of the thing.

Jeptha, let's do importance, and let's agree on the groundrules of how we are going to think about importance. It is, is AMI an important thing to measure resource utilization? And, then, I think it will make this one go much faster.

But, Jeptha, give us the TAP --

MEMBER CURTIS: Yes. So, you can see the votes up there. In general, as opposed to being a high-impact condition with evidence of significant resource use, obviously, a high degree of certainty. There was a little bit more discussion as to whether or not in this application the ABMS staff and workgroups had made a convincing-enough case so that there was substantial real variation in this interval of zero to 30 days.

And so, I think that's why there were some people who weren't as convinced, but
it was fairly balanced between people who were entirely convinced and those who were not convinced.

Again, in terms of resource use, categories are consistent and representative. With regard to the TAP discussion of that, again, we are just considering that as to, were they looking at all the categories of resource use? Were they systematically excluding pharmacy benefits? That's probably not the right one. Or physician visits? You know, we are presenting a broad view of resource use.

And within that, I think one of the concerns again around this measure had to do with the exclusion criteria that was with SNF. I don't think they were able to consistently capture SNF utilization. And we can get into that in the scientific acceptability, but it does overlap with 1d.

CO-CHAIR ROSENTHAL: Again, I think for the purposes of getting through, the
notion of not commingling this with the scientific things, the question on the table here for importance would be, is acute myocardial infarction and its resource use an important thing to be measuring, if assuming, then, in the later pieces of this we can measure it accurately and completely and attribute it properly, et cetera?

Does anybody want to make any discussion points about importance, however, with the caveat that I have put around the topic?

(No response.)

All right. Hearing none, so the vote on this one, this is an all-or-nothing vote and it is yes or no.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Okay, we're up to 17. Oh, there's 18. Ah, did you vote? Oh, Bill's out of the room. Okay. So, we've got the votes. Okay. All right. So, I think we're fine.
Seventeen to nothing, a clean sweep. When you define the question narrowly enough, we can get unanimity here.

Okay. I think, with that, Jeptha, let's now -- I think, again, in a time check, we are supposed to have public comment, then, at 12:25. And so, we have 15 minutes that we can begin the discussion of the scientific aspects of this. And so, let's hear what the TAP thoughts were about that.

MEMBER CURTIS: But the expectation is that we will stop at 12:25 for public comment, no matter where we are? Okay.

So, I think I will do my best to summarize kind of the overall concerns. The measure itself is quite different, mainly in terms of its outcome. And we talked a little bit about all resource use. This is a resource-specific use measure. And a lot of the application goes into how they came about defining what were the resource uses that were reasonably associated with the AMI episode or
It probably, I think, in general, held better with AMI with the short-term outcome as opposed to some of the other ABMS measures. But it is worth considering that as you go along.

So, starting with the population, it was a fairly well-clarified population of 410.X1, I believe. So, it was your standard AMI population.

They did, I think, a reasonable job of applying reasonable exclusions using the NCQA exclusions as sort of a baseline and, then, building off of those. There was one major concern in the exclusion criteria. It was that they excluded patients who died within the hospital.

And that was a source of significant concern across the entire TAP, I believe, that if you died post-op or post-discharge day one, all your costs were included in the measure. If you died in the
hospital, you dropped out of the denominator. The consensus, I believe, was that that wasn't necessarily a valid way of comparing hospital organizations or introduced a form of bias.

The other major discussion point really had to do with, No. 1, how their costing methodology was taking place. And there was some concerns of the accuracy or the up-to-datedness of the codes that were included in the outcome. But, really, the major concern was whether or not this represented truly a comprehensive look at resource use post-MI or in the setting of an MI.

If you go through the packet, there is a description of this iterative process that they went through with the Working Group, which included several esteemed health services researchers and clinicians. So, I think they did a good job, but I think inherent in this, or at least the feeling from the TAP, was that inherent in the selection...
criteria it was, by definition, vulnerable to
errors or decisions that could be construed as
ea errors. I will rephrase that.

    I don't know where in the packet it
is, but there is a detailed list of what codes
were included as being applicable to MI care.
And they capture things like arrhythmias and
heart failure readmissions and anything that
had a primary diagnosis that met one of their
criteria. So, repeat MIs, heart failure, et
cetera.

    But there were a lot of things that
could be reasonably related to the care of AMI
patients that weren't included. So, they made
decisions that struck the TAP as being
arbitrary at times.

    For instance -- and I can't
remember if this is the right example or
not -- but I think that they, for instance,
didn't include renal insufficiency as a claim
that could be reasonably associated with the
care delivered in that episode. So, if a
patient had bypass surgery or coronary angioplasty, had a dye load, had contrast nephropathy, then came back with a readmission for that, that is not captured. That is invisible in this particular approach. And there was some level of discomfort with that, again, what I would call the arbitrariness of those decisions.

Another focused example is in the use of medications that they selected. They didn't include all the medications, but they tried to drill down on specific categories of medications. So, I think they got the big-ticket items. They got the lipid-lowering agents and they got the beta blockers, et cetera. But, for instance, the whole class of anti-arrhythmic medicines weren't included.

And it kind of highlights this, well, why? Why? What was the rationale? And there wasn't a lot of rationale for the specific decisions beyond saying, "Well, we vetted it through the Workgroup and this is
what they came up with." So, I think that was
the other major concern from the TAP.

In terms of reliability and
validity testing, there wasn't a lot. As I
think as the developer mentioned, these
measures are probably a little bit upstream
from where the NCQA measure is in terms of how
much it has been in use and how much data they
had to demonstrate the reliability and
validity testing of it.

And I think this is more applicable
to the other measures where they are
attributing to the individual provider. In
this case, they are attributing to the
hospital-level, but there was concern within
the TAP that the attribution may not have
always been perfect in terms of transfers of
care, like how do you actually make sure that
you are attributing to the right institution?

So, with that, I will pause and
leave it open for discussion.

CO-CHAIR ROSENTHAL: Great. And,
then, if I could, I think we skipped a step on the last one. I think we have people who have specifically reviewed this specific proposal and this specific aspect of it. And in this case, it is Jeffrey Rich.

So, perhaps if you would give a couple of comments and, then, we will open it up to discussion.

MEMBER J. RICH: Sure. Thanks. Just so everybody knows, this is the world in which I live as a cardiac surgeon. And this resonates highly with me. I thought the measure had a lot of importance. I thought there were a lot of great things about the measure and it worked well, and I thought there were some inconsistencies and issues and questions that I had.

But I wanted to bring forward, the first is just the general one. It also applied to the last measure. This is the continuous coverage principle, that you have to be continuously covered for 36 months or 24
months. I will submit that, if you have somebody entering Medicare at the age of 65, you will never measure a 65-year-old patient in this measure unless you have some sort of gap coverage or ways to handle those gaps.

And in addition, I wasn't sure if this is picking up the HMO Medicare patients. That is a very difficult database to tap into. When I was at CMS, we had no access to it. So, I would bring forward at least to the commercial payers at least getting into their databases.

The primary diagnosis is 410-XX. I am not sure if it is your primary diagnosis or it is going to be your discharge DRG. Because if you come in with an AMI and get a CABG, you may not have as the primary diagnosis 410-XX anymore.

Your exclusions, you excluded the uninsured, the deaths, the SNF transfers, the greater-than-85, end-stage renal disease, and end-stage liver disease. And somewhere in the
analysis, I think the TAP said that, when they looked at the reliability and validity, there were 47 percent of patients excluded. So, it becomes, I think, in the general discussion we had early on a very narrow patient population that we are looking at.

Other questions I had is, when you do your analysis and give your reports, how do you control for payer mix and this whole issue of transferring, since, for instance, at Sentara Heart Hospital, we are a hub hospital. So, many of our patients had their AMI somewhere else and get discharged from that hospital, and I know they have controlled for it somehow in the discharge from the AMI hospital. There is, I believe, not an exclusion, but at least a measure there that picks it up. But I don't know if it picks it up on the incoming hospital.

So, when we receive a patient who has had an AMI, they may be coming in just for coronary bypass graphing. So, you may lose a
lot of patients in specialty hospitals if it
is not handled right.

I think you answered the question
when it starts, but I am not sure when the
measurement starts. If it started at the
index hospital, the index event, and you get
transferred to another hospital, does that
event start at the index hospital or does it
transfer over to the other hospital?

And, then, there was just some
basic inconsistencies because it talked about
hospital-level attribution, and, then, in some
of the sections there was attribution at the
individual provider level, which I didn't
think was appropriate because I think Kevin
Weiss said it nicely; this is a system issue.

There is stratification for heart
failure. There was some concern there. And
I would say excluding patients over 85 is a
little concerning. We get more and more of
those patients. Now that is going to be a
high-cost area for us.
On the other hand, when I looked at the reliability and validity testing, looked at all the charts and looked at how the data parsed out between the different cost buckets and things like that, it felt real to me. It felt just like what I see on a daily basis. So, I didn't have a lot of angst about it. There wasn't a lot of variation.

However the measure is being used in that dataset, it is providing, even though it is a narrow population, it is providing a reasonably-accurate picture to me and feedback. It seemed like, yes, I think that is about how much we spend on pharmacy; I think that is about how much we spend on the physician component.

So, I think I will stop there.

CO-CHAIR ROSENTHAL: All right.

Thank you very much.

We are going to try to stick pretty close to the schedule on getting the public comments because, if there are people on the
phone who are waiting or have been sitting in
their office expecting to dial in at exactly
at 12:25, I think we ought to try to respect
that.

So, we have about five minutes that
we can begin the discussion, and there were a
variety of issues raised. We can sort of take
general responses or we could go down them
sort of as they were articulated. And I heard
several.

One was potential biases,
particularly driven by the fact that they
exclude deaths. The whole question of which
exclusions and why. To what degree has there
been validity testing? Whether it is the
admitting diagnosis or the discharge diagnosis
would be a factor. And this issue of how the
transfers are handled. These were the issues
that I heard raised between the TAP and,
Jeffrey, your conversation.

Did I miss one? That sounds like
the key ones.
Maybe we will take them in sort of order. And, then, we can also get feedback from Kevin and his team on possibly answering some of these.

So, maybe we start, because the simplest one of these might be this death-in-the-hospital question. Anybody have any comments or thoughts or observations about that specific issue?

CO-CHAIR STEINWALD: Well, you don't want death to look good, right? Isn't that essentially why you remove the deaths, is that you don't want to give the impression that death is associated with lower resource --

CO-CHAIR ROSENTHAL: The cost, right.

MEMBER CURTIS: There are plenty of deaths in the hospital associated with very high costs as well. I mean it seems just fundamentally wrong. How do you get this in-hospital death as being different than
hospital-stay-plus-one-day death? Why include one group and not include the other group? Are we encouraging people to, then, keep people in the hospital who you think are going to die longer because you want them to die in the hospital, to take the logical extreme of that? And they will disappear from the resource use measure.

MEMBER RUDOLPH: Yes, I think there would be a way to exclude patients who died in the first or second day because that is usually when the main procedure has taken place. But anyone who is there more than two days would be included in the resource use cost, even though they had died at some point later.

CO-CHAIR ROSENTHAL: I think one of the concerns I would have on this one, but, again, I think we ought to ask Kevin what their logic was, and then we can decide if this is important or not. But the concern I would have is what I like a lot about the NCQA
one was the idea of being able to link this up
with quality measures right from the get-go.
And if you have excluded the deaths, now what
do you do when you are going to try to match
this up with mortality and other sorts of
things? You are going -- well, I don't know,
it just seems confusing to me to have done
that. I would take your point.

Kevin, can we ask you what your
group's thoughts were about this exclusion
criteria?

DR. WEISS: Oh, of course, you can.
I am pleased to respond.

So, we looked at the question of a
person dying during the episode. And in order
for us to capture death in the episode, we
would have to capture death outside of the
hospital, and that is not an easy thing to
capture. There just is no easy, reliable way
to do so through the current data streams we
have.

And we do get a nice and clear and
clean piece of data from death in hospital, of course. So, we are left with the inability to have a consistent recognition of that, of the decedent population throughout the episode.

One way -- you know, we have been thinking about this a number of ways, of course, and have since the get-go -- but one way to manage it would be just to make sure in naming the measure that it set up cost for episodes for people who left the hospital alive. That is the only way you can kind of get around this problem of lack of information to identify the cohort of decedents.

CO-CHAIR ROSENTHAL: But your group thought it was more logical or consistent to simply exclude all the deaths than to do what you just said a second ago, which would be death outside the hospital or survival to the point of discharge?

DR. WEISS: Yes, and our group actually didn't, they didn't consider the -- it is only on reflection and after the TAP
meeting that we began to think about, you
know, is there a way to manage that? But our
group was pretty clear that, since we could
not identify the cohort of decedents
throughout the entire period, it made logical
sense to suppress that, recognizing that it
would create a directional bias, but it would
be a consistent directional bias, and easily
identifiable.

And when matched with a quality
indicator of mortality, both in-hospital and
30-day mortality, which would give the balance
that would be needed for this measure, that
you actually would have a nice picture, which
is regardless of resource use, you would still
know independently about how hospitals were
doing in terms of their in-hospital mortality.

CO-CHAIR ROSENTHAL: Yes, any of
these measures that are beyond a hospital
period, if it is Medicare, you could go to
them and get an all-payer, I mean an all-
Medicare for all time thing, but it is not
cheap and you can't get it for the commercials. So, that sort of makes sense.

Anybody else have any comments on this point? Jeffrey, you have a comment on this point?

MEMBER J. RICH: Yes. No, I do think it is important, and I didn't include it in my little analysis because Jeptha did.

But you could get death outside of the hospital like we are doing in the SES database by the Social Security Death Index. It costs about 35 cents. So, if you wanted to include this, Kevin, you could actually add that to your measure, that the Social Security Death Index would track deaths outside of the hospital within 30 days very easily.

I think either a patient who dies with this diagnosis is either your cheapest or your most expensive patient, depending on how long you could get them to stay on.

As a complementary question to that, Kevin, this is for 30 days. If a
patient comes in with an AMI and stays in for more than 30 days, do you truncate the measuring period at 30 days?

CO-CHAIR ROSENTHAL: Kevin, did you hear the question?

DR. WEISS: I did, but what I am going to do is ask Todd Lee to help me with that. I don't recall quite offhand.

DR. LEE: So, yes, the patients who would have lengths of stay longer than 30 days would be, right, truncated at the 30-day period. In our test dataset, you know, I don't think that happened maybe more than one or two times in our whole population.

CO-CHAIR ROSENTHAL: And as a point of order, if this death issue were thought by the group to be a really critical one, can it be adjusted on the fly, in much the same way NCQA kind of, I think, adjusted theirs on the fly in clarifying a point? Or are we limited entirely to what is on the piece of paper?

MS. TURBYVILLE: It is up to the
developer to respond whether or not they could
make that adjustment (a), and, then, it would
have to be within a timely manner in order for
it to be within this project, right? So,
question (a) is, can the measure developer
make these types of adjustments and, if so,
then we would work with them to see if it can
be timely enough to fit in this project or if
it would have to be a future project.

CO-CHAIR ROSENTHAL: Okay. Thank
you for the clarification.

I think, in the interest of
respecting the public time and our getting to
lunch and getting then back to work, I think,
with your guys' permission, we will move the
rest of the discussion to after the lunch
break.

And what's the process now for
getting public comment?

MS. TURBYVILLE: So, we just ask
the operator to please open the line and
provide instructions for those on the public
line to ask questions or provide input to the Steering Committee. And, then, we will go and make sure no one in the audience here in person has input as well.

THE OPERATOR: So, if you would like to comment, make public comment, over the telephone at this time, please press *1. Again, that is *1 for public comment over the telephone.

(No response.)

There appears to be no public comment at this time.

CO-CHAIR ROSENTHAL: Anybody in the room who is a public person, if you have a comment, now would be the time to make it on any of these topics.

(No response.)

Okay. Hearing none, I think this means a break for lunch, and well-earned. We have one half-hour allotted to lunch and that's it, and, then, it is back to the salt mines.
We've obviously got a lot of work to do on this one. I think the good news is, though, 1591 was taken off. And so, we do have a little extra time to pound on the scientific issues on this one, and it will probably be worth our while to do that and really be sure we are comfortable in trying to grapple with these at one o'clock.

So, we're adjourned.

MEMBER CURTIS: Let me just say, though, if we are going to save that hour with that measure being pulled, we should maybe consider bringing one of tomorrow's measures up because tomorrow is a very busy day. So, it would be great to re-use that.

CO-CHAIR ROSENTHAL: Got it. Thank you, Jeptha. We'll do that.

(Whereupon, the foregoing matter went off the record for lunch at 12:31 p.m. and went back on the record at 1:09 p.m.)
CO-CHAIR ROSENTHAL: All right, we will reconvene. Just like there is no such thing as a 10-minute break, there is no such thing as a 30-minute lunch break.

It is like surgeon time. It's like surgeon time, right? I'll be done in 10 minutes.

(Laughter.)

And 10 minutes is a half an hour, and if the guys it's a half an hour, that's bad news.

(Laughter.)

All right. So, we will pick back up on the scientific aspects of this AMI measure.

And, Kevin, are you guys still with us?

DR. LEE: Yes. This is Todd Lee.

CO-CHAIR ROSENTHAL: All right. And did you get some lunch?
CO-CHAIR ROSENTHAL: Okay. I guess we don't have to be concerned as to whether you got lunch.

Actually, Helen was hoping that actually we would weigh-in a little bit on this idea about the death in the hospital as an issue.

I don't think we want to cull it out as a voting item, but is there a general sense that it would be preferable to have the deaths in the hospital in as opposed to the way they have done it? Let's have a straw vote on this. It doesn't mean anything, but it is a straw vote.

Who thinks that the deaths in the hospital ought to be in?

Okay. Who thinks it's fine the way they have it?

And did anybody abstain?

All right. So, Kevin, it was pretty unanimous in the room that, if it were
possible to include the deaths in the hospital and, then, redescribe the thing as applying at 30 days to those who survived in the hospital, who at least survived the hospital -- did I get that right? I didn't say that right. Well, you get what I mean. The group would prefer that the deaths in the hospital be included.

MEMBER CURTIS: Can I just follow up on that?

CO-CHAIR ROSENTHAL: Yes.

MEMBER CURTIS: I think that was the major criteria by which they received lots of low votes on this 2b1, for instance.

CO-CHAIR ROSENTHAL: Okay. All right. Well, all right. So, there we go. Well, let's ask, because, I mean, again, if this is a compelling issue, do you think that's fixable?

DR. WEISS: Yes, it's fixable, and it sounds like it is a compelling issue. So, we will have to take it into consideration.
But we really appreciate the way that you have given us feedback. So, thank you.

CO-CHAIR ROSENTHAL: Okay. Yes, and I don't think anybody is saying that, somehow or another, you have to figure out how to get that 30-day mortality rate in order for it to be okay.

So, let's move on to the question that was raised about the various exclusion criteria.

And, Jeptha, would you mind, I heard several -- there were some classifications of drugs that you were curious about, the question about ESRD and cancer, and, then, the one that Jeffrey raised about exclusions of over age 85.

MEMBER CURTIS: So, there are two parts to that. The first is the cohort definition, and that I think is relevant to the greater than 85, which is similar to NCQA excluding greater than 75 --

CO-CHAIR ROSENTHAL: Right.
MEMBER CURTIS: -- their decision.

But it was more in the what resources are being attributed to this episode-of-care.

CO-CHAIR ROSENTHAL: Oh, you're right. Those are two. One is exclusions from the cohort, and the other is stuff that is either in or not in, once you are in the cohort, right. Those are two slightly different questions.

Open for conversation. Yes?

MEMBER J. RICH: I just wanted to re-up the idea of excluding people who get transferred to a SNF. And I know it is a hard database for you to capture, Kevin, but one of the behavior profiles that you will see when people were being measured for their length of stay is there was a high frequency of transfer to the SNFs to get them out of their facility. So, you wouldn't want to engender behavior that says, all right, if this is a complicated patient, let's just get him to a SNF as soon as we can and it is going to be excluded from
our cost profile. So, let's get him there at
day 29.

CO-CHAIR ROSENTHAL: Or worse, that
you try to get everybody to a SNF. Because
once you get them to a SNF --

MEMBER J. RICH: They're excluded.

CO-CHAIR ROSENTHAL: -- they're
lost.

MEMBER J. RICH: So, the people at
day 29 after an AMI are probably pretty sick,
and most of them could end up in a SNF pretty
easily.

CO-CHAIR ROSENTHAL: Right.

MEMBER J. RICH: So, you don't want
to create the --

CO-CHAIR ROSENTHAL: So, Kevin, can
you give us your response on this point?

DR. LEE: This is Todd Lee. I'll
take that one.

Part of this was a measurement
ability on our end. So, we actually in the
dataset in which we were testing this could
not measure SNF resource use. So, if we include those individuals in the episode that we have currently specified, we would be uncertain as to the impact of that SNF, realizing that may be an important cost center for people in the 30-day period that we're evaluating them. But, right now, it would have looked like a black box to us, or a big black hole, actually, not even a black box, because we just couldn't measure it.

CO-CHAIR ROSENTHAL: Well, but how do you respond to the notion that it ends up, then, excluding, you know, a significant and important cohort? Because I get it that the reason you excluded them was because you couldn't measure it.

DR. LEE: Correct. It may exclude a significant and important cohort. I can't speak to the absolute magnitude of that exclusion criteria. So, I don't know how big it is right now. I'm actually trying to find those numbers, so I can respond to you the
size of that cohort.

CO-CHAIR ROSENTHAL: In other words, you don't know what you don't know. But, Paul, do you want to --

MEMBER BARNETT: Is this cardiac rehab here?

CO-CHAIR ROSENTHAL: No, SNF.


But, Jeff, what's your experience on this?

MEMBER J. RICH: So, inpatient rehab would be the same way. It would be discharged to a different facility. It would be out of the primary facility. You may not get the exact resource utilization, Kevin, within the SNF, but the SNFs are paid under a prospective payment system. So, there is a bundled payment to a SNF. So, you could actually just include the entire bundled payment or prorate it somehow, depending on the number of days they had spent in the
hospital, into your analysis for that hospital.

I could tell you I did the analysis in Virginia for the demonstration project.
Post-acute-care destination discharge varied for a SNF from 3 percent to 19 percent in the State, depending on where your hospital was. Like UVA had a very tertiary care referral pattern, and they sent them back to SNFs because they didn't have control over the patient once they left.

CO-CHAIR ROSENTHAL: Right. Yes, I would have expected some fairly wide variability in that, but it is not trivial. In other words, his question was, if it's trivial numbers, then who cares? But it is probably up to 20 percent of the patients, right?

MEMBER O'NEILL: And it could represent a certain category of folks with a certain level of cardiac-related complications of anoxia or embolic stroke --
CO-CHAIR ROSENTHAL: Yes. Right.

MEMBER O'NEILL: -- and things like that. And if you start taking that subgroup out --

CO-CHAIR ROSENTHAL: Because it excludes the LTACs as well as the SNFs, I assume.

MEMBER O'NEILL: Right.

CO-CHAIR ROSENTHAL: Right. And the LTACs would you get a whole cohort of really high-cost cases.

Yes?

MEMBER B. RICH: I would just emphasize Jeff's point. There's an article I read not too long ago showing that this use of the SNFs is actually increasing, and it is not just a static thing. So, it will become an increasing part of costs we are not going to capture.

CO-CHAIR ROSENTHAL: Anybody else have any observations on this point of who gets in the cohort or what's included in the
bundle, once you are co-horted into this? I
didn't hear any discussion further, Jeptha, on
sort of who is in the cohort, though.

MEMBER CURTIS: The SNF one was
brought up.

CO-CHAIR ROSENTHAL: Was the
biggest one. How about the cancers and the
ESRD, and all that stuff?

MEMBER CURTIS: We felt like those
were reasonably aligned with other measures
that are currently in use and could be
refined, but probably okay.

CO-CHAIR ROSENTHAL: Probably okay.

All right, that's good.

Anybody else want to weigh-in on
this point? Because we've got several more to
get through here. Have we covered this one
adequately at least, that people can decide
whether this is -- yes, I'm sorry, Bruce.

CO-CHAIR STEINWALD: Well, it
sounds like if a patient is discharged from
the hospital, and then within the 30-day
period is admitted to some other facility, then they are excluded. And I don't quite understand why they can't capture those data about subsequent admissions to different facilities. Am I missing something?

MEMBER CURTIS: With the SNF population or overall? They do capture readmissions to other facilities or to any facility. But I think the concern was within SNFs, yes, they could get the bundle payment, but they couldn't get everything else. I think the other resource uses were perhaps invisible to them, if I recall their rationale.

CO-CHAIR ROSENTHAL: So, tell us one more time. So, in other words, if a patient were discharged from my hospital but got readmitted at another hospital, because the data source is the health plan, all hospital days get captured, right?

MEMBER CURTIS: Correct.

CO-CHAIR ROSENTHAL: Okay. So, the
question is, why can't the SNF days get captured?

MEMBER J. RICH: I think Kevin's answer was they don't have access to the data. That is not something that is specified that they can actually capture. They can get the total bundle payment for the SNF, but they won't get the line items, I don't think.

CO-CHAIR ROSENTHAL: Right, but the health plan has it, has the same SNF, the fact that somebody went to a SNF, because somebody is paying the bill, right?

CO-CHAIR STEINWALD: And if they would know how many days they were in the SNF within the 30-day period, it seems to me --

MEMBER REDFEARN: I think that within the validation on the Medstat data, normative database, there may be something unusual about what Medstat captures which is different from what our general commercial carrier would capture.

CO-CHAIR ROSENTHAL: Would you
explain for the group the difference between
that database and --

    MEMBER REDFEARN: Well, the Medstat
Consortium database is just Medstat customers, 
Thomson customers that agree to submit their
data back to Medstat. Medstat standardizes
it, cleans it, and loads it into a database,
and then repurposes it for this kind of work.

    And there may be something unusual
about the design of the Medstat database -- I
don't know. I haven't looked at that stuff in
a long time -- that may limit what they can
see in terms of SNFs.

    MEMBER J. RICH: A SNF, even though
they are under the prospective payment system,
just like the hospitals are for DRGs, they get
a fixed payment, for instance, for CABG. But
every year they do their Medicare cost
reports. So, we at CMS actually knew what
resources were being utilized, and we would
adjust for payments on a DRG basis.

    The same thing happens with SNF.
They have to do their Medicare cost report, and we re-analyze it, trying to adjust the SNF payments. So, I think the data is probably capturable.

CO-CHAIR ROSENTHAL: Well, again, this does get into the question, though, of, what is the data source? Of course, we are preempting this; we will get into feasibility question.

But if the data source is only the Thomson Reuters thing, it may be slightly less feasible than if this were really viewed as health plan data. And, yet, it hasn't been tested in any sort of health plan data source that I can see. But, again, let's not get ahead of ourselves. But I think that sounds like the answer.

But, Kevin, would that be, again, the answer as to how it is that you weren't able to get the SNF data, and, yet, most commercial insurances, and Medicare for sure, capture SNF information, not to belabor this
point?

DR. WEISS: Correct.

CO-CHAIR ROSENTHAL: Okay. Well, that's the answer.

DR. WEISS: We don't have in principle anything from our Workgroup's perspective against the SNF information. It is just, as we were able to develop the measure and test it, we were not able to test it with that information available.

CO-CHAIR ROSENTHAL: Okay. Thank you. Well, that at least explains it.

Steve?

MEMBER PHILLIPS: Yes, I just wanted to clarify, at least in my own mind. Because I heard, also, we are talking about rehab facilities as well.

And is there a number in terms of, or a percentage I guess, of the cases that we are looking at that are going to these facilities that we don't have these costs for?
that's what Jeffrey was saying. I think it is as much as 20 percent. In fact, I think in my experience at our place it may even be a little higher than 20 percent. I think we have 20-25 percent easily.

So, that's what is at issue here, and they may be, and particularly those that go to LTACs, the sickest of the sick. And, therefore, they would be wiped out, and it could, in fact, significantly skew the thing.

All right, I think that we have beaten that one up. And so, let's move on to the next question that I kept track of, which was the degree of validity testing that this measure has undergone and the extent to which that's an issue.

MEMBER CURTIS: Can I just go back one --

CO-CHAIR ROSENTHAL: Yes, yes, yes. Sure.

MEMBER CURTIS: I think we didn't really discuss the completeness of the
outcome, right? So that it is specific to AMI readmissions and not renal insufficiencies.

CO-CHAIR ROSENTHAL: Well, why don't you raise that one one more time then?

MEMBER CURTIS: So, they went through, and, again, we are trying to identify resource use that was reasonably associated with AMI care. And so, they made decisions as to what would and would not count as being associated with that.

The TAP was particularly concerned that that at times appeared arbitrary or, at best, incomplete and I think, also, inconsistent across measures.

CO-CHAIR ROSENTHAL: So, would you give an example? And now you are talking about readmissions, so to hospitals --

MEMBER CURTIS: Any resource use. So, the one that I used an example of earlier was that they don't categorize pharmacy claims for antiarrhythmics --

CO-CHAIR ROSENTHAL: Right.
MEMBER CURTIS: -- as being associated with AMI care.

CO-CHAIR ROSENTHAL: Okay.

MEMBER CURTIS: That just seems to imply --

CO-CHAIR ROSENTHAL: Was there something about hospital use, though, as well?

MEMBER CURTIS: So, readmission for acute renal failure, as I recall -- and correct me if I'm wrong -- would not be associated --

CO-CHAIR ROSENTHAL: Yes, because it is not the primary diagnosis. If it is the primary diagnosis was the way I saw it.

MEMBER CURTIS: Correct.

CO-CHAIR ROSENTHAL: Right.

MEMBER CURTIS: It is not the primary diagnosis. It is one of the codes that they --

CO-CHAIR ROSENTHAL: Okay. So, Kevin, would you all comment on the exclusion criteria of not the patients themselves that
get them into the cohort, but the kind of
stuff that gets excluded as part of the cost?

DR. WEISS: For this, we had a
process, an iterative process, that we worked
through that was data-driven, but maybe we can
have Todd talk about that process.

DR. LEE: Yes. As Kevin was just
alluding to, and as described in our
submission documents, we would provide
feedback to the Workgroup after they had gone
through and specified to us a set of
diagnostic codes to include in the measure.
We would have them look at what's now grouping
to the measure and what's not grouping to the
measure in terms of the imaging procedures
that are done, the other diagnoses that are
happening.

And to take the example of acute
renal failure, that actually would be captured
if there is a qualifying ICD-9 code for AMI-
related care that happened as part of that
diagnosis, as part of that claim. If it is
only for acute renal failure, it would not be captured.

CO-CHAIR ROSENTHAL: But if there was an AMI diagnosis, it was a secondary coded diagnosis, it would get rolled in?

DR. LEE: That's right.

CO-CHAIR ROSENTHAL: Okay. I think that's a little different in your 31-to-365-day measure, correct?

DR. LEE: It is a little bit different in the 31-to-365 where a hospitalization focuses on the primary diagnosis.

CO-CHAIR ROSENTHAL: Okay, but in this one, a readmission, if there's an AMI diagnosis as primary or secondary, the cost would get rolled in?

DR. LEE: Yes, it's codes present in any diagnostic field during the 30-day measurement period for all qualifying ICD-9 codes.

CO-CHAIR ROSENTHAL: All right.
MEMBER CURTIS: So, it is perhaps slightly improved, but, in our view, it is still subject to the vagaries of individual coding and particularly relevant to the outpatient setting, where, you know, what is a physician going to code, how many diagnoses? How consistent are they going to be?

CO-CHAIR ROSENTHAL: Well, but the coding inconsistencies, if we got into that, I would say we're done.

(Laughter.)

We can go home. I mean it's horrific and the variation is ginormous. I think we would be done.

This seems to me to be pretty close. How about the antiarrhythmics? And, then, we will move on.

DR. LEE: Yes, we went through the same process with the Workgroup around categories of medications. You know, I can't remember right off the top of my head why they focused or why they chose not to include
antiarrhythmics. We could go back to the Workgroup and ask for some clarification or at least our Workgroup notes and asks for some clarification. But I understand the TAP's questioning why that is not included as a category of pharmaceuticals that we capture.

CO-CHAIR ROSENTHAL: Okay, but a cohort of world-famous cardiologists sat and opined on which of the pharmaceuticals ought to be in and concluded that arrhythmics didn't need to be in them.

DR. LEE: Yes, and we went through the same process with pharmaceuticals as we did with our ICD-9 codes and our procedure codes, where we showed them what's the most commonly-occurring medications that are being dispensed during this 30-day period and here's what's not grouped into the episode; what are we missing? What should now move into this episode grouping? And that's not one that made the list.

CO-CHAIR ROSENTHAL: Okay. Next we
had the question about admitting diagnosis or
the discharge diagnosis. And, Jeptha, you
raised this one. Can you restate that real
quickly? And, then, the developers can --

MEMBER J. RICH: The admitting
diagnosis often changed based on the patient's
course in the hospital. If you come in with
a broken hip and you have a myocardial infarct
and have to go have a CABG, you will end up
having a discharge diagnosis of CABG.

CO-CHAIR ROSENTHAL: So, is it,
again, at the time of discharge from the index
hospitalization either the primary code or any
code?

DR. LEE: For a qualifying event,
so for somebody to trigger into the episode,
it is the discharge diagnosis, the primary
discharge diagnosis at that index
hospitalization.

CO-CHAIR ROSENTHAL: Okay. So,
Jeffrey is right then. You can have somebody
who comes in with a clear-cut, unequivocal
AMI, gets a CABG, and gets discharged as a CABG?

MEMBER J. RICH: Coronary artery disease.

CO-CHAIR ROSENTHAL: Yes, as coronary artery disease?

DR. LEE: If it is not a 410.XX, then we would not capture them as part of this episode.

MEMBER J. RICH: So, the way you get paid, the hospital's pay is optimized through Medicare groupers. And so, everything that happens to a patient in that hospitalization gets thrown in the grouper, and the grouper spits out the highest payment, DRG, for the hospital, for the benefit of the hospital. And that's what CMS has always taken as a posture. The hospital should --

DR. LEE: So, the DRG could be for a CABG. And, yet, the primary diagnosis could still be an AMI.

MEMBER J. RICH: It might change to
coronary artery disease, though, during his hospitalization.

CO-CHAIR ROSENTHAL: Yes, it is the principle versus -- and I'm not sure why on this one you wouldn't accept this as a secondary code. I mean, even as a secondary code, you are going to end up with an AMI in there.

MEMBER BARNETT: So, presumably, you mean for 10., not 02, because that is a former, prior heart attack, right?

DR. LEE: Yes. Sorry. 410.X, not 2.

MEMBER BARNETT: Yes. And you said if it was in any of the ICD-9 fields, regardless of --

DR. LEE: No. Sorry. That's only for subsequent resource use.

MEMBER BARNETT: I see.

DR. LEE: So, for a qualifying index event, it has to be primary.

CO-CHAIR ROSENTHAL: So, what's
your answer to the concern that, again, you may miss all the cases that have a procedure?

DR. LEE: Well, we had lots of cases in our test dataset that had procedures that qualified under this. I can't give you an answer to the magnitude of potential cases that we missed that may have had a coronary artery disease primary diagnosis and a DRG for a CABG or a PCI.

We did not look at that subset to see how it differentiated. Our Workgroup felt comfortable with the 410.X, not 2, inclusion criteria.

CO-CHAIR ROSENTHAL: Okay. All right, I think we're done with that. I don't know what the answer is, but we're done with that.

(Laughter.)

And, then, the last one that I had on my list was transfers and how they are handled. So, what's the answer to that one?

DR. LEE: Yes. So, transfers,
actually, transfer status becomes a stratification variable for us. If individuals are transferred to another inpatient facility right after their index event, and they are contiguous, then we stratify by people who were and were not transferred as part of our reporting.

The attribution for transfer is attributed to the hospital with the majority of the length of stay. So, if it is a seven-day length of stay and one of the hospitals is four and the other one is three, the resource use is attributed to the hospital that had four days of stay.

And just to give you a sense, when we tested this in our Medicare sample and in our Medstat sample, transfers were under 10 percent of all events.

CO-CHAIR ROSENTHAL: Yes, but they may be 40 percent of all the events in a particular place.

DR. LEE: Agree, maybe, and I am
just trying to give you a sense of overall magnitude when we initially looked at this on the population level.

CO-CHAIR ROSENTHAL: All right.

MEMBER J. RICH: Just as a clarifying, so the receiving hospital where the index event did not occur, if that length of stay exceeds the index hospital, all the costs will be attributed to the receiving hospital?

DR. LEE: That's correct.

MEMBER J. RICH: Including the cost at the other hospital?

DR. LEE: That's correct.

MEMBER J. RICH: Oh, that won't fly. I can tell you that now.

(Laughter.)

And it is really important for a place like our hospital where probably 60 percent of our patients are transferred in from another facility where they have spent a long time and trying to struggle through a
diagnosis and had an AMI diagnosis.

CO-CHAIR ROSENTHAL: Yes, arguably,

with that attribution, well, then, I

underestimated. I said 40 percent. It is in

some places even higher than that.

Yes, I mean, you may be reluctant
to accept a transfer from a place where the

patient has been there for two or three weeks,

arguably.

MEMBER J. RICH: Arguably, but

probably not if that is the way your system

and your community is set up, but the fact

is --

CO-CHAIR ROSENTHAL: Yes, but you

are going to get stuck with all that cost.

MEMBER J. RICH: Right. So, then,
it begs the question. When you do the

analysis, do you stratify, as we were talking
during lunch, do you stratify for hospitals

that have AMIs, but don't have a cath lab,

stratify for hospitals that have AMI and cath

lab capability, and stratify for hospitals
that have an AMI, cath lab, and a cardiac surgery service? Because that last set will be the highest-cost hospital for AMI because they are going to be the ones putting in all the devices and doing the bypasses; whereas, the community hospital with no cath lab or no coronary bypass capabilities, they are going to be a very low-cost center.

So, it will create a little bit of confusion for people who are looking at this transparent data and saying, "Well, I'll go to the lowest-cost hospital." And when they get there, they realize, you know, that there is no facility available or no capability available.

CO-CHAIR ROSENTHAL: Well, that does get to the question of, does this measure, as it is constructed, produce valid -- with the accent on "valid" -- data? If it says this hospital is less expensive than that one, is that believable based on everything that is in here? And at least with regard to
this, you are suggesting that this is a flaw.

MEMBER J. RICH: Yes, I do, and I think it's a flaw in the risk model, too. And I don't know when you want to -- I think that discussion is coming up.

CO-CHAIR ROSENTHAL: I think that is the last one we've got. Well, no, we have got degree-of-validity testing and we have got risk adjustment. So, let's have at both of those.

MEMBER J. RICH: So, I'll begin because Jack asked me this question. The risk-adjustment model I think is short on some important factors. For instance, if this is truly risk adjustment for resource utilization, then one of the variables should be whether you get a PCI and whether you get a CABG because those are huge discriminators between costs for an AMI.

So, it either needs to appear in the risk model and have the risk model redone or else they have to stratify the data, like
they are doing with congestive heart failure and transfer, to include hospitals who do CABG versus those that do not do CABG and are treating the AMI.

CO-CHAIR ROSENTHAL: Comments, then, from the developers on this, on these points?

DR. LEE: Well, we looked at the influence of the intervention and its cost, and you're exactly right that people who had a CABG were more costly relative to those that had a PCI relative to those that didn't have anything.

But we felt like including that in our risk-adjustment model might be adjusting away some of the variability we were trying to capture. At least that is what we heard from our Workgroup, is that this might actually be the choice of institutions. And I am not a cardiovascular clinician, so I may get some of this wrong. I am trying to recall what our Workgroup was telling us.
They indicated that some of this variability might be exactly what we want to pick up with our relative resource use measures, and we so didn't want to include that as part of our risk-adjustment modeling.

MEMBER J. RICH: But your credibility is going to go to zero on this from the hospital standpoint. I mean, if this is going to be the hospital compare for AMI in the newspapers and it will show Hospital A, which just treats AMIs without PCI/CABG, as being low-cost versus my hospital, which is going to be exceptionally high-quality, but exceptionally high-cost because we are providing all the technical backup for treatment of AMI, including left ventricular cyst device and potentially heart transplant patients.

CO-CHAIR ROSENTHAL: Well, yes.

DR. WEISS: This comes to, if I may, a note -- this is Kevin -- that the cost measures by themselves can in all cases be
misleading because they are an incomplete piece of the picture. If one doesn't have quality metrics to balance them, then they will be misinterpreted. I think that was pretty consistent what we heard across our entire project.

So, I want to be mindful that there will always be the ability to misinterpret these.

CO-CHAIR ROSENTHAL: Yes, but I think he is making a more fundamental point, that set aside the quality measures, whether they occur or not, he is questioning whether or not, as constructed, the validity of if a hospital comes out as appearing to be low-cost, that it is low-cost because it simply doesn't have the technologic interventions that are available to the so-called high-cost places. That, in and of itself, will make it misleading. That's the debate.

MEMBER J. RICH: Yes, I know, of course. And come on, let's just fast-forward;
value-based purchasing comes out and you are going to get paid more if you are a high-quality, low-cost center. And all of a sudden, all the high-technology centers who treat AMI and have all the backup technology and operations will appear to fall out of that sort of payment mechanism.

I guess the question is, Kevin, can you risk-stratify when you report results for the different institutional characteristics? AIM without PCI and CABG, AMI at hospitals who have AMI and PCI capability, and hospitals who have AMI, PCI, and CABG capability? That way, you wouldn't have to include it in your risk model. At least in your reporting you would be comparing those three sets of hospitals because they differ very greatly in the way they --

DR. WEISS: It seems very reasonable to look for that kind of a stratification based upon hospital characteristics, particularly if it is only a
three-classification model.

CO-CHAIR ROSENTHAL: Can I assume that when you say you are going to stratify by this transfer question, that basically those cases would be reported completely separately? That's what you mean by stratifying? Or do you just mean that they would be separated into the risk pot?

DR. WEISS: Yes, reported distinctly, that's right. You're right, that's what I was speaking to.

CO-CHAIR ROSENTHAL: Okay. Jeptha, did you have another comment?

MEMBER CURTIS: I think that it is who bears the burden of proof in this case. And I think in this case you are making a compelling case that you have to show us that there is no difference in cost across these characteristics of facilities, right? If the distributions were the same, somehow it was evening out over the course of these 30 days, I think we would buy that, but we haven't seen
that data. So, we are speculating that that's what could be going on.

But I think in your application you said that you couldn't get that data in a reliable fashion because of failure to linkage to AHA or other databases.

CO-CHAIR ROSENTHAL: All right.

And, then, I think the last thing on my list is to hear from our statistician on this because there was an analysis.

And so, Carlos, if you would share with us, in as plain of English as you can for the non-statisticians in the room, what the import of your report is in relationship to the measure?

MR. ALZOLA: Okay. The main thing I noticed was a lot of calibration of the risk score. Ideally, you would want to have the predictive reflect the observed over the full range of the predictive values. So, you want to have that for those for which you predict a high cost of, say, the 95th percentile; you
will likely observe to be about the same value.

So, let's say that for those people who we predict $5,000 cost, and, then, you would like the observed to have on average $5,000, but equally spread between $2500 or $7500. What I saw in your risk scores is that you are severely underpredicting in the high-cost range.

So, that would imply that everybody who has an observed value around the 95th percentile range will be classified as the highest resource use cost.

This is an issue with all statistical models because they don't do well in the tails. But, still, you may not be using all the factors that are driving cost, and one of them could be the type of cost that was mentioned right now.

CO-CHAIR ROSENTHAL: So, in other words, if I am interpreting what you are saying correctly, when you look at the data
statistically, it would appear to support the notion that, in fact, the observed, the expected is underrepresented at the high levels --

MR. ALZOLA: That's correct.

CO-CHAIR ROSENTHAL: -- which is consonant with the observation that was made in the report that Jeffrey made.

MR. ALZOLA: And conversely, at the low end they are overpredicting.

CO-CHAIR ROSENTHAL: Overpredicting at the low end?

Yes, Jack?

MEMBER NEEDLEMAN: Can I ask a question, a clarification? Because this is a regression-based risk model, and it is always going to pull the ends in at the individual level as being predictive. We are always going underpredict the highs and overpredict the lows.

One of the issues I have in thinking about risk-adjustment models in this
context is, should we be looking at the
individual level at which the costs are being
predicted or should we be looking at the
effectiveness of the rollup? When you roll it
up to the unit that the thing is supposed to
be aggregated to, the health plan, the
hospital, and in this case the physician, are
we getting a stable estimate of the actual for
the unit that we are actually doing the
analysis at?

I don't get a good feel from any of
the applications whether anybody is doing the
rollup and actually looking at the stability
of the estimates at that rolled-up level. Do
you know from what they presented whether we
have the same issue when we roll these
estimates up to the hospital level? The low-
resource places are underestimated,
overestimated, and the high-resource places
are underestimated?

MR. ALZOLA: Yes, I think that the
results that are represented are down at the
hospital level. So, that would be the case.

MEMBER NEEDLEMAN: Okay.

CO-CHAIR ROSENTHAL: So, I think the answer was yes. It sounded like it was yes.

MR. ALZOLA: Yes.

CO-CHAIR ROSENTHAL: Paul?

MEMBER BARNETT: Yes, I would just observe, you know, what you said about this problem with the model fit, I think other submissions that we have didn't give us any information about this. So, I would hate to ding these people for being honest about the deficiencies about the models when the other models that we have received haven't told us how well their models performed.

MEMBER REDFEARN: There is one aspect of this that I thought was very interesting. The sample size they are working with from the Medstat is about 11,000 cases. And while that sounds like a big number, that is pretty low for doing this kind of
calibration.

The other thing I thought was very interesting, they are taking the HCC model and recalibrating it. The HCC model predicts total cost. They are changing the calibration for that model to predict their AMI cost. So, they are completely recalibrating a model designed for a different purpose.

And given that kind of a task, I would have been more comfortable with a larger sample size to do the calibration. It is a tough job to do these recalibrations. I have tried to do it myself on millions of cases, and the parameters go all over the place.

It's a tough job.

CO-CHAIR ROSENTHAL: Okay. Well, those were the issues that I pulled out from both the TAP and our own scientific review. Does anybody else have any other scientific issues that they want to raise in relationship to this issue?

MEMBER NEEDLEMAN: Yes. Well, it
gets back to the issue that Jeff was talking
about. I am assuming that patients with more
severe illness should be costing more. They
are getting CABGs as opposed to walking out of
the hospital without any procedures, for
example.

So, on the one hand, knowing the
procedure is telling us something potentially
about the severity of the illness of the
patient, when we don't have other good
measures in the administrative data of the
severity of the illness.

On the other hand, we have got, as
Kevin was saying, we have got this suggestion
that where there is discretion in the choice,
we want to capture the decisions, you know,
the discretionary decisions. Say, if you
chose the high-cost route, we want that to be
reflected in the numbers we are seeing for
you.

So, when I look at the risk-
adjustment model, the question I have to the
clinicians is, do we have enough there to actually distinguish the patients that should be high-cost from the patients that shouldn't be high-cost? So that we actually can then look at the actual resources expended and believe that is a function of discretionary choices in care.

CO-CHAIR ROSENTHAL: Would you like to take a run at that one because my head I am not sure that you can do it based on administrative claims data? The argument becomes circular. And the only way to break through it is you are really going to have to look at different information that is not simply available based on coded information, but I would let somebody take another run at that.

Jeff and, then, Bill.

MEMBER J. RICH: Sure. I think that the one good discriminator they have in there is the cardiogenic shock. That is a huge driver of cost and of mortality as well.
And, then, you get into the diabetes, because those are codeable, all codes in the Medicare claims database. But so is PCI and CABG. So, I think some of the non-clinical indicators, as Jack said, the procedural indicators, to me, the highest-intensity patient will be the one who leaves that hospital with an AMI, PCI, and CABG.

Just like I did a fellow last week. He came in with AMI. He had a salvage angioplasty stent. And, then, I operated on him within 36 hours. Now there is one very costly, sick man.

CO-CHAIR ROSENTHAL: But the problem is, as we know, this PCI and revascularization has at least a certain element of discretionary or gray zone to it. And consequently, the question is, is the procedure indicative of severity of illness or is it an epiphenomenon? And meaning a cost-driver itself.

And I just don't think that the
administrative data is going to provide the ability to distinguish those things. That is my concern.

Bill?

MEMBER B. RICH: Yes, I think the answer to Jack's question is it depends. It depends on the disease and the granularity of the coding system, No. 1. Some diseases have no granularity at all. If you have cardiogenic shock and LID, you can probably impute who is going to be the sicker patient. But some diseases have no granularity. It's not going to be solved by ICD-10. That's only right and left. So, then you are looking at somehow incorporating clinical data, if you really want to get a more robust risk-adjusted model. And I would defer to Jeptha, who knows a lot more about capabilities of administrative databases.

CO-CHAIR ROSENTHAL: All right. So, anybody else have any other scientific questions that they want to put on the table,
other than the ones that we have gone through here in detail? Yes, ma'am?

MEMBER PETER: Yes, I wanted to ask about the observed-over-expected presentation that came up in the statistical report, and whether that is a really significant issue. I thought it was worth discussing.

CO-CHAIR ROSENTHAL: Elaborate a little bit?

MEMBER PETER: Sure. For the expected, they weren't using comparison to an average. They were coming up with some arbitrary or some other benchmark to compare it to. So, I guess a more standard way would be to take the average expected for the peer group and compare it as observed-over-expected for that.

CO-CHAIR ROSENTHAL: Kevin, did you understand the question and can you speak to sort of how you derived the expected mortality for the various cohorts? I think that is the question.
DR. WEISS: For the question, I think we will have Todd, if he is available, take a first crack at that.

DR. LEE: Yes, our O-to-E ratios are individual-hospital-derived observed-to-expected ratios that we then contrast to, at a provider level, we did it with a peer group; in our hospital-level, we did it with all of the hospitals. Again, if we had AHA information, we could have identified like hospitals potentially to do this.

But, then, we looked at different thresholds of O-to-E ratios relative to peers to see what percentage were in the high group relative to the rest of the peer hospitals, which in this case was all hospitals.

CO-CHAIR ROSENTHAL: So, you did do a rather standard identification method for what a hospital's expected mortality was, given its risk profile with the risk-adjusting that you did?

DR. LEE: That's exactly. Sorry,
it is their expected cost relative to their case mix.

CO-CHAIR ROSENTHAL: Right. And, then, the question I posed to the NCQA people was, and I'll ask you as well, how many hospitals end up getting tested out of your 11,000 cases?

DR. LEE: Unfortunately, not that many because we have a hard time with hospital identifiers. And I think Jeptha could say that that is another thing that the TAP pointed out, is that, you know, a limitation of the dataset we were using to test our episode was that it just didn't simply have reliable hospital identifiers on all of the inpatient claims. So, it ended up being tested at about half, I think is what the number ends up being, of the facilities that we have in the dataset.

CO-CHAIR ROSENTHAL: And so, how many would that be?

DR. LEE: I can't remember. I'm
sorry, I don't have that number off the top of my head.

CO-CHAIR ROSENTHAL: Approximately?

Is it 10 or is it 100 or is it 1,000?

CO-CHAIR ROSENTHAL: You know what?

It's not 10. It's certainly more than 100.

I don't know if it gets into the thousands.

I can find that number for you.

CO-CHAIR ROSENTHAL: Well, the secondary question is the other one, though.

Of whatever that denominator is, what percentage fall out as statistically-significantly different, either on the high side or the low side?

DR. LEE: Yes, we have not evaluated it at all of our hospitals. So, I can't answer that, and that was one of the questions that the TAP asked of us, too, is to do some synthesized calculations and power calculations on what we have. And we simply have not had the opportunity to do that yet.

CO-CHAIR ROSENTHAL: Okay. Well,
that says, then, we don't know to what degree
this measure distinguishes in a valid fashion
one hospital from another, right? Okay. I'm
just double-checking my own head.

Yes, ma'am?

MEMBER RUDOLPH: Was there any
thought to using a different database, like
New York State's data, to run the models?

DR. LEE: Well, for our AMI
measure, we actually also tested this in a
Medicare population from 13 metropolitan
service areas. So, we looked at it within
Medicare.

You have to remember we did this
under the auspices of a research grant. So,
we weren't completely at will to test this
across a wide variety of datasets. We had to
work within the constraints of our resources.

Either fortunately or
unfortunately, a lot of what we did was in the
Market Scan database as our test and
development set. And as Kevin noted, it is
now being evaluated in other settings. So, this is going to be an important part of these measures' life cycle as they move forward.

CO-CHAIR ROSENTHAL: All right.

Does the Committee have any other questions about the science?

(No response.)

Hearing none, then I think we will put the question to the vote.

And again, Ashlie, this is yes or no, right?

MS. WILBON: Right.

We have some guidance before we vote on scientific acceptability again that Helen is going to give the group --

CO-CHAIR ROSENTHAL: Okay.

MS. WILBON: -- based on some of the work.

CO-CHAIR ROSENTHAL: Well, you will have to help explain it.

DR. BURSTIN: I will.

CO-CHAIR ROSENTHAL: I was having
trouble following that grid.

DR. BURSTIN: Not only are you the first Committee to go through resource use -- thank you all -- but you are also the first Committee who is using our updated guidance around measure testing and scientific acceptability and evidence.

So, there is a table in your packet that is entitled, it just says, "Evaluation Ratings for Liability and Validity". It is just a two-pager.

The last page of it is a little like a 4x4 table that describes validity rating, reliability rating, and whether or not it actually passes scientific acceptability.

So, if you recall on the last vote -- yes, you've only had one -- this morning, all the ratings were high or moderate, from what the TAP said, and that is reflected in your discussion. So, in general, if you look at this table, you generally rated moderate to high for both of those. And that, therefore,
means a yes, which is, again, consistent with how you voted on scientific acceptability for the last measure.

In this case, you have the TAP's assessment over here on the left. Again, you have talked through many of those issues today that probably, as Jeptha pointed out, reflected some of those lower scores here. I don't know that they have been resolved to your satisfaction.

But, in general, on this table before you, you have a majority of low and some moderate scores, a mix of low and moderate. So, if you look at this table, what you need to do as you think about today your voting, you don't need to go back in and revote on reliability and validity. But I think as you are trying to do this yes/no assessment, you need to feel comfortable that you are at least rating reliability moderate to high and validity high or moderate to make that go forward.
CO-CHAIR ROSENTHAL: All right. I now understand it.

(Laughter.)

Thank you for that explanation. I got it.

MEMBER PETER: I just had a question then. How do you weigh-in the other factors that are in the later parts of 2, like 2b, 3, and 4, 5, and all that? Because that is not validity or --

DR. BURSTIN: I believe what Ashlie has done is actually tried to roll up --

CO-CHAIR ROSENTHAL: No, I think there are two or three that are the subcategories under validity, and there were six under reliability, or vice versa.

DR. BURSTIN: Okay. Right, right.

CO-CHAIR ROSENTHAL: And this is the rollup of all of those. And I guess, according to that matrix, the TAP actually has rated validity low. And according to the grid, a low validity rating trumps everything,
basically, according to the grid.

If you gave validity high or moderate, then depending on the reliability determines, again, the thumbs-up or thumbs-down. So, I think that is helpful because we are not voting on reliability and validity. We just get to vote thumbs-up or thumbs-down, but this is the grid that ought to be in our heads in terms of formulating our yes/no vote on the thing.

So, is that clear?

Thank you. That was very helpful. This is important. Yes, go ahead.

MEMBER CURTIS: I'm concerned.

What am I voting on? Am I voting on the measure that has been presented as we have reviewed or the fact that they have considered the possibility of including the in-hospital deaths and/or transfers to SNF?

DR. BURSTIN: I think at this point you need to vote on it as it is before you. If ABMS can come back, ABMS is welcome to come
back to the Committee, having reflected on
many of the changes you have suggested, and
you will have another chance to reassess
afterwards. But, for today, you are voting on
what is before you.

CO-CHAIR ROSENTHAL: Yes, I am not
sure in my own mind that those couple of
things are really the determining factor about
the validity. Frankly, I think there are
bigger questions about the validity that may
or may not have been addressed.

So, does everybody understand the
grid? It would almost have been easier to
vote on validity and reliability separately,
but I'm not going to suggest that.

(Laughter.)

DR. BURSTIN: I called during the
lunch and said, "I think we need to move
towards voting on reliability."

(Laughter.)

CO-CHAIR ROSENTHAL: All right, but
everybody gets it, and I think most of the
questions that got posed around the table, unfortunately, do have to do with validity, more so than reliability. So, maybe people can have that in their mind. I'm not trying to persuade people on this, but I think you are trying to tee us up so that we vote based on the way the discussion went.

So, I think, with that, a one is a yes, a two is a no. And it's time to vote.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: I'm sorry to say that the vote was 18 against. So, we don't need to consider the usability and feasibility, I understand.

MS. WILBON: That's right.

CO-CHAIR ROSENTHAL: All right?

But I do think we identified some opportunity. I mean the discussion was extremely useful because here's my only editorial for today: I'm sort of disappointed that this didn't pass because this one has a lot potentially going for it. And I
personally would certainly hope that the ABMS folks can go back and address some of the questions that got raised because this would be, this is a really important one, and it would be really ideal to figure out some of the stuff that was raised here.

Jeffrey?

MEMBER J. RICH: No, I agree. I think it is a great measure. It can be a great measure if they go back and find some of the things we talked about. It feels right, it's important, and I think, for resource use, episodes-of-care are a lot easier to tackle than longitudinal. And this will have a lot of importance, I think, in the provider community if we get it right.

CO-CHAIR ROSENTHAL: All right. With that, I think are we ready to move on to the next issue, the next measure, which is 1571, which is the companion to this one, which is acute myocardial infarction episode-of-care for post-acute period days 31 to 365?
I have a feeling this conversation will go a little faster than the last one.

But, Kevin, would you all describe this one for us and your thinking about it?

And, then, Jeptha, we will ask you to comment.

MEMBER B. RICH: Sure. Hopefully, it seems pretty clear that this was meant to take a look at once the patient leaves the acute phase and into the chronic phase of their care for at least the first year, that there was a sense from the Workgroup that there was a lot of opportunity to look at variability in practices, specifically around medication use and diagnostic imaging.

There’s a number of guidelines in terms of how care should be managed in this point. There was a big sense from the group that, in fact, there was a tendency to in many cases overuse periodic assessment, and that there is a real opportunity to assess resource use and actually variability to some
significant improvement.

CO-CHAIR ROSENTHAL: Jeptha?

DR. WEISS: The other part to this was that the time period was very consistent with the ability to look at this in terms of pairing this eventually with quality measures for these patients.

CO-CHAIR ROSENTHAL: Sorry I interrupted.

DR. WEISS: One final note is that --

CO-CHAIR ROSENTHAL: Yes, I interrupted again.

DR. WEISS: -- you will see the issue of attribution here was one where it was directed towards the individual physician. It was believed that, once one got through the acute period where it was system-driven, that a person would ultimately land with a physician or physicians who would take care of their chronic care needs for this condition over this period of time.
CO-CHAIR ROSENTHAL: I'm waiting longer this time. I wasn't sure where the pauses were.

Is that pretty much your summary?

DR. WEISS: Yes, it is. Thank you.

CO-CHAIR ROSENTHAL: I'm sorry. The phone makes it difficult because there's no body language to judge what's going on. I apologize.

Jeptha, now the first item, again, will be importance. So, comments on importance, and, then, we will quickly move into the scientific portion of this.

MEMBER CURTIS: Right. So, I think, again, the rationale for importance is almost exactly the same as it was in the last measure from ABMS. Actually, I think the thought was that this was, as he alluded to, a more interesting timeframe. We are out of the acute period. You are in more stable, where the gray zone effect is more prominent and you may actually be able to detect
differences in discretionary resource use as opposed to being driven more entirely by patient severity.

CO-CHAIR ROSENTHAL: All right.

So, I will quickly, unless somebody has a burning desire to discuss the importance question, seeing none, let's all vote on the importance of the 31-to-365-day heart measure. And it's one, yes; two, no.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Okay. We all think this is important.

Okay, scientific. Jeptha, the TAP analysis?

MEMBER CURTIS: So, again, this is really a paired measure. So, really, the criticisms and strengths and weaknesses of the measure are essentially identical. They use really the same codes to identify the cohorts. They have largely the same exclusion criteria. They do not have, I believe, the same SNF exclusion criteria in this case, which is
reasonable given that it is outside of that first 30-day. Correct me if I am wrong, Kevin.

But, overall, the same things we are applying, somewhat the arbitrariness of the codes that were being used, some concern about some of the exclusion criteria, and there was some concern about using the NCQA exclusion criteria. You know, renal patients, are they really that different that they should be excluded? But, generally, fairly accepted.

And you guys got ahead of me on this because we moved so fast that I couldn't think through everything that we did.

(Laughter.)

I think, overall, though, the reviews were quite similar in terms of the scientific acceptability.

The biggest, I think, hotspot on this particular measure was the attributions at the physician level. And again, it may
have had most to do with the data that they had available to them. But if you look at the attribution, No. 1, when they were trying to get down to the physician level, they made somewhat arbitrary rules as to how to attribute. So, greater than 60 percent of the claims were associated with a single physician. That's who got attributed to all the resource use. If it was greater than 30 percent, but less than 60 percent, you know, it could be attributed to two. And a lot of people didn't get attributed at all. So, there was some concerns with that.

And, then, in terms of the data that they had, there just were missing identifiers. So, they couldn't attribute lots and lots of the individual cases. So, the reliability with which those attributions are being made was suspect.

CO-CHAIR ROSENTHAL: And I am the scientific acceptability reviewer on this one. I had basically the same issues and one new
And, actually, I would assume that all of the issues about validity testing that we reviewed on the last one, including 11,000 episodes, et cetera, et cetera, are also applicable to the longer set of cohorts.

But the very first one I had was the procedures drive most of the cost difference in this cohort as well. And the question is, is this a reflection of illness burden or inefficiency? And so, the question that we posed in the first session arises again, and I don't know the answer.

I also focused on the attribution question. I certainly couldn't argue with the idea that in this chronic phase physicians make more sense than hospitals as the locus of attribution, but I think it was 47 percent of the events could be attributed. And of those, three-quarters had a single provider attributed and a quarter had multiple providers, but there were 4 or 5 percent of
the events that got attributed to ER docs, surgeons, nurse practitioners, and a variety of miscellaneous folks that I thought was probably not terribly meaningful in the context of this.

And the only other new issue that I had arises, I think, in this cohort, but not in the 30-day one, is the issue about transplant. And I think they exclude transplant appropriately. But there is a cohort that gets missed in that, and that is people that are evaluated for transplantation and put on a transplant list, waiting list. And there is no code for that that I am aware of, and, yet, those people basically can be in an intensive care unit in a hospital for nine months, clearly the highest cost drivers, and would not be identified as a particularly high-risk patient in the various modification schemas that exist there.

But, otherwise, I had all the same ones that were identified previously in the
previous discussion.

MEMBER CURTIS: That particular effect would be mitigated to a certain extent by the capping of the cost that would be applied, right?

CO-CHAIR ROSENTHAL: It capped out at what again?

MEMBER CURTIS: I think like 100,000 or so. I can't remember exactly.

CO-CHAIR ROSENTHAL: Yes, you're right, it would. But I would argue that capping it out at 100,000 is way too low. I mean because there, frankly, are patients that would be in their inclusion criteria that could easily use up more than $100,000. If you are in his hospital for a couple of weeks -- (laughter) -- you're going to chew up some big dollars, and that ought to be in there.

But you're right, that would deal with the concern that I raised. But that was my review on the thing.
So, this is now open for discussion from the group. Yes?

MEMBER J. RICH: I agree with those points. I had a couple of other questions. One, was there any discussion of using the E&M codes for attribution rather than cost? It seems to me like people with AMIs end up getting a lot of diagnostic tests ordered, and they are probably the bigger cost drivers over the course of the year, rather than E&M visits, but I may be wrong. This is just a question, and I don't have a big angst about using E&M codes.

But the concern I have here has to do with physician behavior and acceptance. If you start at 31 days, I couldn't tell from here, and the patient is still hospitalized, let's say, for the next 30 days, so those 30 days of inpatient hospitalization costs get attributed to that poor cardiologist who agrees to take this patient when he gets discharged from the hospital. And my concern,
if that is true, is that there will be a huge behavior change about accepting complicated patients who are being discharged from the acute care facility who have been in prolonged hospitalization.

MEMBER CURTIS: I think they did address that in the sense that the 30-day window starts at the time of discharge. It is a nuance to the measure we didn't actually discuss in the previous one. But the clock starts. So, there is that 30 days. They wouldn't be in the hospital at day 30, I think is what your question is.

CO-CHAIR ROSENTHAL: So, if the patient was in the hospital, say, for 90 days, really what is being measured is 91 through 365, or is it 91 plus 365 minus 30? In other words, is it a comparable time measurement?

(Laughter.)

That's the question.

MEMBER CURTIS: Right. It would be 120 plus 365.
CO-CHAIR ROSENTHAL: Your other question that I don't think got answered was --

MEMBER J. RICH: The E&M codes, using the E&M codes.

CO-CHAIR ROSENTHAL: Yes, so this would be for Kevin and the group. Other attribution models use E&M codes and which providers have the most E&M codes to drive who the attribution goes to. Did you contemplate that instead of the cost?

DR. LEE: Ours is actually an E&M-code-based attribution model. It is all around the E&M codes and physician visits. We felt that that was, through our deliberations with our Workgroup, that was the strategy we wanted to go because those are the times the physician is contacting the patient and felt like that individual provider may be the one most responsible for the services that are being used.

CO-CHAIR ROSENTHAL: Right. We
misunderstood that.

MEMBER CURTIS: But, Kevin, the fact that it was missing in 47 percent of cases, is that a reflection of the data that you had available to you or is that a problem that would be present if you applied it in different datasets?

DR. LEE: This is Todd Lee. I don't want me to be misinterpreted as Kevin.

It was a function of the data and the provider IDs that were missing, not the E&M codes that were missing; rather, the provider IDs, for the reason that we couldn't attribute the majority of the non-attributable cases within our dataset.

MEMBER B. RICH: You know, I wonder if you might --

CO-CHAIR ROSENTHAL: No, go ahead.

MEMBER B. RICH: -- expand on that a little bit more? Because that is a problem through all the chronic care ones.

If you are going to eliminate 47
percent — I don't understand how you were missing provider numbers. Could you go into that a little bit more?

DR. LEE: Yes. I mean it was a function of what we had available in the data that we were using to test these measures. The provider numbers were missing in a lot of cases within the dataset.

You know, potentially, this is resolved if this is used in alternative datasets. Because we have not yet tested this outside of the Market Scan database, I can't give you a sense of how pervasive this issue would be in other systems. I doubt if it is as large of an issue, but I don't have any evidence to support that statement.

MEMBER REDFEARN: It is likely to be a problem in commercial databases, too. It depends on what kind of provider ID you want to look at. What we struggle with in California, if you are looking at an individual physician, you have to go down to
the California State Medical License. If you want to get specialty, if you want to use speciality to build peer norms, you have got to be at the individual level. Tax IDs, everybody's got tax IDs. But if you are in a State like California, in which we have group practices, the same doctor can have multiple tax IDs, and one tax ID can represent 1200 physicians, like at UCLA.

CO-CHAIR ROSENTHAL: Yes, and, as I recall, the comparison group ends up being peer-based. It was cardiologist to cardiologist and primary care to primary care, right?

DR. LEE: That's correct.

CO-CHAIR ROSENTHAL: Right. So, the issue of who is a cardiologist would come into play. And, actually, for me, that made me a little nervous about the risk-adjusting methodology because I would assume if the risk-adjusting methodology were robust, you would be able to account for the fact that it
was a cardiologist taking care of the patient
versus a primary care physician. That one
made me a little nervous.

Jeffrey, do you have --

MEMBER J. RICH: A complementary
question. That is, I agree that it should be
physician-level, but I didn't know if it
should be group physician because the delivery
model in our community is that a group of
cardiologists takes care of these patients
longitudinally, including my mother who sees
a group of cardiologists and not a single
individual cardiologist.

So, attributing it down to the
physician level, you may be losing some of
your capabilities. If you group the
physicians together, you may get a more
accurate picture of resource utilization, and
that is occurring within a particular group of
physicians versus another.

CO-CHAIR ROSENTHAL: Yes, but their
database wouldn't identify that the five
cardiologist that are showing up in these
claims fields are all part of the same group
necessarily, would it?

MEMBER J. RICH: Unless you use a
tax ID number.

CO-CHAIR ROSENTHAL: Unless you use
a tax ID number.

MEMBER J. RICH: A tax ID number.

CO-CHAIR ROSENTHAL: Yes.

And can we just clarify, the same
questions that arose in the last -- this is
still 11,000 episodes across "X" number of
hospitals, is that correct?

DR. LEE: That's correct.

CO-CHAIR ROSENTHAL: All right.

Yes, Paul?

MEMBER BARNETT: And the risk
adjustment is just the HCCs prior to their
AMI?

DR. LEE: That's correct.

MEMBER BARNETT: So, there is not
any severity of their cardiac illness or what
procedure they had, or any of that goes into this?

DR. LEE: That's exactly right, and that's one of the reasons we felt peer groups might be the right comparator groups, because we realize there is going to be some severity differences between somebody who is -- there's potentially severity differences between somebody who is managed by a cardiologist versus a family practice physician.

CO-CHAIR ROSENTHAL: Actually, we did a study on this looking at heart failure, and it didn't make any difference at all whether they were a cardiologist. But in a big dataset there may be differences.

MEMBER BARNETT: Yes, but that is totally endogenous to the efficiency. I mean, if your health plan sends everybody to a family -- yes, it is a totally endogenous --

CO-CHAIR ROSENTHAL: We're saying the same thing.

MEMBER BARNETT: Yes.
CO-CHAIR ROSENTHAL: And if you are worried about accounting for that and stratifying it by which doctors they saw, you probably don't have a huge amount of confidence in your underlying risk-adjustment model.

MEMBER BARNETT: So, I mean, the fundamental problem in this whole area is that the things that we really think matter, like are they STEMI, heart attacks, how many vessels are involved, all of the underlying risk factors aren't in the administrative data.

CO-CHAIR ROSENTHAL: Bill?

MEMBER B. RICH: Actually, they are; they are just not captured in this dataset. You know, there are codes for acute MI. There is granularity in the coding. It is just not captured in this dataset.

CO-CHAIR ROSENTHAL: Well, some of what he is saying is accurate and some is complete -- it captures some of it, but it
doesn't capture a lot of the things that you would want to know clinically that would distinguish a really, really sick heart patient from a not-so-sick heart patient.

MEMBER B. RICH: One other question, just a point of information, to go back to what you said, Tom, I didn't understand why that cutoff was 100,000 because I practice in a tertiary care hospital where a great number of these patients are referred in and they routinely have costs more than that. Why did they pick 100,000? Did they explain that to you?

CO-CHAIR ROSENTHAL: Well, let's ask them. Or, Jeptha, do you know?

MEMBER CURTIS: Yes, you would have to ask them.

CO-CHAIR ROSENTHAL: Well, let's ask them.

Kevin, can you explain the $100,000 truncation at the top?

DR. WEISS: For the
hospitalization?

CO-CHAIR ROSENTHAL: Yes. Well, for the whole cost.

DR. WEISS: That is right around the 98th percentile of the distribution.

CO-CHAIR ROSENTHAL: Well, I guess our places are in the 2 percent. That's the problem, all three of our places.

(Laughter.)

We are well-represented; the 2 percent are well-represented in the room.

DR. WEISS: But, remember, this is post-acute. So, this is mostly care happening after that acute event.

So, I mean, I don't know if your patients are $100,000 in this 31-to-365-day period.

CO-CHAIR ROSENTHAL: Well, we get some of them, and that is the point. They exist.

But, Doris, I think you were next.

MEMBER PETER: Yes, I just had a
question about minimum sample size. Since this is at the physician level, I was just concerned about that.

CO-CHAIR ROSENTHAL: A question about the sample size, Kevin.

DR. WEISS: Like what is the minimum sample size?

CO-CHAIR ROSENTHAL: Well, yes -- MEMBER PETER: Yes.

CO-CHAIR ROSENTHAL: -- and do you have enough cases in your database to have gotten it down to an individual physician level accurately? And, then, I guess we will get Carlos' input on this question as well.

DR. WEISS: Yes. Again, we don't come out and recommend an individual, sorry, a minimum sample size necessary. We can calculate that within our database. I don't know how generalizable it is. It is not something we have done to date. That, again, is one of the things that the Technical Advisory Panel asked us about.
You know, one of the things we don't know is what is the minimum clinically-important difference or economically-important difference between groups. I think there is a lot of work to be done with these measures and understanding what the right difference is for being able to determine what a sample size, what a necessary sample size would be.

CO-CHAIR ROSENTHAL: So, in other words, at this point in time we don't really know --

DR. WEISS: Yes, that was a very long-winded answer to say we don't know yet.

CO-CHAIR ROSENTHAL: Carlos --

DR. WEISS: And we don't have a response to tell you what we believe our minimum sample size should be yet.

CO-CHAIR ROSENTHAL: Okay. All right. Thank you.

Carlos?

MR. ALZOLA: No, the point is correct. The real point is, a sample size for
what? What is a clinically- or financially-
significant difference? Once we determine
that, then we can determine, estimate the
sample size to determine what the standard
deviation is.

CO-CHAIR ROSENTHAL: And did you
test this for skew, like you did the previous
one?

MR. ALZOLA: No, I did not.

CO-CHAIR ROSENTHAL: No, you
didn't?

MR. ALZOLA: No.

CO-CHAIR ROSENTHAL: Okay. Are
there other questions, aside from the ones
that have been raised up until now, that we
want to discuss or get input from the
developers?

Bill?

MEMBER B. RICH: To follow up on
Dolores' question, and that was going to be
part of my presentation tomorrow, even though
you are not recommending any specific sample
size for the physician or the group, if you
look at your dataset that you analyzed, and
you are down to 47 percent, how many were
attributable just to the number of physicians
that you looked at?

DR. WEISS: How many of our overall
episodes were attributable?

MEMBER B. RICH: No. How many per
doc?

DR. WEISS: Oh, what's the range of
attributable episodes for a physician?

MEMBER B. RICH: Correct.

DR. WEISS: Yes. Again, I am going
to have to apologize. I would have to dig
that number up. It ranged anywhere from 1 up
to 50, 60, 70.

In our example report, our sample
report that we have here, for example, the
physician that we grabbed randomly had 21
episodes.

CO-CHAIR ROSENTHAL: Jeffrey, do
you have one other?
MEMBER J. RICH: I have a clarifying question. Is the $100,000 cap for the inpatient index hospitalization or for the following year, the following 365 days?

DR. WEISS: Yes, that's about the 90th percentile during the followup period. There's also during the index hospitalization, but that doesn't count in this episode.

MEMBER J. RICH: Okay. So, I want to pull a Bill Golden here. I want to bring this back to 35,000 feet and ask a question. (Laughter.)

So, if we paired these two measures, and we are really trying for the healthcare delivery system to figure out how much it costs to take care of patients, both acute hospitalization and longitudinally, and we have a gap for the sickest patients that truncates the measurement of resource use at 30 days and doesn't pick it up until they leave the hospital, what are we accomplishing for the healthcare delivery system for the
sickest patient population that we take care of?

I mean there is a huge gap between those two measures, and it is not relative to either measure. It is just the way they are specified.

And I don't know if I have an answer.

CO-CHAIR ROSENTHAL: We will accept that as rhetorical, but Bill may have the answer.

(Laughter.)

MEMBER GOLDEN: No. I have a question for the Committee, the Technical Committee.

Was there any discussion about cutting off catastrophic cases at some limit or something, that there was such an outlier that they become distorting?

MEMBER CURTIS: Maybe I'm wrong, but I think that is what the $100,000 cap represents, is an attempt to minimize the
chances that a single case would skew the
sample for the payer, or whatever.

CO-CHAIR ROSENTHAL: Others? Yes,
Steve? Steve?

MEMBER PHILLIPS: I have a question
about the patient who kind of disappears from
the physician's office until they now suddenly
have another event and are admitted to a
hospital. I mean I guess I am wondering how
they are attributed here because it seems like
that is one thing that we would want to get
at, is where the patient, you know, there is
no encounter until they have an event again.

CO-CHAIR ROSENTHAL: Kevin, did you
follow that question? It sort of addressed,
it is asking about people that are lost to
followup or semi-lost to followup or lost to
followup until something hideous happens.

DR. WEISS: Yes. It is all based
on the number of E&M visits, codes, that they
have within the database. If they see a
provider shortly into the 31-to-365-day period
and, then, don't have any followup care that results in a claim with an E&M code on it, and then have a rehospitalization, you know, 320 days later, the way that our model is specified, it would be attributed to the doc who the patient saw shortly after the beginning of the period.

CO-CHAIR ROSENTHAL: Yes, I assumed that that was how a few ER docs got to be the attributed physician.

DR. WEISS: That's absolutely correct.

CO-CHAIR ROSENTHAL: It is almost that exact scenario. And, suddenly, they show up in an ER, and that's the only E&M codes they got, and the whole business gets attributed to an ER doc.

MEMBER PHILLIPS: Yes, which raises some question. I mean, should they be the attributable doctor or the one who hasn't seen them up until that event?

CO-CHAIR ROSENTHAL: It is hard to
know who you don't know, which is the
challenge in a lot of this.

Jeptha?

MEMBER CURTIS: So, just to, again, maybe bring it up a level or maybe not, one of the things that we have really focused on in the TAP is this attribution, and that is kind of essentially where we stopped our evaluation because we got so hung up on it.

And one of the questions that I think is worth reflecting on with this group is, if this were a different target, if it were a medical home or an accountable care organization or some other categorization of patients or rolling up patients into a larger group, and you get more stable estimates, some of these problems about outliers sort of disappear as you get increased case numbers.

Is this, then, a more reasonable measure at that level? Is it just that they are proposing to apply it to the level of the individual physician?
CO-CHAIR ROSENTHAL: All right.

Any other scientific questions that haven't been posed or thoroughly discussed?

Yes, ma'am?

MEMBER RUDOLPH: Well, I suppose this is a usability question, but it is sort of, how is this measure designed to be used? Is it designed for quality improvement, for public reporting?

Obviously, if it comes to endorsement, we make the assumption that it is designed for public reporting. And that sort of, in my mind, raises the bar a bit for making sure the attribution and other things are really on target.

CO-CHAIR ROSENTHAL: I think your description of that is exactly correct.

MEMBER RUDOLPH: Okay.

CO-CHAIR ROSENTHAL: By definition, it is for both. And consequently, the bar is as high as it exists in any of our minds for what is necessary to be accurate for both of
those uses.

Any other scientific questions?

Yes, David?

MEMBER REDFEARN: I have a question for the developer. Rather than just calculating an observed-to-expected ratio and, then, for example, doing a confidence interval around that, they do something a little differently. They calculate the percentage of the ratios exceeding 75 percent of the peer group. I just wondered why they chose that particular methodology.

DR. LEE: Yes, it's a fair question. It is not a methodology that has been evaluated in terms of a benchmarking or performance measure.

After we had gone through this exercise with several of our Workgroups, they asked us, "So, can you help us differentiate the sort of high resource users from the non-high resource users in these episodes, in these example reports?"
And so, we chose a 75 percent threshold. Again, there is not a lot of strong rationale as to why that is the right benchmark, the fact that it sort of began to differentiate the sort of individuals that had a higher-than-expected proportion of O-to-E ratios above that threshold.

CO-CHAIR ROSENTHAL: All right. Do we have enough information about the science of this to make a judgment or do we need any further conversation?

(No response.)

Okay, I think we have got enough.

So, I will re-refer us to the grid, and I will, also, then, re-refer us to now the TAP scores that are behind us. And actually, it is interesting, this one didn't score quite as bad as validity, but I think we identified a few validity questions today that perhaps the TAP didn't, frankly, quite get to.

But, interestingly, on this one, this one skews negative on reliability. And
again, according to the grid, low reliability also gets you a negative score. So, either low reliability or low validity gets you a no. So, the same kind of thought process in factoring both of those factors into your vote applies to this one, as it did the last one.

And so, let us -- it is, again, one, yes, and two, no. So, let us vote.

(Whereupon, a vote was taken.) Somebody is making up their mind.

There we go.

Okay. So, the vote is in. Zero, yes; 18, no.

So, we do not need to discuss usability or feasibility.

But, again, I think like the last one, I think the group is really enthusiastic about these measures, despite the votes. Again, I will editorialize, but I am getting the sense from the whole group, it would be really wonderful to have a few of these things worked out and these measures resubmitted.
And a couple of the sets of advice that I heard that might be particularly useful is the idea of trying to get other data sources than the one that was used because of some inherent difficulties in that database that might be remedied with some larger and more robust datasets that could probably remedy a few of the things that were significant issues in the discussion.

Paul?

MEMBER BARNETT: There is also a national registry of cardiac cath data.

CO-CHAIR ROSENTHAL: I have got to believe they know about that, right? And maybe it doesn't have all the stuff in it, but who knows?

MEMBER B. RICH: To follow up on Barbara's point, there is robust literature out there to look at minimum sample size at the physician level. Bill Thomas in Maine has published extensively on this.

And since this is available to
public reporting, I would like to see some
discussion about the sample size. Obviously,
if we are down to 27 now, that is going to be
an issue. So, it would be nice so we feel
comfortable if we get a measure that addresses
the scientific and reliability and validity
questions, that that is part of the
discussion. But there is a robust literature
out there.

CO-CHAIR ROSENTHAL: Well, and the
one other thing I would add on this one -- I
know everybody is trying to not get to the
break -- but the idea of some doctors are, in
fact, just individual doctors and need to be
analyzed at such. But today, fortunately,
lots do practice in groups. To have a
methodology that would allow either for an
individual attribution or a group attribution,
because in those groups, frankly, in our place
the peer pressure of the group is way more
powerful than one guy being called out who
then, in fact, says, "Well, those weren't my
patients and I am just going to ignore it."

Frankly, we don't really care. We look at the whole group and say, "You guys are not doing good, and we don't care which ones of you did it. Figure it out."

And so, the idea of being able to have the possibility of doing both by using these administrative datasets, and if it is looking at tax ID numbers, or however the methodology, I think that would be another powerful aspect of the thing.

But I think, with that, unless anybody has any further comments on this, I think it is time for a short break. And our break is scheduled for an hour and a half.

(Laughter.)

Ashlie, how much? 2:45, okay, a 15-minute break. I'm going by the thing. I'm going right by the thing here. Ashlie did correct that earlier on. I apologize. Sorry.

Okay, about a 15-minute break and, then, we will reconvene.
Oh, and when we come back, we are going to do 1572 from tomorrow, another cardio measure. Well, it is a good thing somebody asked what we are doing.

So, you have got 15 minutes, Dolores. Good luck.

(Laughter.)

I know you were planning on doing that tonight, but now you can have a drink at dinner. It will even be better.

(Whereupon, the foregoing matter went off the record at 2:32 p.m. and went back on the record at 2:53 p.m.)

CO-CHAIR ROSENTHAL: All right, what is on the agenda for this afternoon, we will start with 1572, which is episode-of-care for management of chronic coronary artery disease. This is an ABMS measure.

And if we have time, depending on how we are able to grapple with this one, we hopefully will have time, also, then, to do 1604, which is another HealthPartners
measure,, which I believe is the companion to
the HealthPartners measure that we already
considered as a group on the extensive phone
call that we had. So, that is what we hope to
do this afternoon.

So, Kevin, are you guys still on
the phone with us?

(No response.)

Oh-oh.

DR. STROUPE: I am Kevin Stroupe,
who was also a measure developer for this
particular measure.

CO-CHAIR ROSENTHAL: Oh, terrific.
So, thank you for sticking with us -- we
appreciate it -- and enabling us to move
forward with this measure this afternoon.

Would you mind giving us a little
summary of this one? And I think a suggestion
was made that perhaps you can identify for the
group the ways in which this one is similar to
the two previous ones, and I am talking
similar sort of methodologically, and possibly
ways in which it is different. And that compare and contrast might facilitate the group's ability to understand and make a good decision about this one.

DR. STROUPE: This measure was developed to examine resource use and cost associated with the management of coronary artery disease over a one-year period. The patients were identified with a diagnosis of CAD during a 12-month, one-year period prior to the measurement year, and, then, measurement resource use and cost are assessed during the measurement year.

So, this is a measure looking at a chronic condition. So, we are trying to assess the resource use and care that occurred during a one-year period of time for these individuals who had been previously identified in the prior year with coronary artery disease.

As with the other ABMS measures, an inclusion criteria includes having continuous
medical and pharmacy benefit enrollment preceding the measurement year and during the measurement year in order to have adequate data available to examine the population with this condition.

In addition, for this specific condition, we were looking at individuals whose age was greater than or equal to 18 years of age. And, then, we identified patients who had a diagnosis using ICD-9 codes for coronary artery disease.

Exclusion criteria, then, were in the year prior to the measurement year having acute coronary syndrome, acute myocardial infarction, or having a prior revascularization through either a coronary artery bypass graft or through percutaneous coronary intervention.

In addition, there were exclusion criteria that had been used throughout the ABMS measures based on prior NCQA work, including active cancer, end-stage renal
disease, organ transplant, HIV/AIDS, and, then, for this particular measure, vasculitis.

So, in terms of this particular measure and how it would be contrasted with the other ABMS measures, we are using -- the inclusion criteria in terms of identifying this specific patient population would be the coronary artery disease would be unique to this particular measure, as well as the exclusion criteria, the acute coronary syndrome, AMI, revascularization as exclusions.

So, defining the particular patient population that would be of interest or applicable for this particular measure would be unique to this particular measure. What would be similar with the previous ABMS measures includes the fact that this was developed using and tested with the same dataset and, similar, the costing methodology was applied similarly throughout the ABMS measures. So, it would be similar to the
prior measures in that respect, as well as the risk-adjustment approach would be similar.

However, for the particular measure, a unique function was developed based on the input from the Clinical Advisory Workgroup that was involved in the development of this measure. So, the overall process, the Workgroup process and the development process was similar across the measures as well.

There was an in-person meeting of clinical advisors who provided input on the particular aspects of the definition of the population for whom we should be looking. And then, based on that, their initial input of the conditions and the other types of care that we should be looking at to define that, the resource use during the measurement period, the development proceeded, then, with identifying the specific codes to address the conditions that they indicated to be measured, as well as the procedures and the medications, and so forth.
And, then, through an iterative process, the data were tested using the Market Scan data, and information, then, that was obtained was provided back to the Work Group, who then looked at, evaluated the information to determine whether there were additional conditions or coding and that sort of thing that should be incorporated into the measure. And, then, that was reassessed then and retested using the Market Scan data.

So, although we would have a different coding for a different condition and different ICD codes and different CPT codes that were identified for the relevant procedures and diagnoses, and so on, a similar iterative process was used with the prior ABMS measures.

And, then, again, the costing, a similar costing methodology was used for this measure as well as the prior measures.

And, then, as far as the risk-adjustment procedure, the Workgroup Committee
members would through that process identify particular conditions that were of interest. And, then, the model was developed. Then, their feedback was obtained, and so on, for the final risk-adjustment approach that was specified.

So, basically, there were similarities along the way in terms of the methodology, but this particular measure would be unique in the disease state that was examined and the particular codes, health conditions, codes, procedures, medications, and so on, that were identified as being relevant to measure for the episode-of-care of coronary artery disease.

CO-CHAIR ROSENTHAL: All right. I think that is a good summary. Thank you very much.

DR. LEE: Oh, this is Todd Lee with ABMS. Because Kevin didn't have the advantage of bring on the earlier
conversation, I wanted to sort of follow up with a bit of context that might help to drill us down to a very what's different level in terms it is really the disease is identified in a year, not based on an acute event, and, then, as Kevin noted, followed chronically forward.

The methodological issues were exactly the same across the two episodes, including attribution.

CO-CHAIR ROSENTHAL: All right, thank you very much for that summary. We much appreciate it.

Jeptha, the TAP?

MEMBER CURTIS: Yes, I think really this was one of the first ones we went through, but, overall, it is regarded the same.

So, with regards to importance specifically, I think that the same criteria applies.

CO-CHAIR ROSENTHAL: All right. I
goofed again.

We have to vote on importance. So, with no further discussion, a one is a yes and a two is a no, and then we can get to the scientific questions.

(Whereupon, a vote was taken.)

Now we've got 19. Fantastic. We are getting better every time. And it is important, unprecedented vote.

Did we have 19 people here for the earlier things? Okay, I'm taking your word for that.

Welcome back.

(Laughter.)

Okay. So, it is important.

Now scientific discussion.

MEMBER CURTIS: As they outlined, I think, overall, the approach is almost identical to what was taken before. As you can see, there are kind of these four complementary measures that they have tried to develop to capture kind of almost stratifying
this cardiology population. So, am I early, am I late, chronic disease with and without revascularization. So, this is the chronic coronary disease without revascularization.

The one thing that I think is particularly notable about this is the code used to identify the population is 1. It is 414.XX. And it is simple and straightforward, but it carries with it the assumption that every patient with chronic ischemic coronary disease is going to have that particular code. And it has some face validity to it, but there wasn't a lot of confirmatory evidence to suggest that that is capturing everybody, as opposed to using other ways to identify this population.

CO-CHAIR ROSENTHAL: If they were going to try to confirm it, what would they have to do to do that?

MEMBER CURTIS: Well, I guess you would wonder, for instance, off the top of my head, if you have a patient who then undergoes
PCI in the index year without that code, and without a diagnosis of chronic or acute coronary syndrome, whether or not that patient had it before, or if you went to the year prior and explored it in the year prior, in any given 12-month period of time, how reasonable is it that you are going to have that code documented? Obviously, people don't lose the chronic condition after 12 months.

CO-CHAIR ROSENTHAL: Right, but you have to look at codes before or after or something to try to find out why it dropped off, or you would have to do chart reviews of some sort to actually confirm it, yes?

MEMBER CURTIS: Yes. Yes, something like that.

CO-CHAIR ROSENTHAL: Okay.

Anything else from the TAP?

MEMBER CURTIS: I think nothing that we haven't already discussed.

CO-CHAIR ROSENTHAL: Okay. So, in other words, the various methodologic things
that we discussed in the previous ones are relevant to review of this one or not?

                  MEMBER CURTIS: I think everything is relevant. I think the major difference is the numbers that they had available to them for derivation and validation were significantly higher than with the MI measures.

                  CO-CHAIR ROSENTHAL: All right. We will ask them in a moment what that number was.

                  Dolores, I think you were our internal reviewer. Sorry, we only did give you 15 minutes' notice on this, but I am sure you have copious notes from before.

                  MEMBER YANAGIHARA: As they said, there are a lot of concerns that came up with the other two that still exist. I have a question about if the 414.XX was sufficient to get the full population. So, I think those were the biggest things.

                  I think that the validity and
reliability testing did look more robust in this particular case. I didn't have a chance to dig into it in detail, but it seemed like Carlos' summary -- I don't know if he is still here or not -- but that it looked like he felt like that was much better than the other ones, but with some reservations as well.

CO-CHAIR ROSENTHAL: Kevin, can we ask you what the number of episodes were on this one, and, then, their sort of distribution, like we talked about on the other ones?

DR. STROUPE: The testing for this, initially, 308,000 were identified, CAD patients were identified. And, then, after applying the exclusion criteria, there were 108,000 patients, then, that were identified in the denominator of the measure, then, as it was tested.

CO-CHAIR ROSENTHAL: And, then, I am looking at slide 21 from the packet. And it looks like there's a slightly higher number
that were attributable. I think the previous
one was 47 percent and this one is 57 percent.
So, a little bit higher, and like the other
one, three-quarters of the attributions are
attributable to a single provider and 26
percent to multiple providers. There wasn't
a slide on this one that I could see that told
you which kind of doctors the things got
attributed to, but maybe I missed it, vis-a-
vis the attribution question.

All right. So, with our two
internal reviews, or our internal review and
the TAP Committee, let's open this up for
discussion.

MEMBER PENSON: Can I ask two
questions, primarily of Jeptha I think,
because I'm obviously not a clinical expert in
this?

But on the bottom, No. 1, what we
know about these measures compared to the
other ones, they are constructed the same way.
Do you feel that this one is able to overcome
the problems of the other ones or are we still in the same place? I mean, basically, I am asking you to tell us how we should vote, based on the way you painted the program before.

(Laughter.)

But, frankly, I think it is a very valid question, and it may save us a lot of time, too.

The other question is, assuming you say, yes, it is acceptable, could you just say a few words because I didn't push it before because it was pretty clear to me the discussion wasn't going that way, but, you know, the risk adjustment in all these things is kind of hinchy to some degree, and it is a new risk-adjustment methodology. The HCC, you know, they are testing it. Did the TAP feel comfortable with the risk adjustment?

MEMBER CURTIS: So, I think that the issues are slightly different than they were on the previous ones. I would say,
overall, that for the chronic conditions we
felt, as a group, more comfortable with these,
but not super-comfortable. I don't know if we
can put up the summary scores.

So, I don't think that we,
personally, as a group, I don't think we had
sufficient confidence to say that this should
go forward from the TAP perspective. But the
issues were slightly different. It was partly
the attribution and partly the fact that you
couldn't get 50 percent of the cases to be
attributed to a single or multiple providers
There was still that concern that I still have
about the arbitrariness of the designation of
codes that are related to chronic CAD or not.

And so, we scraped off some of the
really big ones. Like discharge to SNF, that
is not an issue here, but you are left with
still some things that are terribly
concerning.

You know, we talked a lot about the
risk-adjustment methodology. I think we felt
that it was difficult to assess in the application specifically because they talked about how they developed 18 different models and then selected the one that had the best characteristics, but there wasn't a lot of detail on that. And I think Carlos had referred to that in his review as well.

They had subsequently come back with kind of more information about the models that they selected, but I didn't have that information for this measure specifically. So, I can't comment as to whether or not it was really suitable.

There are limitations to any administrative risk adjustment. I think, speaking to Bill's point from before, you know, yes, it's not clinical, but we have shown, at least our group believes that you can risk-adjust using administrative claims data as long as you validate it against a chart-based model or a gold standard model.

In this case, they haven't taken
that step, but, as proof of concept, yes, you
can fairly risk-adjust to the hospital level
using administrative claims data.

    I will leave it at that.

CO-CHAIR ROSENTHAL: Carlos, do you
want to comment, then, on this?

    MR. ALZOLA: Yes. No,

unfortunately, there wasn't any information
for me to evaluate the risk-adjustment model.
Like Jeptha said, they said 12 models, but no
details were provided.

    CO-CHAIR ROSENTHAL: So, the one
that had the skew problem was the 30-day one,
and that is the only one that you
identified --

    MR. ALZOLA: Only the AMI models
had the detailed information.

    CO-CHAIR ROSENTHAL: Yes.

Did that answer your question?

MEMBER PENSON: Well, I appreciated
Jeptha's candidness, too. So, yes, it did.

    CO-CHAIR ROSENTHAL: Okay. Other
questions?

CO-CHAIR STEINWALD: I have one.

CO-CHAIR ROSENTHAL: Yes, please.

Of course.

CO-CHAIR STEINWALD: The over 50 percent that can't be associated with a primary care doctor, now the assumption is that these patients are actually having visits. So, what is lacking is an ID, right? And I heard around the table that this is a common problem.

Is there a reason to think that it is a source of bias as well as missing information? Or is there a reason to think that it is not a source of bias?

MEMBER O'NEILL: It could be, I guess, a source of bias in that it would be a characteristic of an organization or a system to have missing data elements. I mean I think that there are some systems that are more reliable in terms of making sure all the data is present. Don't you think that's true?
CO-CHAIR ROSENTHAL: Bill, do you want to weigh-in on this?

MEMBER B. RICH: Well, I learned a lot talking to Joe and Barbara. They might want to elucidate this.

We have all been waiting for these data aggregation groups, the value exchanges, and they explained, quite well to me anyway, why there has been a big holdup.

MEMBER RUDOLPH: Yes. Actually, the National Association of Health Data Organizations, which works with the All-Payer Claims datasets, has identified this as a serious problem for doing any physician-related reporting, and is partnering with the Centers for Disease Control to send a letter to CMS requesting that CMS really begin an initiative to find a true, unique patient identifier, not an identifier that has embedded in it location and other kinds of things.

So that there would be one ID,
provider ID -- did I say patient? -- I'm sorry, provider ID that would have be unique to that provider. And that is what would be used in claims databases.

Until that happens, it is a big problem. Individual states and other sort of multi-state claims systems are having to come up with their own provider directories, et cetera, build them from scratch. And it is a big problem.

CO-CHAIR ROSENTHAL: Joe, do you want to elaborate on that? Or would you?

MEMBER STEPHANSKY: In Michigan, we have been trying to have some physician-level reporting on our hospitalization data, so that an individual hospital in this case can see where else a particular physician is referring patients. It is only partially successful.

It remains a real issue, and it is extremely expensive to maintain. We are constantly updating those lists of physicians, and there are constant ones that are falling
MEMBER REDFEARN: WellPoint is working on a process to impute provider, unique provider IDs. We have software to do that, and just like we are doing that for members, to keep track of members.

Because you can have a member who comes in under a Social Security number and, then, they go out and they come back as a spouse under a different number. So, we impute, are trying to impute IDs for members and are doing the same thing for physicians.

MEMBER J. RICH: So, Barbara, I have got a question for you. What about the NPI? Where does that come in here? I mean I have an NPI. Everybody has an NPI.

MEMBER RUDOLPH: Unfortunately, some physicians have multiple NPIs if they work in a number of different clinic locations, et cetera. They will have an NPI that has them appearing here and one over here, and then you have to verify whether, in
fact, that is the same physician, which is problematic because then you have to go to state license and tapes, et cetera.

And so, the NPI does not help us.

MEMBER J. RICH: But it helps some?

MEMBER RUDOLPH: Some. Some.

MEMBER J. RICH: Some?

MEMBER STEPHANSKY: But there are still a lot of errors in that data if you assume that a doctor is only supposed to have one ID. It doesn't work out that way.

MEMBER J. RICH: And is that true with electronic payment claims? Is that true in general?

MEMBER STEPHANSKY: Yes, even claims submitted to a single payer, they have difficulty sometimes. They will have multiple NPIs for a tax ID or multiple tax IDs for a single NPI.

MEMBER O'NEILL: And there are databases that were set up before the NPIs, and there's not always fields. I mean, you
know, it may not have a field for physician ID.

MEMBER STEPHANSKY: Or the field will be one digit too small.

MEMBER O'NEILL: Yes.

MEMBER STEPHANSKY: And, then, you've really got problems.

(Laughter.)

MEMBER PETER: Can NPIs also be at the group level or not, at the individual level?

MEMBER RUDOLPH: Yes.

MEMBER STEPHANSKY: Some were created that way, yes.

MEMBER RUDOLPH: Some were, uh-hum. So, it is really a complex process to try to figure out who the physician is.

MEMBER REDFEARN: But this error rate seems a bit higher than what I have seen in my experience.

CO-CHAIR ROSENTHAL: Well, again, is that some function of the fact that this is
a culled or a combo dataset that has been
extracted from other datasets that might
accentuate that?

MEMBER REDFEARN: Very likely, it
is sort of lowest common denominator --

CO-CHAIR ROSENTHAL: Yes. Right.

MEMBER REDFEARN: -- when it
consolidated. So, that's right.

CO-CHAIR ROSENTHAL: Okay. Other
points of discussion on scientific validity?

MEMBER NEEDLEMAN: Yes, I have a
question. I am trying to understand the
population here and the exclusion of patients
who in the identification year have some kind
of revascularization or have a heart attack.

We are talking about a chronic
disease here, somebody who had that heart
attack two months before the identification
year or had revascularization two months
before the identification year is going to be
in the group.

Does it make sense to exclude these
patients or stratify on these patients?

MEMBER CURTIS: It only makes sense in the sense that there are four measures that are all complementary. So, I think if you take all four of the ABMS coronary atherosclerosis measures collectively, really, very few people drop out. So, you have got, again, the MI, early MI, late. You have got chronic with revascularization, chronic without revascularization.

So, in any given 12-month period, throughout all these four measures, you should capture just about everybody, with the proviso that the specific codes used for inclusion may or may not be comprehensive enough.

MEMBER NEEDLEMAN: So these measures can't stand alone?

MEMBER CURTIS: No. Well, I would argue that they cannot.

CO-CHAIR ROSENTHAL: A companion question to that, and I'm not sure it is germane to answering the question about
scientific validity, but this measure is somewhat similar to what NCQA showed us earlier, at least in intent. Are there substantive differences in inclusion and exclusion criteria? My mind can't work fast enough to sort of track those, but --

MEMBER CURTIS: I think the only main one is, again, that assumption that everybody is captured by the 414. That is the biggest difference.

And so, getting back to the point that was raised earlier, is this closer? I would say it is substantially closer.

CO-CHAIR ROSENTHAL: How is NCQA getting them, again?

MEMBER CURTIS: So, again, that was, because it is not four different measures, it is one measure, so they could enter based on history of AMI, history of --

CO-CHAIR ROSENTHAL: Oh, that's right, they had multiple triggers. They've got the multiple cohort. That's right. Yes,
yes, yes, yes.

Do people feel like we have discussed this thoroughly enough in the context of the others and that we have guidance from the TAP on the direction that they are advising us?

Oh, here is the reliability and validity. Well, this scored a little better.

But your answer to Dr. Penson --

MEMBER CURTIS: Again, that was my personal answer, as I have tried to distinguish it from the TAP.

CO-CHAIR ROSENTHAL: Is it possible -- and I do mean this, because, again, part of our goal I think as a Steering Committee is to pay some deference to the TAP. You guys have spent really deep dives and a lot more time than we are. So, we probably, as a general rule, probably should not substitute our judgment for yours.

But I do get the sense that we uncovered a few things in the methodology that
perhaps might not have been the focus of the TAP. Is that fair to say or am I overstating it?

MEMBER CURTIS: I think we pretty much covered the same things that you covered. I don't think there are any major differences.

CO-CHAIR ROSENTHAL: Okay. All right.

MEMBER CURTIS: We didn't take into account necessarily could you consider this measure in isolation, which I think, by the nature of the fact that it is submitted in isolation, you would have to think of it by itself: is this capturing what they wanted to capture and is it providing a good view of the care of these patients?

CO-CHAIR ROSENTHAL: Right. And your answer on that question?

MEMBER CURTIS: Personal answer.

CO-CHAIR ROSENTHAL: A personal answer.

MEMBER CURTIS: I would say that it
is, again, close, but there's enough problems
for me that I would not --

CO-CHAIR ROSENTHAL: Okay. All
right. And that is why I am belaboring this
just a little bit, because the second MI one
kind of, not completely obviously, but fairly
obviously followed the first one. This one
has some subtleties to it that warrants us not
just immediately knee-jerk going it's the same
as the other one. So, that is why I am trying
to be respectful and not just sort of rush to
judgment on the thing.

Bill?

MEMBER B. RICH: Just a quick
question for Jeptha. In what sequence was
this code in the order in which you considered
codes at the TAP?

MEMBER CURTIS: I can't remember.
We can look that up. My recollection was that
we did one of the chronic ones after we did
the second MI, but I wouldn't --

MS. WILBON: Are you asking --
CO-CHAIR ROSENTHAL: Now we are trying to see, if I do a meta-analysis of the TAP --

MEMBER CURTIS: Based on the numbers of the reviews that are available in this rolled-up part, I think this might have been in the phone call, the followup, but, again, I am having a hard time separating this from the related measure of chronic CAD with revascularization, which was, I think, the second measure that we reviewed.

CO-CHAIR ROSENTHAL: You can adjudicate that factoid in your head any way you want.

MEMBER PHILLIPS: Tom, I was wondering if --

CO-CHAIR ROSENTHAL: Yes, sir?

MEMBER PHILLIPS: -- we could maybe hear the measure submitters' response, if they care to, whether this could stand alone.

CO-CHAIR ROSENTHAL: All right. I think that's fair.
Kevin, standalone?

DR. STROUPE: I think that the intention was this was, in particular, looking at a population who was in a stable management phase of CAD. And so, that was, in particular, why the exclusion criteria for the previous AMI or the previous revascularization, that that might be capturing a less homogenous population.

And so, from that perspective, this was intended to be a standalone measure, where we were looking at specifically patients with CAD and sort of a stable, chronic management portion of their condition, and, then, looking sort of subsequently at what care and cost they accrued during a 12-month period.

CO-CHAIR ROSENTHAL: So, it sounds like your answer would be that (a) you believe that the measure could stand on its own, but it is interesting, the contrast is relevant, I think, with the NCQA one, in that, in fact, revascularization and prior events were key
triggers in the prior year to getting included in the cohort that we identified this morning. I mean maybe it is okay to exclude them, but we would have two measures purporting to measure the same thing that would, in fact, have quite different cohorts.

I don't think that because we approve the other one means that we have to be necessarily consistent in approving this on that basis, but this one would pull in a different cohort.

Other questions? Yes, sir, go ahead.

DR. STROUPE: The intention of the Clinical Workgroup that was involved with the development of this was, as I said, for a patient population that would have been in a more stable management phase. And so, that is why those other conditions that would have indicated that they might not be necessarily in a more stable management phase of their condition were to be excluded.
However, it should be noted, though, that certainly that one-year period, if the individuals did have a revascularization or something, their disease progressed to the point -- that that would be captured as part of the measure.

CO-CHAIR ROSENTHAL: Okay. Thank you for the clarification.

Any other questions about the science? Comments? Jack?

MEMBER NEEDLEMAN: As a non-clinician, I am heard some concern about restricting the inclusion to 414. I didn't hear any of the other clinicians in the room comment on that. All I've got is the voting from the TAP. So, that seems to be the biggest issue here.

So, I would like to hear some discussion that would help inform my decision on that.

CO-CHAIR ROSENTHAL: Well, let me ask the question, Jeptha, is it your sense
that is the key methodologic issue around
the science or are there also issues about
attribution and a variety of other factors?

MEMBER CURTIS: I think there are
issues along every step of the pathway, but I
think it starts with the code. And they made
a decision to go with one restricted code in
contrast to what was taken by NCQA, which is
trying to get to a comparable or somewhat
similar population by using the other code.

So, if you look at that list of
codes and you contrast that with 414, I am not
sure if you are really capturing the full
spectrum of chronic coronary disease patients
clinically. That's my sense.

The second piece is, again,
decisions to apply or to attribute subsequent
care to chronic CAD or not chronic CAD based
on a list that they did, we had concerns about
the completeness of that list and the
arbitrariness of that list.

If you look at the packet they
submitted with it that showed the diagnoses
and the codes and the related costs of related
and non-related care episodes, then it sort of
highlighted that. It was closer, but it
wasn't -- there was one that caught my eye in
terms of lipid testing I think lots of times
is related, but a lot more times it is
actually unrelated. To me, that is completely
related. That was just one thing that threw
that to the forefront of my brain, that,
again, this is not a perfect way of
attributing whether or not it was related to
the CAD.

And, then, you get into the issues
of attribution, which although improved
because of the size of the dataset, still are
equally problematic as they were for AMI. If
you can't attribute 50 percent of the episodes
to a single provider, how are you going to
characterize provider care? From my
perspective, that is impossible. That becomes
probably the single greatest problem of this
measure.

   Now, again, could you do it at a higher level? Yes, but that is not what we were asked to evaluate.

   CO-CHAIR ROSENTHAL: Yes, and I think one of the things in my head, just trying to compare and contrast, one of the critiques of the NCQA one, of course, was all of the errors that get built into sort of stuff that happens at day 364, they solved the ambiguity by rolling everything in and saying, well, we will report it at the health plan level.

   Here you are trying to make judgments about what's in and what's not in. I think they probably made as good a set of judgments as anybody is going to make. But, nonetheless, they are still subject to that we need the right set of stuff to have in and out. And as you get out to 364 days, it starts to get very fuzzy as to what really is related to the episode-of-care and who is
responsible for it.

And we heard very significant concern around this generalizable kind of method in relationship to applying it at any level below a health plan. And yet, the intention here is to apply it to individual doctors.

I think it is the combo of facts. And to tell you the truth, though, on the question of whether 411.XX -- is that which one it is? -- 414.XX, whatever, it probably is less a clinical issue than it is people here that are really familiar with coding and accuracy of coding. So, I would perhaps defer it to somebody from a health plan.

I don't know. David, you have insights, either Penson or --

MEMBER REDFEARN: No insights. I mean all I can say, in general, is that diagnostic coding has improved across time. But, then, for example, we don't pay based on diagnosis. We pay based on procedure codes.
And anything that is not related to payment
tends to be lower quality.

CO-CHAIR ROSENTHAL: And there was
a choice made in picking that, and I think,
Jeptha, you are saying, but correct me if I'm
wrong, that this is the part that linked all
of these together because they made a choice
about how to incorporate these that were based
on a sort of combo of the measures. Do I have
that right? Is that what you're saying? Or
am I missing --

MEMBER CURTIS: Understanding that,
as a clinician, you have discretion. Like you
could arbitrarily go with 414, you could go
with 413, you could go with 411, you could go
with 429. You know, you just have this range
of --

CO-CHAIR ROSENTHAL: But the
developers made a choice in picking 414.

MEMBER CURTIS: They made a choice
to go with one that was very specific.

CO-CHAIR ROSENTHAL: Right.
MEMBER CURTIS: And they had a rationale for it. I think it was that it was simple and that it was the most commonly used. But it raises the issue of the completeness.

CO-CHAIR ROSENTHAL: Okay.

MEMBER B. RICH: I had the same concerns that Jeptha did. It is the restriction to the one code.

But, also, one of the goals of looking at groupers is the decreased variation eventually. And there is a great deal of variation. When CATs are done and PCIs are done, if you just throw them out, you fail to address possible variation or deviation from ACC guidelines.

Was that discussed, Jeptha, within your group, just eliminating PCI? Out of the Chronic Care Group, did you guys talk about variation or these always changes in patient population and sicker patients?

MEMBER CURTIS: I'm sorry, I was having a sidebar with Ashlie. So, I missed
the first half of your question.

    MEMBER B. RICH: Is it appropriate
to eliminate all PCI patients from chronic
care? Is there enough variations and
indications within groups? I am not a
cardiologist, but I do read the front page of
The Washington Post.

    MEMBER GOLDEN: I think you are
referring to the COURAGE study.

    MEMBER B. RICH: Partially, but
that was for ICDs, wasn't it?

    MEMBER CURTIS: Certainly, there's
like the patients with chronic coronary
disease, you get revascularized. And I think
it made sense as a paired measure, right, to
take it in isolation without this suite of
measures going forward. And it sounds like
they have actually withdrawn the
revascularization one for other reasons. This
one makes less sense to me. But, again, I
don't know if we can consider the broader
scope or if we are stuck in this is the
measure that we are evaluating.

CO-CHAIR ROSENTHAL: Yes, Doris?

MEMBER PETER: I just had a quick question about 57 percent of the data are missing provider IDs. Do we know what percent of the costs that represents, the missing costs that are not attributed?

CO-CHAIR ROSENTHAL: Kevin, I think that question will be for you. Can you identify, of the episodes that are missing and not attributable or unattributable, is that a proportionate amount of the cost?

DR. STROUPE: I don't have that particular, the number, directly at hand. I would assume that would represent a substantial portion of the overall cost.

CO-CHAIR ROSENTHAL: You would guess that perhaps it is at least in the relative range of proportionality. There isn't any reason to think that the ones that you can't find are either more or less. But I think the answer I heard was "I don't know."
And that's fair. That's perfectly fine.

Bill Golden?

MEMBER GOLDEN: To follow up on Bill's comments, COURAGE, there are a number of -- in fact, Washington Medicaid and your Technology Commission is looking at it -- the number of people without symptomatic angina getting stents.

And so, I guess the question for you or to ABMS is that you could probably remove your revascularization of your stents with a co-morbid diagnosis of unstable angina. But if you take them out when they are being put in for asymptomatic coronary disease, that would be potentially confounding what you are trying to measure.

MEMBER B. RICH: That's what I was trying to express. Thanks.

MEMBER CURTIS: Correct. So, if they were paired with a revascularization measure, it makes more sense. In isolation, I think it loses --
MEMBER GOLDEN: Does it? Because if they are paired, you are still not determining whether the revascularization was for symptomatic disease or asymptomatic disease.

MEMBER CURTIS: That's true, but --

CO-CHAIR ROSENTHAL: None of these purport to measure that. And that is the $64 question on much of the stuff related to stents and CABGs, right, Paul?

MEMBER BARNETT: Yes. So, this differs from the NCQA in really just a couple of dimensions. One, it is more inclusive, excuse me, less inclusive than NCQA, right? And, then, the other is that it attributes to the physician rather than the group. And otherwise, the case mix, it is still the HCC is the risk adjustment, which doesn't really include, as far as I understand it, the severity of cardiac disease. So, otherwise, they are really quite similar.

And the specificity I don't think
is, rather than having this very broad
category, is that big a deal. I think the
thing that bothers me about it is the
physician attribution and the problem with the
data. And that is going to be true with any
measure that we look at, evidently, we are
learning, that tries to attribute to
physicians, if the data is not there in any of
the systems. And so, maybe that is kind of
the key problem.

CO-CHAIR ROSENTHAL: Well, we
actually have approved the scientific basis of
a HealthPartners one that attributes at the
individual physician level. So, we have one
exemplar where that is doable.

MEMBER BARNETT: Well, so that is
an interesting question, which is, if somebody
comes to us with a measure that is developed
with a dataset that is good, and then we
endorse it, and then we have to apply it to
the real world where there is no such data,
that is an interesting problem.
CO-CHAIR ROSENTHAL: Well, it is an interesting problem, but in that one, I mean in the one, and not the preempted, because we are going to get to another one of theirs, but there was a real dataset in the real world.

MEMBER BARNETT: Yes, but I am just saying maybe their dataset was special. And I think that may also be true of the NCQA data because, you know, they have this Audit Department that evidently is part of the process.

CO-CHAIR ROSENTHAL: Yes, I think that is a very fair point.

I do think the contrast, though, between this and the NCQA one is not insignificant because, if we remember this morning's discussion, several of us were rather militant about the idea that that one only made sense in the context of it being applied at the health plan level.

And the other difference is they have five years of real-world experience of
actually measuring that thing and applying it, where here there is no real-world experience of applying it. This one is a pure Gedanken experiment, and not that that is automatically disqualifying, but it seems to me, in relationship to the myriad of several other problems -- I don't think just because we did this morning's with NCQA that we would say there's minor differences and, therefore, this one can go through on the basis of our being internally consistent. Because I think they are pretty substantial, those two differences are, to me, pretty substantial.

Doris?

DR. WEISS: Kevin Weiss on the phone. If I could perhaps just note an important difference that hasn't been reflected?

CO-CHAIR ROSENTHAL: Yes, please.

DR. WEISS: Sure. Hi. And I apologize, I had to step away from the call
because of a scheduling conflict.

From what I gather on this one, it is important to keep in mind that the Workgroup were very clear that they wanted to look at, the best that I can describe it, the meat-and-potato person with hypertension. They were not trying to take all people with hypertension, recognizing that there is so much variability in that. I'm sorry, CAD, I apologize.

The other part was that they wanted to get very specific with cost. They did not want to look at total cost. They thought that there was so much noise in a total cost measure that it was really not actionable in any sense. I think that may have reflected the discussion you, as a Committee, had earlier today. And so, they went for a condition-specific cost.

And those are two of the key constructs of this measure. That is, to look at this same population and to try to make it
as relevant to disease in terms of cost

attribution or cost inclusion, I should say,
because attribution has a different meaning in
the context of our conversation here.

I hope that that is helpful to
create the clear distinction in why this
measure is a different measure than NCQA.

CO-CHAIR ROSENTHAL: Yes. Thank you. I think those are relevant points.

Jack?

MEMBER NEEDLEMAN: Yes, one comment on the discussion, and, then, I have actually got a question for the developers.

Somebody made the comment that the HCC risk-adjustment model here was like the NCQA model, and it isn't. This uses the HCCs to identify categories that then get put into a regression-based model to estimate the weights on each of the relevant HCCs for the patient. And the NCQA is using the HCCs to group patients into different tiers and then look at empirically costs in the standardized
plan or across all plans for each of those
tiers, and then reweight those average costs
against the actual cost of the plan.

So, very different risk-adjustment
models here and very different concepts behind
each of them.

CO-CHAIR ROSENTHAL: For those of
us that are maybe not as grounded in the
subtleties of that, could you give us some
flavor of what the implications of that are?

MEMBER NEEDLEMAN: Well, as I
understand it -- I am basically learning from
reading this stuff -- the HCC model that NCQA
uses says let's use these weights and we will
figure out which weights apply to which
patient. And there is a hierarchical
component to that which is common to both
systems.

But let's create the weights.

Let's get a total weight for each patient.
And, then, we will group them into tiers.

Then, we will look at the average cost per
patients in each of those tiers across all the plans we have using the standardized costing. And for the average weighting across the tiers for the plans, we can now get an average weight for our average plan.

What they are doing with the plans is they are getting the costs within each of the tiers for the plan they are rating. Then, there is a proportion of patients in each of those tiers unique to the plan which is different from their average.

They reweight their average plan cost for the percentage of patients in the tier for the plan they are studying to get the expected cost and then take the ratio of average actual to expected for the plan. That is my understanding of what NCQA is doing in their risk adjustment. Because they have got a price for each of the tiers, there is some opportunity there, I think, for less compression than you see in the regression-based models.
What these folks have done, if I have got it -- and correct me if I'm wrong -- is they have identified for the patients what they think are the relevant HCC categories and included those in a regression model of the costs for the patients, standardized costs of the patients, to get a standardized adjustment to the expectation for each patient, rather than saying, what tier are they in and what is the average expense to the tier?

So, it is a very different model of risk adjustment. I like regression-based risk-adjustment models. I find it a perfectly fine one. But it is very different from what NCQA is doing, and if NCQA is using the standard CMS weighting model, it is very different from the standard CMS weighting model, and we should appreciate that as we move forward.

CO-CHAIR ROSENTHAL: Kevin, do you guys want to make any comments on why you did it the way you did it?
DR. WEISS: If Todd is on the phone, he can help us here.

DR. LEE: Yes. I mean we took this approach largely under the direction of our Technical Advisory Committee, thinking that when we drove this down to the patient level, we wanted to implement these using these regression-based models, so that an implementer would be able to, hopefully, take our regression weights and apply it to their population and be able to calculate these observed-to-expected values at an individual patient level.

CO-CHAIR ROSENTHAL: All right. That makes sense.

And I don't think we remembered to ask the same question that we asked on others of, how many episodes, approximately, on average, per physician ended up getting attributed in this run, in this model?

Obviously, you had more episodes, and I am assuming approximately the same number of
physicians or perhaps even a few more. Can you give us a flavor of what the average number of episodes per doc ended up being?

DR. LEE: I don't think we have that number right at hand again. I am trying to dig it up as we talk.

CO-CHAIR ROSENTHAL: All right.

DR. LEE: I don't know if Kevin Stroupe has that number close.

But I think the answer is we don't know for sure.

CO-CHAIR ROSENTHAL: All right. I apologize for asking such a detailed thing, but it is sort of relevant from before. I guess we could do a back-of-the-envelope. If it was 20 per physician before, and there are double the number of episodes and slightly more physicians -- 10 times as many episodes? Okay. So, then, it would be 200-ish. Well, you can't do it. That's not right. That's not right because that was for MI and it was per hospital, and we have no clue what the
number of physicians is.

Never mind. Sorry. I was trying to do the back-of-an-envelope to help, but it was no help.

(Laughter.)

MEMBER NEEDLEMAN: Tom, I did have a question.

CO-CHAIR ROSENTHAL: Okay. Yes, yes, for the developers.

MEMBER NEEDLEMAN: Right. So, my question, can you explain how you -- I have read the description, and I'm not getting it. So, can you explain how you do the standardized costing for the inpatient component of the care you are looking at?

DR. LEE: We follow the same model that NCQA described this morning where it is a DRG-based model. We actually use the NCQA price weights for our inpatient costs, standardized costs, where they are available.

So, if it is a DRG that groups to our episode, and based on its length-of-stay
category, and NCQA has a price for that DRG, that's what we used. If they did not, we developed our own by averaging the DRG payments within our dataset and creating a standardized price for that DRG, which is divided based on the length of stay.

CO-CHAIR ROSENTHAL: All right. I would like to suggest that, unless there is any other burning scientific question that we have not pounded our heads on, that it is time to get the clickers.

Jeffrey, last comment?

MEMBER J. RICH: Just a quick one. I thought the physician attribution and the coding issue were the two biggest ones. But just on the coding issue, you mentioned a bunch of other codes that weren't used, but could apply. Did they look, if they included those, how big would the population grow to from 308,000? Are we losing a lot of patients by not using those?

MEMBER CURTIS: I don't think they
provided that information.

MEMBER J. RICH: And I don't recall; how did the NCQA measure get to the group level? And we can't get there here. I know we are talking about individual physicians and not being able to code for it in our last conversation. But how did they achieve the group-level identity in the NCQA measure?

CO-CHAIR ROSENTHAL: It's a health plan. It's a health plan. It's a health plan; it wasn't a group. It's a health plan. They know who Blue Cross is.

All right. If there are no other pressing issues that we have not thoroughly discussed, I think it is time to call the question. And so, again, the same grid applies between reliability and validity. And you've got the TAP ratings here in front of us. And again, low in either one gets it out.

So, we are voting now again for -- this is yes and no, right, Ashlie? Okay, it's
yes and no. Sorry. I'm wearing out at the end of the day.

Okay. So, one is yes, two is no, and it's time to vote.

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: And we have 19. Two, yes; 17, no.

I think, again, with some discomfort.

CO-CHAIR STEINWALD: I would like to add something. If this missing ID problem is going to be really an endemic problem, then it seems to me -- and this could have been a factor for me -- it sort of has to be treated like non-response bias in a survey. You know, you really need to demonstrate that the missing IDs don't constitute a source of bias, if it is possible to do that.

In the absence of that, then I think it is hard to vote yes.

CO-CHAIR ROSENTHAL: All right. I think we are done for this measure for today,
and I hope we will see this again.

And with that, do people need to stand up at their chair for five minutes? I think we certainly want to do intellectual justice to 1604. So, perhaps a five-minute, stand at your desk and do jumping jacks, or something, for a few seconds. And, then, we can do another hour's worth of work on the last measure for today.

(Whereupon, the foregoing matter went off the record at 3:51 p.m. and went back on the record at 4:02 p.m.)

CO-CHAIR STEINWALD: Let's begin.

The people from HealthPartners, would you reintroduce yourself, and I understand you have a slide presentation for us?

MS. KNUDSON: Yes. Thank you, Bruce.

I am Sue Knudson with HealthPartners. I lead our Health Informatics.
Along with me is Chad Heim from Health Informatics at HealthPartners as well.


Well, good afternoon.

I'm Sue Knudson with HealthPartners. I lead our Health Informatics effort.

MR. HEIM: And I am Chad Heim, Senior Director of Health Informatics.

MS. KNUDSON: So, yes, we did prepare some slides for you today. We have six slides just to provide a brief overview.

And in talking with Bruce on the break, because this is a new measure, as you might recall, when we went through our resource use measure that we have already vetted, we had initially filled out the application with a companion measure of total cost of care. And so, that is what we would like to review with you today, is that separate measure.
Just by way of just a reminder of a background of where we are from, HealthPartners is a consumer-governed, nonprofit, integrated healthcare delivery system in Minnesota, which means we operate a health plan. We own and operate care delivery in terms of a large multi-specialty group as well as a large hospital and some smaller, community-based hospitals.

HealthPartners also operates in a market that is an open-access market, which means from a health plan product point of view, we do not work with assignment. So, members aren't assigned to us. So, we are very similar to other markets in that regard.

So, that is just a little bit of a background.

Also, a reminder, our submission is for the commercial population. So, this is a population-based measure.

So, if you could advance to the next slide, Ashlie?
Oh, very difficult to see. So, let me just walk you through this.

We wanted to give the framework for where this total cost-of-care measurement comes into play. So, if you could read the single box out to the lefthand side, it would be titled, "Healthcare Value".

And with that, the top portion of the slide where you see the three rectangles, that second one in is quality. In quality, we have got two domains, one of clinical quality measurement and the other of patient experience.

And so, why I wanted you to have this context, it is in how we use this measurement and how we propose its future use. So, we do not use the total cost-of-care measures or the resource measures standing on their own, but we use it in combination with those quality results. So, really more in terms of a Triple Aim view of performance.

The other thing, in the larger
rectangle on the bottom -- oh, and one more thing before I move past the quality. What we do is subset those into domains. So, in the clinical quality, we look at acute and preventative care. We look at care for chronic conditions. We also look at health information technology use and safety measures, performance.

And in the experience domain, feedback and information around care and communications with patients and members, as well as access to care.

So, then, that larger rectangle on the bottom, this is where this measure comes into play. The darker blue that is a subset is the resource use component. So, that is the component that we have already gone through with you over the few hours of conference call meetings that we have had.

I should also say it is very nice to put some faces to names and voices.

(Laughter.)
So, that broader box, then, is where this measure of total cost of care comes in.

And so, on the break, Bruce had asked if we could point out, as many of you may have planned to do some reading on this measure tomorrow in anticipation of this discussion taking place tomorrow, to really highlight what the differences are with the resource use measure.

Well, the key, and really only main, difference is that this total cost-of-care measure does not employ a standardized pricing methodology. It is actual cost, but expressed as an index. And so, we get into that a little bit more.

So, as we know in the previous discussions, for a resource use measure, the way we are viewing it is that does require an approach to standard pricing; whereas, this does not.

I guess the only other thing, just
observing the previous discussions, if it is at all helpful, is on the importance realm.

In our previous application as well, a lot of the citations that we had noted all refer to we are really fortunate to have several measures in the quality and experience domain, fewer, if any, thus, the work of this Committee and the NQF to have standardized measures in this realm. So, we really see that as kind of the third leg of the stool in filling out that Triple Aim.

Next slide, please.

So, the next couple of slides I am going to go through really what we have outlined in terms of specifications, and the next slide more so in terms of what we have teed up in terms of what are our guidelines, as it relates to this measurement approach.

So, first, the specs. This is an illness burden adjusted per member per month, which, as we are measuring it, the smaller font is showing that some may refer to this as
allowed. And just to be explicit, we are saying it is both what the plans are liable for as well as any patient or member liability. So, it is inclusive of both of those pieces, simply divided by the member months, which is your membership over a 12-month period, if that is the study period.

So, what we want to emphasize is that this is a measurement that is really standard in the communities already. Many stakeholders are routinely measuring this from health plans to consultants working on behalf of purchasers, et cetera.

In its core, it uses administrative claims data as well as eligibility data and a risk-adjuster, as we have previously discussed in our other measure as well.

And so, our comment on the risk adjustment, it is key for it to be robust as well as capture disease prevalence of a commercial population.

In terms of the population-based
measure and cost, it, again, is all care for
the population being managed. So, that is
inclusive of all inpatient care, outpatient,
professional services, those of the group who
might be primarily responsible for the
patient's care, as well as any of their
referral partners, pharmacy, and any other
ancillary services.

And so, what we do, as well as what
I mentioned before, is we are displaying this
as an index for benchmarking. And so, that
computation is simply we will refer to a total
cost index, and it is simply the risk-adjusted
PMPM. Our unit of analysis is the group level
divided by the peer group risk-adjusted PMPM
to get an indexed rate.

So, just to highlight again one of
the values of a measure like this,
particularly at this unit of analysis, or,
frankly, even at a plan level, would be that
it takes into account not just care for those
folks with chronic disease, but it also takes
into account effectiveness in terms of it expressing itself with cost as it relates to prevention. So, if you have effective prevention programs in terms of keeping folks healthy and ensuring optimal life, either through disease management programs, other interventions, that you really get credit for that by way of those members and patients being evaluated on their costs as well.

Next slide, please. Thank you.

So, this slide, just reflecting on some of the previous discussion we have had, again, our attribution method that we have put in guideline is at the group level, but reflecting on previous Steering Committee discussion that we have had on our phone calls, we wanted to talk about the unit of analysis a bit.

And so, what this is illustrating is that there can really be two different levels. On the left side of the diagram, we are really illustrating that that unit of
analysis could be at a plan level, the community, or a regional level, and it could include the full populations. In those applications, really, this measurement could be done without attribution. It doesn't require attribution there.

Where we have been using the measure is on the right side, and that is in attributing to our provider groups. And so, again, in our market where we have an open-access, non-gatekeeper market, we are using this attribution model. And it assigns a member to the provider with the largest portion of office visits during that measurement period.

And we are finding, as I had mentioned before on the calls, that this synchs up very nicely with what our medical groups are finding as they reconcile to their medical records.

Next slide, please.

So, another guideline area that we
wanted to talk about, although reflecting again as well on the previous decision in the other measure, that it went forward with our risk adjuster that we are using, which is the Hopkins ACG method. But we wanted to again just illustrate our guideline recommendation around risk adjustment.

We are using ACGs, as I had just mentioned. It has been in the commercial market, the public, and the research settings, and has had numerous peer review journal articles over the last 20 years.

Since our last discussion by phone, we have augmented the micro-website that we have made available to you all with a technical guide from Hopkins as well as easy links to get to this information, so that you are comfortable with the transparency around this tool. And there is easier access to that information.

We also wanted to note that ACGs was reviewed alongside 11 other commercially-
available risk adjusters by the Society of Actuaries, all resulting in similar predictive accuracy.

And so, our guideline in our application is really to say we are using ACGs because it is standard in our community, our local Department of Human Services as well as Department of Health. We have a history as a payer, as well as the other payers in our community, using ACGs. So, our community is used to that model. But knowing that the Society of Actuaries tested it along with the others, our assessment was any of them could be applied, given their robustness in that testing.

And, then, one other development that I think I mentioned on the last call, but just to reiterate, is back in May Johns Hopkins did announce that they will provide a free version of the ACGs to the health information exchanges under contract. So, that was a new development as well.
So, what we have done on the slide deck is provide, within the micro-website, a specific link to where these new tools are for you to take a look at.

Next slide.

So, these remaining couple of slides are just reflecting on previous discussion around transparency. So, we wanted to share with you how we are using this in the transparency realm.

So, every year we provide performance information to the providers in our network. So, this is a snapshot of that. And so, this is a summary of how I described in the first slide our Triple Aim approach to measurement and evaluation and assessment. And so, this just kind of gives you a snapshot of what some of that detail might look like. Again, this is just a little sliver of that information.

But in the upper left is a high-level assessment of the clinical quality
information. On the right on the top is that patient experience information. And, then, the box on the bottom illustrates the transparency on the overall cost.

So, consistent with that discussion, it may be difficult for you to see, but you can see that we are using an icon approach. So, we use stars in terms of quality. We use dollar signs for the overall cost assessment.

And if you could go to the next slide, please, Ashlie?

This is just a little snapshot from our website to show you how that is drillable. So, depending on the user, someone may be interested in overall cost. But if I am a consumer, maybe I want to just drill into something that is specific.

And so, this is just to illustrate that this is out there at the group level, and you can click into the detail and see all the individual measures and the performance behind
those as well.

So, this was just to give you an example of that transparency and how it is actually displayed.

Is there one more slide? Is that it? That's it? Okay.

Thank you.

So, that is sort of just a key difference, to give you a feel of how we use it. Not only in transparency, but the other thing I would say is we have a very collaborative approach to this. And so, we work pretty directly with providers, not only our own. I was chatting with some folks on the break, letting them know yesterday I spent at least three hours with another group in the Twin Cities, not our own, kind of going through this data around improvement opportunities.

And so, it has really been an opportunity for us to have dialog in a collaborative environment around where there
might be some practice opportunities and opportunities for systematic improvement.

So, with that --

CO-CHAIR STEINWALD: Thank you.

Before we get to importance and other criteria, are there any questions for HealthPartners about the measure itself?

Yes, Bill?

MEMBER B. RICH: Just out of curiosity, what tools are you using to collect patient satisfaction? Are you using CAHPS surveys, and how are you collecting the data?

MS. KNUDSON: Yes, that's a great question. Historically, we have had a health-plan-specific survey, but in our community we are using Minnesota Community Measurement. We are just closing out a pilot on having standardized CAHPS. And so, that is a great source because that means everyone in the community would then be using the same result.

MEMBER B. RICH: Are they collected remotely by telephone or are they done at the
MS. KNUDSON: You know, I will have to follow up on that question. I don't know that specific.

Do you know, Chad?

MR. HEIM: No.

CO-CHAIR STEINWALD: I have a question about index construction. So, your total cost measure is reduced to an index and then compared to a peer group. So, any variations in input costs should be factored in that peer group comparison, is that true?

MS. KNUDSON: We are benchmarking to our plan average. And so, that is really the basis. And so, if the unit of analysis is if a health plan is doing this, it is understanding variation among the groups within that plan.

Would you add to that, Chad?

MR. HEIM: Yes, that's correct.

MEMBER BARNETT: So, how are you going to compare Alabama to, say, Boston, if
they have quite different salary structures?

Well, it is the question of I don't see how

the geographic --

CO-CHAIR STEINWALD: Go ahead and

answer that. However, geographic variations

in the costliness of care are factored into

the index construction.

MS. KNUDSON: Well, let me take a

shot at that, and Chad can augment my answer.

So, where we have done the testing

is within our plan at the group level, and

that is what our submission is on.

But reflecting back on that

attribution side, it is not what we have

tested by way of this submission, but there

could potentially be, first, if you have

access to that data and it is clean and

scrubbed, you could use a national database

and compute if you had a database with

allowable or PlanPlus member liability and

simply compare. There's no standard pricing

here. And so, really, to me, it is the access
to the information which is key in responding
to that question.

MR. HEIM: The only thing I would
add is a lot of it is kind of dependent on the
ultimate business application. If we are
working directly with some employers, they
want to understand the differences in certain
geographic. So, we have done some internal
benchmarking where we will look at the metro
and then also compared to different regionals,
to kind of help inform when we are working
with employer groups.

But from a consumer transparency
perspective, you want to try to account for
that. So, it kind of depends on the business
application, the approach to it, but there is
flexibility to define it as appropriate, where
you want that geographic adjuster applied.

CO-CHAIR STEINWALD: I think we
might come back to this when we discuss
usability or maybe even feasibility.

I'm sorry, Jack has something.
MEMBER NEEDLEMAN: Two quick questions, and maybe at least one of them should be deferred to feasibility.

But, first, when you say this is based upon actual payments, not standardized prices, are the actuals what the plan is paying or what is being billed?

MS. KNUDSON: It is what the plan is paying, plus the member liability.

MEMBER NEEDLEMAN: Plus the member co-pay?

MS. KNUDSON: Correct.

MEMBER NEEDLEMAN: Okay. So, that is one question.

The second question: you have checked, you have tested the feasibility of this off of your own plan. All the pharmacy costs, behavioral health costs are completely currently under your control, no carve-outs, I'm assuming? So, have you had any conversations with any other plans about how feasible this is in an environment in which
they carve those costs out and subcontract it
to some other group?

MS. KNUDSON: So, what we do, let
me take a shot at that in a couple of parts.
What we do is we calculate -- and this is in
the spec -- we calculate the medical PMPM and
we calculate the pharmaceutical PMPM
separately, and they are added together. So,
that accounts for that pharmacy carve-out
piece.

For us, we do not have a carve-out
for behavioral health. So, we would include
in our medical and pharmacy. It wouldn't be
separate.

And what I guess we would say for
others who may have behavioral health carve-
outs is that consistency is really the key to
this. So, whatever your analysis is, and if
you are using this for comparative reporting,
for example, they either need to be carved out
of all or in all. And so, that is part of the
data scrubbing and knowing your data going in,
which is really for any of these measures a really critical aspect.

MEMBER NEEDLEMAN: Okay. I just need a clarification.

MS. KNUDSON: Okay.

MEMBER NEEDLEMAN: You say you do have a pharmacy carve-out? Is that what I heard you say?

MS. KNUDSON: On occasion, we have an employer within our plan -- so, for example, like the self-insured examples that were brought up this morning. So, say we may have an employer who carves out and has a different pharmacy administrator other than us --

MEMBER NEEDLEMAN: Okay. So, when you are trying to figure out the cost per member per month, is that the average premium they are paying for the carve-out for every member who is in that group or is it specific or is that being adjusted to reflect that?

MS. KNUDSON: So, in the pharmacy,
the numerator would be the plan and member
liability with the denominator being just
those with the pharmacy benefit. So, that
accounts for the carve-out.

Any clearer way to --

MR. HEIM: Yes, so it is basically
adding two PMPMs together. So, if there is a
pharmacy carve-out, that particular member's
cost, we are only calling the medical, but
when you have both of them, they are both two
different denominators. So, it is adding --

MEMBER NEEDLEMAN: Yes, but I am
trying to understand what is in the numerator
here. So, you have got a carve-out because
one of your employees just loves Medco, and
they are paying Medco $20 per month, $30 per
month, whatever they are paying per member per
month. But some of those members are chronic
artery disease people and have lists of
prescribed drugs like that, and others are not
having anything.

So, are you particularizing it to
the individual member and what is actually being spent on them through the carved-out plan or are you just using the average per-member per-month premium that is being paid to the pharmacy benefits manager?

CO-CHAIR STEINWALD: You know, I think we are going to have to defer this and give them some time to think about the answer to your question --

MEMBER NEEDLEMAN: Okay.

CO-CHAIR STEINWALD: -- when we get to feasibility.

MEMBER NEEDLEMAN: Okay.

MEMBER BARNETT: I think they said they weren't considering that, those people, at all. They were left out of the statistics.

MS. KNUDSON: It is not that they are left out. It is that we are doing the per-member per-month denominated by the people who have the benefit, and we are looking at them both discretely, medical and behavioral together, denominated by those that have that
benefit, and, then, the pharmacy.

And so, then, by adding them together, then that is an accurate reflection of the overall PMPM for those with that benefit. So, it really does account for that component at this aggregate level.

CO-CHAIR STEINWALD: All right.

Let's go on, please. And you can re-raise it. You will have ample opportunity.

Importance, would anyone like to speak to the importance, or lack thereof, of measuring total cost per member per month in the environment that we are talking about here?

MS. TURBYVILLE: May I do a point of --

CO-CHAIR STEINWALD: Order?

MS. TURBYVILLE: -- process or order?

CO-CHAIR STEINWALD: Sure.

MS. TURBYVILLE: As a reminder, for these measures, which are non-condition-
specific, so didn't benefit from a Technical Advisory Panel, you will be rating first on the subcriteria. So, starting with here; I have posted up on the screen 1a, "Is this high impact," et cetera, through the criteria for importance, and the same thing for the other criteria.

CO-CHAIR STEINWALD: An important point of order.

So, we are functioning first as our own Technical Advisory Panel and, then, going on to be the Steering Committee. So, it's harder.

(Laughter.)

Bill?

MEMBER B. RICH: May I ask a question? So, the numerator for the cost is everything for the patients in that group, whether they are psychiatrists, total cost --

MR. HEIM: That's correct, yes.

MEMBER B. RICH: Okay. So, you basically are taking the total cost for the
population of that group, irrespective of attribution or anything else, correct?

MR. HEIM: Yes, all the costs of all the members. So, it is 100 percent of all services.

MEMBER B. RICH: Isn't this only valid, then -- and again, this is going to help us address -

CO-CHAIR STEINWALD: Validity.

CO-CHAIR ROSENTHAL: This is all the science.

CO-CHAIR STEINWALD: Yes.

CO-CHAIR ROSENTHAL: So, let's do the importance.

CO-CHAIR STEINWALD: Yes, let's do.

CO-CHAIR ROSENTHAL: And, then, we can talk about it.

CO-CHAIR STEINWALD: And we do have to vote on the criteria individually.

MS. TURBYVILLE: So, if you pull out -- if you recall the side-by-side table, but we are glad to verbally remind you. So,
1a is the measure focus, addresses a national health goal priority identified by DHHS or the National Priorities Partnership, or is a demonstrated high-impact aspect of healthcare.

CO-CHAIR STEINWALD: Unless we have comments specifically on that criterion, why don't we vote?

And we have one to four, is that right?

MS. TURBYVILLE: Sorry. So, one equals high, two is moderate, three is low, and then you have the opportunity for insufficient, if you feel that the application submitted doesn't provide the information you need to assess this.

(Whereupon, a vote was taken.)

CO-CHAIR STEINWALD: Anyone on the phone?

(No response.)

The second criterion?

In each case, I am going to let you --
MS. TURBYVILLE: Okay, I'm fine to do that.

So, 1b is about the demonstration of a resource use or cost problem, and that there is opportunity for improvement; basically, looking for data that demonstrates variation in the delivery of care and resource use.

(Whereupon, a vote was taken.)

CO-CHAIR STEINWALD: Okay.

MS. TURBYVILLE: We had 14 high and 4 moderate.

Moving on to 1c, which is the purpose or objective of the resource use measure, and the constructs are clearly described. So, the purpose has been clearly communicated in the application.

CO-CHAIR STEINWALD: Prepare to vote. Go ahead.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, we have 11 high and 7 moderate.
Moving on to 1d, which is thinking about the resource use service categories and whether they are consistent with the conceptual construct represented.

So, go ahead and vote.

CO-CHAIR STEINWALD: Prepare to vote.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: Similar to 1c, 11 high and 7 moderate.

So, that is it for the importance subcriteria.

And so, now, right -- but thank you, though, because I could easily forget -- so, that is it for the subcriteria. So, now we will ask you to vote on the overall criteria. So, you have a yes/no, is this an important measurement area of focus for resource use?

(Whereupon, a vote was taken.)

MS. TURBYVILLE: Eighteen, yes, important to measure.
So, now we can move on to scientific acceptability. I will hand it back over to you, Bruce. Or do you want me to read off the subcriteria?

CO-CHAIR STEINWALD: No.

Does everyone have the sheet in front of them, so we can be looking at it?

MS. TURBYVILLE: It's in your folder.

CO-CHAIR STEINWALD: It's in the folder, right.

MS. TURBYVILLE: So, we are at 2a1 now.

It is a side-by-side table. So, it has two columns.

CO-CHAIR STEINWALD: It is essentially the measure is well-defined and precisely-specified.

Would anyone like to make a comment or raise a question for the developers?

Bill and, then, Paul.

MEMBER B. RICH: Again, since you
are looking at total costs, is this only valid with the same population? In other words, if you are going to compare people in Minneapolis to patients in Memphis with different racial groups and things like that, are the total costs really comparable? Or is it only valid within the same well-defined patient population? In other words, how can you compare a group or a physician's total cost with no attribution in Minneapolis, everyone is healthy, to Memphis?

MR. HEIM: Well, what this measure will demonstrate is that there is a cost differential between those geographic areas. So, then, the next question would be, do you want to account for that or not by a geographic adjuster?

So, if you subset that by the two different regions, you will have two different, I guess, costs indices. Say Minneapolis is at 10 and Memphis might be at 1.20. This measure will actually measure that
difference.

And the next question is, how do you want to use that in a business application, whether you want to adjust for that or not?

CO-CHAIR STEINWALD: Paul?

MEMBER BARNETT: Yes, my question had to do with attribution. If I am understanding this right, attribution is, members get attributed if they have a primary care office visit and they get attributed to either a family practitioner or an internist, a peds, geriatric, or OB/GYN.

And so, I wondered, I think there are many patients who get their primary care, say, from a cardiologist or somebody with HIV from an infectious disease specialist. So, they might not be counted or they would be attributed to — say they go to a family practitioner for something unrelated to their chronic condition, and that family practitioner would end up with, say, oh,
$25,000 of HIV care that the patient was receiving in the IV clinic.

So, I am just wondering why you decided not to allow for specialists to be the primary care providers in this.

MS. KNUDSON: Well, when we look at our attribution facts, we find that we have about 75 percent of our population that is attributed to primary care. So, they are visiting a primary care group.

We see, then, in the remaining, about 15 percent using specialists only. But that is largely they had an ED visit or, actually, in our population we have studied, we see a lot that only see PT as well. And we are not attributing to ED physicians our physical therapists. We do not see a lot of people going just to the cardiologist without a primary care guide in what we have tested.

And the remaining 10 percent are non-users of the system. So, they are not attributed.
So, those are the facts from our attribution study. We have done some benchmarking based on our own performance on this measure with a consulting firm and an engagement over the past year or two. And we know we have a lower non-user rate, but we understand from their large national dataset that nationally what they see is about a 17 to 18 percent non-user rate. So, that might be one of the keys earlier as well.

Generally, at least speaking from our plan in terms of benefit design, we try to remove barriers to obtaining preventative care to make sure folks are coming in, and what have you. And what they are finding is, in studying our results, was that those plan designs motivating folks to get and removing barriers are likely leading to our lower rate of non-users.

MR. HEIM: The only thing else I would add is the measure helps in primary care to help out with that coordination factor.
So, the primary care doc coordinating closely with the cardiologist, but, then, also, for those cases, the ACG is helping to adjust for those additional costs where they are coordinating the primary care with the cardiologist.

MS. KNUDSON: And again, that is why we put attribution as a guideline versus a specification, knowing that if some adaptation for attribution in other markets that might have other -- I think, again, as long as when you are doing it in a comparative basis, as long as all of the methods and the inputs are consistent, that is the key in that.

CO-CHAIR ROSENTHAL: A comment. I can understand why an individual health plan would want to use a total PMPM measurement and, in fact, benchmark or compare your own groups within your plan. But I don't think this measure is generalizable in any way, shape, or form, for several of the reasons
that have been brought out.

The way medical care is delivered in one set of communities in Minnesota is not generalizable across the country. And it seems to me the kind of endorsement that we are doing here, if we say that this is a measure, that it has to, in fact, be able to compare the medical group number in your place with the number that would show up in Florida or in Louisiana or in Oregon.

And it is not up to like somebody else to figure out the geographic adjusters or the market factors or the 50 other or 100 other things that go into determining what the PMPM is in a particular community. I don't think this is in any way, shape, or form generalizable in its form.

The previous one was generalizable because you used standardized, because you used and we agreed on the notion about standardized pricing. But it seems to me this is not generalizable.
CO-CHAIR STEINWALD:  Jeff and, then, Mary Kay.

MEMBER J. RICH:  Yes, I was going to bring up the same point.  I think the answer to the question really had to do with market basket economic indicators and geographic adjustments that we do.  We do that in Medicare with the Wage Index for Hospitals and GPSIs for docs.

But what your question was, what about population makeup?  I don't know how you are adjusting, they are adjusting for population makeup, based on Bill's description of it and your comment as well, Tom.

CO-CHAIR STEINWALD:  Mary Kay?

MEMBER O'NEILL:  Well, I think your comment that the way medicine is practiced in different places indicates that there may not be adjustments that can be made on these types of measures between communities.  You know what I'm saying?

I mean, if the practice pattern in
Memphis and Minneapolis are so different, what measure are you going to use to compare how folks are cared for?

And what this is doing is giving, in my opinion, real information, particularly real economic information, that the other measures do not give that will help guide people's choice of where care is sought.

And so, within the context of the Minnesota market, you could see who is more efficient, who has higher quality indicators, who has higher patient satisfaction indicators, and what the cost is going to be to see these folks. And you may not be able to compare the cost in Minneapolis to Memphis, but the Memphis market could to the exact same thing.

And there are only going to be a few categories of care, such as transplant and some higher-level cancer treatment, that people are willing to travel for.

But I'll tell you, if you are
interested in what employers are interested
in, they are looking at both international and
domestic tourism, and they are going to want
the kind of information that shows you that
there is difference in actual real dollar,
out-of-pocket cost in different markets for
certain levels of care.

CO-CHAIR ROSENTHAL: But I'm
cnfused at our task. I don't see why
Minneapolis or Minnesota or Memphis, or any of
the markets, don't have the ability to do
that. But we are asked to approve something
that is a generalizable thing.

And if we want to say what we are
approving is giving permission to a local
market to establish local guidelines, okay,
but that's not what this is purporting to do,
as far as I can see.

MEMBER O'NEILL: But if it is a
generally useful measure, but the use is
local, but we have endorsed the measure
itself, does that make it not a candidate for
this group?

CO-CHAIR ROSENTHAL: Well, I don't know. As I understand our -- well, I don't know. We would have to ask for clarification, but my understanding is that it has to be generalizable.

CO-CHAIR STEINWALD: Barbara?

MEMBER RUDOLPH: Yes, I don't think that necessarily is a criterion in the way that you are putting it out. Because if you think about it, there's other things, too.

Some of the measures are useful only for administrative data. Well, what if I don't have administrative data; I have a different kind of data? I have clinical data. I can't use the measure? Maybe not.

So, I think generalizability in terms of like this is going to fit for every single use across every single geographic zone just doesn't seem like a criteria that we need to use for endorsement.

Many of the measures are very
narrowly-focused and are available only to
people who hold registry data, et cetera. I
mean I don't see how this is different. I
mean, if a national plan can't use it exactly
as it is, does it matter?

CO-CHAIR STEINWALD: David?
MEMBER PENSON: So, I don't know.

I'm not agreeing with you on this, frankly.
I mean, first of all, there is a criteria,
usability, which I think this speaks to. And
the question is, is it usable not just for
Minneapolis versus Memphis, but in Minneapolis
and Minnesota between HealthPartners'
patients, because this is a closed system, if
I understand it, versus not HealthPartners'
patients?

I mean this is helpful to you guys,
to HealthPartners. There's not
HealthPartners' patients in Minnesota. Can it
be exported elsewhere? And I am falling down
with Tom on this. I just don't see it. And
I do think there is a criteria here, which is
usability, which this comes into.

CO-CHAIR STEINWALD: Go ahead.

MS. KNUDSON: Could I clarify? We are actually an open-access market. Our own medical group is not a staff model assigned market. It is open.

MEMBER PENSON: So, if you have a patient who is seeing a HealthPartners' physician, can you, then, make a comparison to a non-HealthPartners' physician using this?

MS. KNUDSON: Yes, and we do.

MEMBER PENSON: Okay.

CO-CHAIR STEINWALD: Dolores and, then, Mary Kay. Well, all right, Jack, go ahead.

MEMBER NEEDLEMAN: At some point, I am going to get back to the carve-outs, but not now.

(Laughter.)

We previously considered a measure with standardized pricing. Is everything else in the way that measure is constructed, except
for the multiplier on the unit of service that is billed, the same in that standardized pricing model, that measure with standardized pricing and this one?

MS. KNUDSON: Right. This measure, the main difference is no standardized pricing.

MEMBER NEEDLEMAN: Okay.

MS. KNUDSON: No other differences.

MEMBER NEEDLEMAN: So, if I had the cost with the standardized price per member or allocated per physician, and I had this one, any difference between those two is a reflection of the difference in the charges that are being reimbursed versus the standardized charges?

MS. KNUDSON: That's right. And because it is a total care measure, though, it is not only that group's price, but it is the aggregated price of, you know, that relative price of what hospitals they admit to, what referral provider partners they have. So, it
is really that aggregate.

So, that is why, in terms of use,
in terms of improvement, we find them useful
because, in trying to drive to better
affordability, the resource use measure really
helps us to understand practice opportunities,
as you are all discussing. And, then, the
price component is just that. It helps on an
index basis to understand price.

Now what I can say in terms of us
being able to work in a collaborative
environment with the providers in our market,
we will drill down. We will talk with them
about the profiles of the referral providers
that they are using, to help them understand
as well their cost, quality, performance as
well, all under the purview of we have
transparency in all of this. So, it is there.

We are not disclosing anything that
we haven't shared with every individual
provider already. And we do that under a
pretty rigorous approach, where we release
results first to every individual provider, give them a notice period, have them vet them. And so, by the time they are final, they are not a surprise.

MEMBER NEEDLEMAN: But if I have your standardized measure for Memphis and Minneapolis, and that seems to be our comparison here, and then I have this measure, I can sort out what the differences are. I can separate the total cost measure, this one, and I can sort out what is accountable for differences in pricing in the two markets versus the resource use of the two markets.

MS. KNUDSON: Right, it would be a relative price difference.

CO-CHAIR STEINWALD: Dolores and, then, Mary Kay.

MEMBER YANAGIHARA: So, David asked the question, would other places be interested? Yes.

(Laughter.)

All the California HMO plans are
very interested. We have been doing parallel work to what HealthPartners has been doing, trying to come up with a standardized total cost of care using actual cost risk-adjusted, I mean very similar.

And so, there is great interest. And if you look at the ACO movement, not only just what is happening in Medicare, but just in the commercial market, it is all about accountability for total cost of care. And so, having a standardized measure is really key.

And having something that people can actually go to, an NQF-endorsed measure, and say, "Great. We can use this one," instead of trying to create their own, and we have spent a couple of years working on trying to develop something.

There are adjusters that can adjust for geographic differences, but it is kind of interesting to know, what is the difference between Memphis and Minnesota of Minneapolis.
and San Francisco, or whatever. So,
understanding those differences, and then you
can adjust for that, if you want to. I mean
there are adjusters, HWI and the GPSI. I mean
those things can be applied.

CO-CHAIR STEINWALD: So, you are
saying, for some purposes, you don't want
standardized pricing?

MEMBER YANAGIHARA: Correct.

CO-CHAIR STEINWALD: You want the
actual --

MEMBER YANAGIHARA: You don't. You
want the cost to the system.

CO-CHAIR STEINWALD: Right, right.

MEMBER YANAGIHARA: And that is
what ACO is all about, is the cost to the
system, and being accountable for that cost.

MEMBER O'NEILL: And I was just
going to say I think there is a difference
between the general applicability of the
measure, you know, and can it be used across
the country, versus are the results going to
be the same in different geographic locations. So, you can use their model and measure anywhere.

But, for example, I spent last week in Alaska talking about their prices, which are ridiculous, but they are normal in Alaska, right? So, it doesn't mean that we couldn't measure them the same way. The results of the measure are going to be different. So, the high and low in Anchorage is going to be very different than the high and low in Seattle.

CO-CHAIR STEINWALD: Doris? I'm trying to keep track of --

MEMBER PETER: Sure. Maybe this might be premature, but one statement was about the fact that practice patterns are going to differ and, therefore, these results will not be comparable geographically. But you also have the issue of the risk adjustment which is based on diagnoses. And so, if you have areas that are higher-intensive, you are going to have more diagnoses; that is actually
going to make them look better. So, there is
that issue as well.


MEMBER HENDRICH: I was thinking
the same thing, that we are going to go into
that knowing that there is going to be great
geographic differences, and that is a given.

My question was around the
methodology of cost. I am not remembering
this in the detail. How would you, though,
control for the variability of what true cost
is between the groups or practices? Actual
or --

MEMBER O'NEILL: What is being
measured is cost to the system and not what
the true internal costs that are --

MEMBER HENDRICH: Which is charge-
to-cost ratios or how?

MEMBER YANAGIHARA: The actual
amount paid --

MEMBER HENDRICH: The actual amount
paid.
MEMBER YANAGIHARA: -- by the health provider or --

MEMBER HENDRICH: Thanks.

MEMBER YANAGIHARA: -- the member.

CO-CHAIR ROSENTHAL: I will try one more time. We are buying this lock, stock, and barrel as is. If we endorse it, this becomes the endorsed method for this. And we all agreed; the importance was virtually unanimous. Nobody is debating the importance of this. The question is, is this the right one? I would have expected NCQA to come in with something like this and have figured out -- and I am troubled by, again, two pieces of the thing.

One is the attribution part. Seventy-five percent of the care delivered in Minneapolis is delivered by primary care physicians. That is not true everywhere in the country. You are buying this attribution model, and that won't be applicable in other sorts of places.
And secondly, any health plan is free to figure out this today, but I can't think of a single quality measure where we go, well, it's applicable in Minneapolis, but it's not applicable in other parts of the country.

And I do think there are ways that -- I would expect somebody to have come in and said, "We're going to have a PMPM cost difference the same way Medicare is trying to figure out cost differentials between one part of the country and the other." And they would have figured out which wage adjuster they were going to use or which market adjuster they would have used.

I mean, why would we make the way Minneapolis is accounting for their PMPM cost to be the standard for the entire country? That just doesn't make sense to me.

CO-CHAIR STEINWALD: Bill?

MEMBER B. RICH: Well, I brought up those two cities specifically for that. Now, if you are looking just at Memphis, you don't
even need a geographic price adjuster as long
as you are comparing the relativity of cost
within a similar patient construct and pricing
structure.

So, I don't know how we define this
and it's applicable nationally. But on a
regional basis -- I don't how to verbalize the
issues that you raise. It's perfectly
legitimate to do this, I think, in Minneapolis
or here in D.C., where probably only about 50
percent of interactions start with primary
care docs. It's okay as long as you are
comparing the groups in D.C. to the other
groups in D.C.

Do you understand? And I think
that is what you are trying to verbalize. I
don't know how we put that in.

But how it is used, as long as the
relativity is the same, I guess you don't even
need a geographic price adjuster. Or am I out
to lunch?

CO-CHAIR STEINWALD: You have an
answer to this question?

MEMBER BARNETT: Well, I think we have a good idea about where people stand on this geographic variation. I don't think we need to pursue anymore.

I wanted to raise a different issue, which was the exclusion of the members who don't incur any costs. So, I am a little bit worried about this for two reasons.

One is, at the outset, you mentioned that the importance of this would be that it would encourage preventative services, having this measure available. And if the preventative services result in the member not getting any services, then they are going to be left out of the matrix. So, you actually don't get any credit for that.

And, then, the other thing that worries me about this, and this comes from our own experience in VA, is that there then becomes an incentive to make sure everybody gets in for at least one visit a year. And
so, VA had a capitation plan where it resulted in some of the clever regional networks creating health fairs for veterans where they would enroll veterans for an eye check or a blood pressure check, and they would get credit for those people. So, they were able to game the system that way.

And so, quickly, our capitation system changed, so that we had a stronger threshold. But this is just one visit. So, it would be easy for someone to really get a much better per-member per-month score if they could just get every member in for a blood pressure check once a year, and they would be able to game this.

I guess that has more to do with the feasibility than scientific acceptability.

CO-CHAIR STEINWALD: I have a question about the attribution. As I understand the measure, everybody -- well, to verify this, the non-users are not in the denominator? That's right? That's correct,
right?

MR. HEIM: If you need to attribute, that would be true. You don't need to attribute all the time. So, if you are in a member-assigned environment, you don't need to do any attribution.

CO-CHAIR STEINWALD: But when you calculate a per-member per-month figure, you are not including the non-users in the denominator when you calculate that?

MS. KNUDSON: You know, again, this is just to clarify. We had submitted attribution under the guise of the guideline and explained how we did attribution.

So, say your unit of analysis was a health plan, which in this it's an index measure, so that would be the 1.0. You would use all of the members --

CO-CHAIR STEINWALD: You would?

MS. KNUDSON: -- if you were comparing different plans.

CO-CHAIR STEINWALD: A second
question, how does the attribution, whether
the patient is attributed to an internist or
an OB/GYN, how does that affect the
calculation of the index? It doesn't seem to
me that it should, but am I missing something?

MR. HEIM: It doesn't adjust for
that.

MS. KNUDSON: It doesn't affect it.

MR. HEIM: I mean it doesn't affect
it at all.

CO-CHAIR STEINWALD: Where are we
now?

CO-CHAIR ROSENTHAL: Could I ask
one more question?

CO-CHAIR STEINWALD: Yes, sure.

CO-CHAIR ROSENTHAL: You are
attributing it to groups, and in your
environment what is the definition of a group?

MS. KNUDSON: Well, we are largely
in a group-practice-organized market. But
Minnesota aside, I think the point about
creation and evolution of ACOs, this would
have application nationally.

And, then, also, just reinforcing from our perspective, for the majority of services, given the need for this measure, consumers do largely get healthcare services locally.

CO-CHAIR ROSENTHAL: Yes, but that wasn't my question.

MS. KNUDSON: I'm sorry.

CO-CHAIR ROSENTHAL: I'm sorry, maybe I wasn't clear.

You have specified that this measure can be applied by a health plan to groups of doctors.

MS. KNUDSON: Yes.

CO-CHAIR ROSENTHAL: So, it is a group of doctors to which you attribute it. It gets to this question of, what about gynecologic services or what about OB services or what about cardiology services? Are they in your groups? Are those doctors in your groups? Or are your groups primary care
doctors?

MS. KNUDSON: We have done our attribution around primary care as the specialty. Even within a multi-specialty group practice like our own, we are attributing to the primary care physicians, based on that definition the gentleman had said earlier, internal medicine and family practice, OB/GYN.

CO-CHAIR ROSENTHAL: All right. So, this would really only be applicable, even, then, in the Memphis/Minnesota scenario that we keep constructing, for health plans where, in fact, the care is delivered by groups of doctors, and particularly of primary care doctors, because that it is specified as? Or am I missing it?

MEMBER BARNETT: So, it says in here that you have the option of assigning it to a health plan, an employer group, or to a provider. Those are the options that are offered in the --
CO-CHAIR ROSENTHAL: But we have talked about it being valuable --

MEMBER BARNETT: Not a group practice, it doesn't say group practice in here.

CO-CHAIR ROSENTHAL: Well, what is the group of doctors to which we are referring then?

MEMBER BARNETT: It says the employer group.

CO-CHAIR ROSENTHAL: I thought I heard them talking about provider groups.

MEMBER BARNETT: I don't see that here.

CO-CHAIR ROSENTHAL: Well, they just said it did. So, I am trying to clarify that because that is the part that concerns me. That concerns me.

MEMBER BARNETT: It is a little fuzzy about what --

MS. TURBYVILLE: Can I do a point of clarification? So, in response of the
level of analysis, which is S11.3 on page 15, selected was group practice clinician and community population. So, I don't know if you were going to stay with that level of analysis, but that's what --

MR. HEIM: That's correct.

CO-CHAIR ROSENTHAL: Yes, that is, 11.3 is what I was looking at. And therefore, I am, then, trying to find out, since this is based on an "N" of 1, meaning their experience in this health plan, what do they mean by group practice? And again, then that would assume, I would assume, then, that those same conditions have to be relevant or prevalent in any other health plan or community that would use this measure.

CO-CHAIR STEINWALD: Is that a question for HealthPartners?

CO-CHAIR ROSENTHAL: Yes. Well, I am trying to still find out what they meant by group practice.

MR. HEIM: So, group practice,
then, would be at least two docs, internal med
or whatever practicing specialty specified,
geriatrics, OB.

MEMBER B. RICH: Could we clear
this up a little bit if you said this could be
attributable to, you know, whatever, groups,
docs, plans in the same region, just leave the
verbiage at that?

CO-CHAIR STEINWALD: I think some
of the discussion around the table is that the
value of the measure is comparing across
regions.

MEMBER B. RICH: I don't think so.

CO-CHAIR STEINWALD: I didn't hear
that?

MEMBER NEEDLEMAN: The issue of the
provider level is very relevant. So, if a
young woman has not an OB/GYN as their primary
care doc, but an internal medicine or a family
doc, and then gets pregnant and has OB/GYNs,
you know, obstetrical services, if it is at
the group level, those two will be combined
for purposes of attribution. And if it is at
the individual practitioner level,
provider/clinician level, those are two
separate clinicians.

So, the issue of what level you are
aggregating to for purposes of attribution and
for computing is definitely relevant. Where
will it be allocated? Is it to the family
medicine physician or is it to the group that
has both the family medicine physician and the
obstetrician in the same group?

MR. HEIM: So, just to play out
your scenario, if the OB doc and the family
practice are within the same provider group,
it is assigned just to the one provider group.
If they are separate, going along with primary
care in Clinic A, and I start with OB services
at a different provider group, then we are
going to who has the most office visits to
determine which provider group would we go to
then. And, then, if there is a tie, it would
be the most recent experience, then, would
basically get the member.

CO-CHAIR ROSENTHAL: Could I just clarify? So, then, again, I accept that in your region that is meaningful and accepted because the preponderance of care, as you guys described it, is delivered by primary-care-oriented people. But that is not necessarily true in every community in the country. In fact, it is largely not true.

And certainly, there are certainly not multi-specialty groups. And so, the obstetric/internist scenario is likely to segregate in most communities; whereas, I accept perfectly that it works in yours. It may work in others.

MR. HEIM: So, just to clarify then, let's go in a different market where you don't have provider groups then. So, now we are at the different physician levels. The same thing is kind of occurring there if your plan is who has the most office visits, and if it is a tie, it goes to the most recent. So,
it would play out.

CO-CHAIR STEINWALD: Bill?

MEMBER GOLDEN: Yes, just to get a
sense of this, Minnesota is sort of like
Wisconsin; the doctors are in large groups.

I am from a part of the country, and many
others, where everyone is in two- and three-
person practices.

How would this operate if you had
a large population of just two-doctor
practices? Would it be a very different
operating characteristic?

MR. HEIM: Yes, in a group
practice, and what we have kind of recommended
as a guideline, is an "N" of 600 patients to
start making those comparisons.

CO-CHAIR STEINWALD: So, I recall
when we discussed the other measure this same
issue was on the table. It seemed that even
the previous measure was most applicable to
large, multi-specialty group practices.

I don't know that we approved the
measure with that proviso. Paul, do you remember?

MEMBER BARNETT: I thought it was as they defined it here, and I think the definitions are exactly the same, that you can use it for the employer group, the health plan, or the provider. And the rules for attribution were exactly the same as they are in this measure.

CO-CHAIR STEINWALD: Dolores, were you -- no?

All right. Well, I hate this feeling of being kind of at an impasse. So, let's see if we can rectify that. We still have to evaluate the criteria. I think much of our discussion over the last 20 minutes has covered of at least the subcriteria. And so, I am wondering if we can go on to vote on individual subcriteria until we get to the point where we really have to have more discussion.

Where are we?
MS. TURBYVILLE: So, we could start on 2a1, which is about whether or not the measure is precisely defined and specified so that it could be implemented consistently within and across organizations.

And if you recall, there are eight subcriteria on reliability and validity, and then there are some others. So, I think there will be opportunity for your concerns and your positives for the measure to come through in the ratings of the sub-subcriteria.

So, let's go ahead and start 2a1.

CO-CHAIR STEINWALD: 2a1, and is it one --

MS. TURBYVILLE: So, 2a1 is under reliability, but it is focusing on the specifications being defined precisely enough that it could be implemented consistently. And you have high, moderate, low, or insufficient information has been submitted to allow you to assess that. So, one being high, et cetera.
CO-CHAIR STEINWALD: Go ahead.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, we had 5 high, 8 moderate, 4 low, and 1 insufficient. I think that reflects -- at least from what we heard on staff, I don't think we need any more input.

So, moving on to 2a2, which is reliability testing, the question is about whether or not the testing submitted demonstrates that the results are repeatable and producing the same results a high proportion of the time when assessed in the same population, in the same time period, or that the measure score is precise.

And again, this is a high, moderate, low, insufficient rating.

CO-CHAIR STEINWALD: Do we have an analysis from Carlos on this separate from what we had before?

MS. TURBYVILLE: Yes, he did a 1604 review. Do you want to pause and have him
1 speak to it?

   CO-CHAIR STEINWALD: Yes, please do.

   MR. ALZOLA: Hi.

   The reliability analysis that I did was a little different from all the other measures. It was more based on simulations in which they restricted to each different provider. They simulated the variability within that provider and compared that to the observed variability, as a way to measure signal-to-noise ratio.

   And they also compared how the ratios changed from one year to the next, again, by provider. And the differences that they found were really insignificant. So, in terms of signal-to-noise ratio, there was a really reliable, I can say it is really reliable.

   CO-CHAIR STEINWALD: Paul?

   MEMBER BARNETT: Just for clarification that, what they saw, the
reliability was not at the level of the provider, but at the level of plan, right?

MR. HEIM: I'm sorry. It was at the provider level. We were comparing the actuals to those simulated populations and, then, recording the differences.

MEMBER BARNETT: So, from year to year?

MR. HEIM: Yes. We did three years. We stayed within the year doing those simulations to see on that year what the actual index was compared to the simulated population. As Carlos highlighted, that was pretty small differences. And we did that similar methodology for three years to see the consistency over the time, if there were any changes.

MEMBER BARNETT: So, just to understand, is this where you took the 90 percent sample and you did that with the three years of data, instead of just one year of data?
MR. HEIM: We did that reliability
test three times, one for each year.

MEMBER BARNETT: But did you
compare the result you got in year one with
the result you got in year three?

MR. HEIM: So, we did a 90 percent
sample, a bootstrapping approach. That's the
with all replacement. And, then, we did a
similar bootstrapping with replacement. And,
then, we did a third one where we did look
over time, specifically looking at a
provider's TCI and, then, see how that changes
from one year to the next. And, then, if
there was an appreciable difference, we
commented on what those differences were,
reflecting that the measure was working.

So, in short, in answer to your
question, yes.

(Laughter.)

MEMBER BARNETT: Yes. So, I didn't
find that last one, which is the one that is
interesting to me.
CO-CHAIR STEINWALD: Any further discussion on this one?

(No response.)

And hearing none, could we put it up for a vote?

MS. TURBYVILLE: 2a2.

CO-CHAIR STEINWALD: 2a2.

(Whereupon, a vote was taken.)

CO-CHAIR STEINWALD: On to 2b.

MS. TURBYVILLE: So, 2b1 is the measure specifications are consistent with the evidence presented. And it ties back to what was submitted under importance, so is the measure measuring what it is intended to, and the way the measure is being proposed to be implemented as well.

So, it is the kind of high-level validity. As a reminder, we do hold face validity as the minimum threshold. They did provide their own findings for that.

Oh, I'm sorry. Thank you. Thank you.
Before we move on to 2b1, we do require to assess the overall reliability of the measure. So, if you could quickly vote? And this is also on a rating from high to low, including insufficient.

CO-CHAIR STEINWALD: You mean we are voting on 2a1 and 2a2 together?

MS. TURBYVILLE: Right. 2a1 and 2a2 together, so that you may weight how you found one of those differently. So, we request that you rate the overall reliability of the measure.

CO-CHAIR STEINWALD: Okay. Ready?

Go.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: Okay. So, we had 8 high, 6 moderate, and 4 low on reliability. Now we can move on, right, to 2b1. Again, that is whether or not the specifications are consistent with the evidence presented to support the measurement focus area.
CO-CHAIR STEINWALD: Carlos, would you provide us with your summary statements about validity, please?

MR. ALZOLA: Sure. Again, validity, in this case they not only tried to prove face validity, but they also looked at the correlations between the TCIs and the observed actual costs and the risk-adjustment groups.

And the correlations were, for the most part, were high. And what I found interesting, and I thought it really indicated that risk adjustment was doing its job, is that, once you included the risk adjustment, the correlation between the actual cost and the -- let's see. Right, one includes the risk adjustment; the correlations between the total costs and the TCI really goes down, meaning that the risk adjustment is doing its job. It didn't go down as much as I would like it, but it went down by a really significant amount.
CO-CHAIR ROSENTHAL: I'm a little confused. If people have concerns about the attribution part of the thing, where would that get scored? Because it isn't clear to me exactly in which of the validation ones we had contemplated those kinds of questions.

MS. TURBYVILLE: Great question.

Thank you.

I would recommend putting it in 2b1. So, constructed as it is presented for its reliability, is it going to be measuring what it is intended to measure at that conceptual level? So, this is the measure that says, is the conceptual measurement that they submitted meeting, how it is actually being proposed, specified, and, thus, would be implemented, and that would include the important specifications for attribution.

CO-CHAIR ROSENTHAL: All we needed was a rule.

CO-CHAIR STEINWALD: Or guidance.

CO-CHAIR ROSENTHAL: Yes.
CO-CHAIR STEINWALD: So, the attribution issues are included in the criterion we are discussing right now.

Any further discussion? Yes?

MEMBER J. RICH: So, I am a little confused where to ask this question. But I am looking at their application at S9.6. It includes inpatient services and ambulatory services.

And when I got to the attribution model on page 15, exclusion criteria is everything that doesn't occur in the office. So, in the attribution model you are saying that it is only office-based, but in the included services you are saying that it is everything on the inpatient side as well.

CO-CHAIR STEINWALD: HealthPartners, can you clarify, please?

MR. HEIM: For assignment, we are looking at office visits only to actually get the member assigned. And, then, when we are doing the calculations, we are inclusive of
all the costs. So, therefore, there's no
exclusions there, if I am tracking with the
question.

CO-CHAIR STEINWALD: So, it is
comprehensive of cost measurement. But in
order to put the patient in a category, you
are using office visits to do that?

MR. HEIM: That's correct.

CO-CHAIR STEINWALD: Okay.

Discussion? Yes, sir?

MEMBER BARNETT: Yes, and this is
like the other measure; that office visit
could have happened after the hospital stay?

MR. HEIM: Correct. Anytime during
a 12-month period, we look at all the office
visits and determine which provider saw them
the most or most recent.

MEMBER BARNETT: So, in other
words, a provider could be responsible for a
hospitalization, the cost of a hospitalization
before they had ever seen, when they had never
seen the patient before?
So, I would just observe there is a disincentive to take on patients who have recently had an expensive hospitalization without any primary care provider.

CO-CHAIR STEINWALD: Right. But, in order for that to happen, the primary care provider would have to provide enough services to the patient after the hospital stay to overcome --

MEMBER BARNETT: Just one visit is all it would take.

CO-CHAIR STEINWALD: Just one?

MEMBER BARNETT: Yes.

CO-CHAIR STEINWALD: But that is only if there were no other physician before the hospital stay then?

MEMBER YANAGIHARA: This is not an individual physician-level measurement. This is a group-level measurement. You have to keep that in mind.

MEMBER BARNETT: No, I think it has been stated that that is really not true. It
will get down to attributed to as low as two
general internists in a practice, who will
then be responsible for obstetric care and
hospitalizations, and a dozen other things.

MEMBER YANAGIHARA: There are two
physician groups in California that contract
as a group and take risk for the care of
populations on these.

MEMBER BARNETT: Yes, but I think
it is a good point, that a lot of the problems
that are raised here would go away if there
were actually just attributing to the plan or
multi-specialty group, or something like that,
rather than down to the individual provider.

MEMBER REDFEARN: If a specialist
admits the patient to the expensive hospital
stay and then a PCP sees the patient following
discharge, then it is going to be assigned to
the PCP.

My concern about this is in terms
of the attribution, which they have indicated
could be varied. It doesn't work this way in
California. There is an awful lot of the episodes that we look at that are managed almost exclusively, and sometimes exclusively, by specialists. So, I don't know what happens to that care. How do you force one of those episodes-of-care into a PCP, if basically a PCP has not been involved? And that concerns me. Again, this is a geographical issue because it just works differently. What happens to that utilization and how do you assign it?

CO-CHAIR STEINWALD: See if this accurate. If a patient is seen by a cardiologist for an entire year, is admitted to the hospital, discharged, followed up by that cardiologist, and never sees a primary care physician, that patient's utilization never gets included, is that correct? Because there is no primary care doctor to attribute to?

MR. HEIM: That's correct. As a primary care total cost-of-care measure, yes.
CO-CHAIR STEINWALD: Any more questions or comments?

All right. Barbara does? Yes, ma'am?

MEMBER RUDOLPH: Well, just a comment. There are a number of places that have larger practice groups than onesies/twosies that would love to have this measure. I think to think that any measure is going to be 100 percent useful across all places is not a good approach to endorsement. There are going to be places where this works really well and other places where it doesn't work as well. And that is the case with many of the measures that are endorsed now.

MEMBER NEEDLEMAN: Yes, but, okay, this issue of the specialist, so the cardiologist is one example. A person with HIV whose primary care doc is an infectious disease specialist and is not part of the GIM group in whatever group they are is another example. We have got some very expensive
patients who are getting their primary care,
getting all their care managed by specialists.
And by saying those folks don't get counted
here, we are excluding some very expensive
patients from the measure of resource use.
And I don't know what percentage of patients
those are, but they are among our most
expensive and the ones that most need managing
of their resources.

And I am a little concerned when I
hear that they are not showing up in the data
in this measure of resource use in this plan.

CO-CHAIR ROSENTHAL: Again, I think
we are confusing the importance of this
measure with the validity of it, and perhaps
expressing our frustration that there is not
another PMPM measure that, in fact, accounts
for these things in a way that we could be
more confident and comfortable about.

I wish there was another PMPM one.
I live in California. We use PMPMs. We've
got 80,000 capitated lives. We get the value
of this.

The question is, is this the one that we need to use, given a number of problems that are not the fault of the Minnesota group. This I'm sure works beautifully and perfectly well in their environment. And I don't think they need our endorsement to continue to use it.

It is a question of, is this the one that really is going to -- and 2a1 really said, so that it can be implemented consistently within/across organizations. And I think we are sort of fudging on that by saying, well, no, no, no, it really doesn't have to be; as long as somebody can use it, that that is good enough. I just think it is a problem.

CO-CHAIR STEINWALD: I think the HealthPartners people have alluded to this. It is sort of up to the user to determine whether the measure is of utility within their own environment. We might not like that, but
that is essentially how it works.

You have customers, basically, who are using it, and, presumably, those for whom it is not useful are not your customers. But I don't know if you have any -- and that would imply to me that very small practices probably wouldn't find it that useful, but maybe I missing something there, if you would like to comment?

MS. KNUDSON: That could be, and I think it is this discussion sort of bears out exactly why we set up attribution as a guideline, knowing that other areas of the country are not organized in a similar way, but knowing there might be very likely some application to have a standardized approach to this with the evolution of ACOs. That will be, to take the example of, if someone wants to create an ACO, which is kind of think of that in terms of a large group practice for an accountable care group of practices or individuals that might work together as a
group, and then that attribution could be set up accordingly, based on how that system is set up.

So, that is one, you know, just playing out a potential scenario that we were anticipating. But in following the guides of the application, we have tried to be rigorous with how we have used and tested it thus far.

MEMBER PENSON: So, I wonder, I hear that, and I mean we are endorsing it as is. You may be flexible with attribution, and other places they may do it differently, but it changes the measure inherently.

So, I think at this point, I mean, I had said we call the question because everyone at the table has an opinion now and we should just see where we all sit and go from there. Because we can't go by, well, in California, if you tweak it a little differently with the attribution -- this is what has been submitted; this is how the endorsement process works. We've got what
we've got. Let's just vote.

CO-CHAIR STEINWALD: Everybody okay

with that? All right, let's go.

MEMBER YANAGIHARA: I'm sorry, I

had a question.

CO-CHAIR STEINWALD: I'm sorry.

Yes, Dolores?

MEMBER YANAGIHARA: So, when things

are submitted -- I know we had a lot of
discussion about this early on -- a guideline
versus part of the specification, so if the
attribution is being presented as a guideline,
how do we judge that?

MS. TURBYVILLE: I can give you

what we had interpreted from the Steering
Committee. And, then, clearly, your

colleagues may comment.

The attribution section itself, I
don't think we allowed for guidelines. There
were other parts of the reporting area, for
example, identify and define peer group, and
you will see when you see a guideline
beforehand, that is actually something they
toggled on, which was based on what the
Steering Committee said that it will be the
specifications may need to adjust here and
there, but there has to be something well-
thought-out that is provided for users to
react to.

So, that is what we took away from
with the application. So, you can clearly see
in the application where that may be an
option, and they did select that option at
various points, as you can see from their
submission.

So, how you interpret it and weigh-in on your ratings, I think that leave that to
all of you.

MEMBER YANAGIHARA: But attribution
was not one of the ones that could be a
guideline? I thought I heard them say that it
was a guideline.

MS. TURBYVILLE: I believe S11.1
was not, and we can verify that. Right, it
wasn't; I'm getting the confirmation.  S11.2
was.  S11.3 was not.  So, you have a level of
analysis.  It has to be a specification.  11.4
could be a guideline, I think.

And so, that is how it worked, and
it was based on the input of this Committee.
Then, we took it to the CSAC to vet it out as
well.

CO-CHAIR STEINWALD: And so, the
attribution methodology is part of the
measure.  Okay.

Are you ready?  Let's go.

MS. TURBYVILLE:  2b1, we are on
2b1.

CO-CHAIR STEINWALD: Right.  That
was the guidance from NQF.  It has to go
somewhere.

(Whereupon, a vote was taken.)

(One high, 6 moderate, and 11 low.)

CO-CHAIR STEINWALD: All right.

Yes, Bill?

MEMBER B. RICH: It is apparent
that one of the problems is with the outline that we have. We are trying to fit in a measure that is applicable for ACOs on a regional level. And it just not fitting into our criteria. I think that is --

CO-CHAIR STEINWALD: I'm sorry, Bill?

Oh, you're just talking to him? Talk to all of us. Come on.

(Laughter.)

I have that same feeling of a measure that has great potential value, but we are trying to put --

MEMBER B. RICH: Trying to put it in a box.

CO-CHAIR STEINWALD: Yes. Go ahead.

MEMBER BARNETT: But I think that, if it were to be resubmitted, that at least the proponent has some idea of what the concerns are, and those could be addressed.

CO-CHAIR STEINWALD: Yes.
MEMBER NEEDLEMAN: Apropos of our conversation, it was, is there an incentive here to not take somebody who is really sick into your panel? And I was saying I thought the ACG risk adjuster should effect that.

Our problem is we have got sick patients that are being given their primary care not by primary care docs, and the attribution model here doesn't seem to accommodate that terribly easily.

MEMBER YANAGIHARA: Yes, what I am wondering, I mean, one comment that they made was this is really a primary care total cost-of-care index. I mean I wonder if this measure, if it moves forward, the title should clearly state that. And maybe there needs to be a companion measure that has a broader attribution that would include that specialty care.

So, anyway, just a comment.

CO-CHAIR STEINWALD: Anything further until we move on to 2b2?
(No response.)

Do you have something for us?

MS. TURBYVILLE: We may have to do a revote on 2b1. So, we need to circle back on whether or not -- my understanding was the attribution was not meant to be a guideline. However, it looks like on the submission form we were vague about that language. Whether or not it would change how you just voted on 2b1 is not for me to decide, or any of us.

So, we want to make sure that we are capturing your sentiments about the measure. So, I apologize for the confusion, but we want to make sure that we are being fair and consistent.

CO-CHAIR ROSENTHAL: I think we should certainly be fair and consistent with this submission because, in fact, if we were vague, they shouldn't be penalized.

But I would say, given the importance of the conversation that we just had, I don't see how we could actually in
reality have a situation where the attribution model can be vague and a guideline, because it is important. It is critically important, as several of the discussions today have articulated. So, I think we have to clarify it going forward.

And my position would be that it can't be a guideline. It has to be specified.

CO-CHAIR STEINWALD: Bill, go ahead.

MEMBER B. RICH: Bruce, a question for the developers. Is the title of this really appropriate?

To go back to Dolores' point, if you read the definition, it just says, "Total cost-of-care population PPPM index." It doesn't say anything about primary care or anything.

Is the intent that this be a primary care population-based PPPM? The descriptor is quite different than what -- that it may address some of the issues.
You know, it is always helpful to get others' feedback on that. I think we would be open to changing the title of it to be more descriptive of exactly what it is.

I think, also, perhaps on the confusion on the attribution, that was obviously our misinterpretation. The guideline buttons start on the next. And so, if you want to continue the review with that being a part of the specification, you know, and the retitling, we're fine with that.

Sally, your advice? Given what you just told us, do you think that we are obliged to revote?

I think it would be easier to interpret the votes if we do revote, understanding that we did allow for attribution rules to be submitted as guidelines. That said, I think your sentiments about that, and kind of going back to one of the first slides actually presented,
but to make sure we are being fair to the measure developers, that we are learning, also, from the process. So, the conversation has still been very informative.

But, yes, we did allow them to submit attribution rules as specifications or guidelines, but it is still up to you to weigh-in on how that plays itself out.

CO-CHAIR STEINWALD: Right. I am going to call the vote again.

Lisa, do you want to have a comment first?

MEMBER GRABERT: Yes. I was just wondering, since the developer did test at both the level of the plan and providers, I don't like changing measures on the fly for what they are intended to do. But since you tested at both levels, and it seems to be a bit of a sticking point where the level of attribution is, are you amenable to limiting the attribution to just the plan level and then revoting?
CO-CHAIR STEINWALD: I'm not sure that our process permits that.

MS. TURBYVILLE: It does --

CO-CHAIR STEINWALD: It does?

MS. TURBYVILLE: -- but we would want the recommendation to come from the Steering Committee. And it is up, then, to the developer to decide if they want to meet any requests like that, even changing the requirement that there is a PCP visit.

You can say, "Would you consider...?" We try to avoid changing measures on the fly, but it is always up to the developer whether or not that is something they can do. You did that for the ABMS measures earlier, you know.

Well, this is not a trivial change. And so, well, how quickly could they test? You need to vote on what the measure is right now, right? But, then, whether or not the measure developer comes back, given your feedback in this project, in time with testing
data or in a future project is something we
would certainly continue to encourage.

**CO-CHAIR STEINWALD:** Dave?

**MEMBER PENSON:** Again, I mean I
think we have to vote on it as it is now. I
mean because we went through this yesterday,
and we are functioning as a TAP right this
minute, effectively -- some of the other TAPs
aren't going to be able to do this.

So, I think we have to vote on it
as it is written and say to the measure
developer, you know, if you did this, this,
this, the Committee might be more amenable.
I'm not sure that is true or not, Lisa, but,
I mean, I'm not comfortable --

**CO-CHAIR STEINWALD:** Bill?

**MEMBER GOLDEN:** A question for
Helen, kick it upstairs. Other NQF reports,
when they go through all these measures, say:
we endorse the following measures. We didn't
derose these measures. And these measures
are promising and need more work.

You know, we are going to be seeing a lot of measures here like this where there is some interesting conceptual things, but they need some work or the idea needs some further work. So, I am just curious, we haven't talked about things in that perspective. I mean here's a measure here that has some potential, but it needs some shaping. It needs some caveats.

Where are we? How should we proceed with that? Or where does that fit into this framework?

DR. BURSTIN: Yes, I mean, you are certainly welcome to put in the report whatever you think the Committee wants to put forward. In the discussion of this measure, these things were very promising. The Committee continued to have concerns about A, B, C, and D. Those are fair game.

I was also mentioning to Bruce and Tom earlier that there is always a final
section as well where the Committee kind of thinks prospectively, based on what we have seen. "We wish we had seen the following." So, those sections are still important.

And again, just going back to the point Sally was making, you know, it is always fair game to recommend minor changes to the developers, but if it is a significant, wholesale change, it is probably not appropriate.

But, again, I think you do need to vote on the measure as it is before you today. If they want to go back, ponder what I just missed while on a conference call, and bring it back to you, that is certainly their prerogative. But you still need to vote on it.

CO-CHAIR STEINWALD: We voted on 2b1 with the understanding that the attribution was part of the measure. We learned later that it is a guideline, not part of the measure. To me, that means we need to
revote, even if it comes out the same way.

So, can we do that, please? So, we are back to 2b1.

(Whereupon, a vote was taken.)

CO-CHAIR STEINWALD: All right.

MS. TURBYVILLE: So, we had 4 high, 5 moderate, and 9 low.

CO-CHAIR STEINWALD: Can we move on to 2b2? This is the more traditional validity testing topic.

And we already heard from Carlos on this, I think.

Did we vote on 2b2?

MS. TURBYVILLE: We started the conversation, and I interrupted you. Sorry.

MEMBER O'NEILL: For planning purposes, how late are we going?

CO-CHAIR ROSENTHAL: About another 10 minutes to finish up the votes on this section, don't you think?

CO-CHAIR STEINWALD: Yes. I think that's right.
CO-CHAIR ROSENTHAL: I would suggest that we try to get through the scientific thing. We've got three more votes to do on this or four more.

MS. TURBYVILLE: Six more.

CO-CHAIR ROSENTHAL: Oh. Well, contentious, if the rest of them are, I would say if we limit the discussion at this point, I think we have discussed everything.

MS. TURBYVILLE: It's up to you guys. So, there are six more subcriteria for validity and scientific acceptability. If you want to plow through them now, we are willing to stay here and support that. So, I think it is up to you and the Committee members.

CO-CHAIR ROSENTHAL: We will have to start over on this tomorrow morning if we don't get through it.

CO-CHAIR STEINWALD: Yes, let's try to do that.

Okay. So, we are up to 2b2 now.

Can we have it up on the screen? Great.
CO-CHAIR ROSENTHAL: This is more standard validity testing.

CO-CHAIR STEINWALD: Right.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: Okay. So, for the testing component, 7 high, 5 moderate, 5 low, and 1 insufficient.

So, moving on to 2b3, which would be about exclusions are supported by the clinical evidence. Otherwise, they are supported by evidence of sufficient frequency, so some empirical information, and that the measure specifications for scoring include computing exclusions so that the effect on the measure is transparent.

CO-CHAIR ROSENTHAL: So, as a point of clarification, would this include the exclusions that Dr. Needleman was alluding to earlier? Or is that a different kind of exclusion?

MS. TURBYVILLE: It's all exclusions that are of interest.
CO-CHAIR ROSENTHAL: Okay.

MS. TURBYVILLE: So, once you have your inclusion criteria -- yes.

CO-CHAIR ROSENTHAL: Okay. So, his would be relevant in the scoring of this section.

CO-CHAIR STEINWALD: You mean the carve-outs, in particular?

CO-CHAIR ROSENTHAL: Yes, the fact that, in particular, all of the cases that don't have a PCP are excluded.

CO-CHAIR STEINWALD: Oh, okay.

CO-CHAIR ROSENTHAL: This is where that would be scored? Okay.

CO-CHAIR STEINWALD: Presumably, yes.

Is it up?

(Whereupon, a vote was taken.)

CO-CHAIR ROSENTHAL: Oh, 3, 6, and 9.

CO-CHAIR STEINWALD: Three high, 6 moderate, 9 low.
MS. TURBYVILLE: So, 2b4 is the risk adjustment that they have proposed as specified, and, then, if there were any stratification methods. So, it is for the outcome measure. In this case, it is a resource use measure when indicated. There is an evidence-based risk-adjustment strategy, and we don't want factors related to disparities that would be of interest to expose.

CO-CHAIR STEINWALD: Okay, put it up.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, for this subcriteria, we have 7 high, 7 moderate, 2 low, and 2 insufficient.

So, the next subcriteria is that the data analyses that are provided demonstrate that the methods for scoring and analysis allow for the identification of statistically-significant or/and practically-
and clinically-meaningful differences in performance.

CO-CHAIR STEINWALD: Go ahead. Go ahead and put it up. And, then, hold it up.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, for this subcriteria, we have 7 high, 5 moderate, 2 low, and 4 insufficient.

2b6, I believe you are only specifying for commercial administrative claims data. So, as we have been working with TAPs, as well as the Steering Committee is a TAP because it is specified and, hence, would be endorsed only for commercial administrative claims data, it has been not applicable. It would be applicable if they were including clinically-enriched data and other data sources, but that is not included in the specifications.

CO-CHAIR STEINWALD: So, we don't need to vote.

MS. TURBYVILLE: Right. So, unless
there is something someone here wants to call
to the attention that we might have missed?

(No response.)

Okay. So, that would be not
applicable.

And, then 2c is --

CO-CHAIR STEINWALD: Wait. Don't
we have to do 2b?

MS. TURBYVILLE: No. Oh, sorry,
2b, validity overall. Holding 2b6 not
applicable, how do you rate the validity of
this measure as specified.

CO-CHAIR STEINWALD: Okay, put it
up. Hold it up.

(Whereupon, a vote was taken.)

MS. TURBYVILLE: Oh, sorry. So, we
have 4 high, 6 moderate, 7 low, and 1
insufficient.

So, we made it through reliability
and validity.

Yes, we are going to move on to 2c,
but before we move on, I just want to, for
validity, I believe staff captured the comments and everything. If anyone voted low on validity and has a rationale that wasn't discussed, if you could provide that now, that would be helpful, so we have that feedback. But if it has already been discussed, we can move right on 2c. But since there were quite a few low, I want to make sure we are capturing all the rationales.

MEMBER GOLDEN: The only thing that I want to add is that the notion that this would be able to give you statistically-significant differences in primary care performance, given the attribution of specialty costs to the primary care docs, gives me great pause.

MS. TURBYVILLE: Okay. Thank you. That's helpful.

MEMBER B. RICH: And I think the fact that it does not exactly -- the intent is for primary care purposes, but the measure description doesn't state that.
MS. TURBYVILLE: Okay.

CO-CHAIR STEINWALD: Paul?

MEMBER BARNETT: In the validity testing, it appears to me, going back to the website and pulling up the document that they gave us before, that they did the validity testing across three years and doing the bootstrapping for 19 provider groups, and not for individual primary care providers.

MS. TURBYVILLE: Okay. Anything else?

(No response.)

Okay. Great. So, moving on to 2c, which is the disparities have been identified. And, then, for those that are identified, the specifications, scoring, and analysis allow for the exposure, and so the identification and stratification of results. And, you know, we are talking about race, ethnicity, socioeconomic status, gender as relevant.

And I think this is an area that I don't know if Jeptha and Dave want to provide
some context of how the TAPs thought about
disparities when they were doing the ratings
on other measures because I think we haven't
kind of landed on a firm place on how it
relates to the resource use measures.

MEMBER PENSON: So, we basically
went to the document here with regard to
disparities, which really sort of -- I'm
looking for the actual line on disparities.
So, if disparities of care are identified --
the measure specification scoring analysis is
to allow for identification of disparities
through stratification of results.

And the key there was by race,
ethnicity, socioeconomic status, or gender.
So, we looked at it, at least in the Cancer
TAP, looking at it by disparities by patient
characteristics primarily and things like
gender, race, things that identify at-risk
populations. Or, if there was no mention of
it, was there a rationale not to have it?

In the Cancer TAP, you know, it
wasn't feasible because in many of them administrative data doesn't let you have anything in the way of at least race in SES. And for the most part, I don't think that was a deal-breaker, but it was definitely noted by the TAP.

MEMBER CURTIS: I think within the CV/Diabetes TAP really we didn't spend a whole lot of time discussing it just because we were talking about so many other things.

But I would argue that (a) these are differences, not necessarily disparities, and (b) that on average the "N" within any group that you are measuring is too small to really consider stratification.

CO-CHAIR STEINWALD:
HealthPartners, can you give us any information on whether the measure has been used or is being used to identify disparities?

MS. KNUDSON: I hope this directly answers your question and, if not, let me know.
We do not make any adjustments in risk adjustment for that, based on what Sally said when we teed-up the review, because we don't want to adjust away those factors.

Frankly, how we address disparities as a system is we started with data collection of race/language information, and have started with a lot of concentrated work on segmenting our measurement in the quality and experience domain and setting goals for eliminating disparities. We have not stratified this measure in the same way. That has been our emphasis in reducing disparities in actual care process.

Does that answer it?

CO-CHAIR STEINWALD: I believe it does.

MS. KNUDSON: Thank you.

CO-CHAIR ROSENTHAL: And I do think we have been operating under the principle that this could score low, but it is likely to score low because it isn't being measured or
collected anywhere virtually and wouldn't necessarily be the defining moment of our scientific acceptability.

Is that a fair --

MS. TURBYVILLE: Right. I think what we heard from the TAPs when the measure especially was being endorsed for use in the commercial population only, and understanding that a lot of the commercial administrative databases did not have the disparities information, I think we were even maybe voting moderate and some insufficient. I think it was up to the interpretation of the members.

And again, this measure is being presented as it has been tested. And so, it would be endorsed for use in commercial populations only. I think it was David who pointed out, how feasible would it even be? So, it is up to your interpretation on how that influences your ratings.

CO-CHAIR ROSENTHAL: All right.

CO-CHAIR STEINWALD: Okay. Can we
put it up?

(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, for this, we have 1 high, 8 moderate, 3 low, and then 7 insufficient.

CO-CHAIR STEINWALD: Yes, sir?

MEMBER PENSON: Can I ask a question? I know we are going through this as a TAP, but, I mean, I think we have had a very long, contentious discussion this afternoon about the scientific acceptability and validity. Is it possible we could do the yes/no vote now, so that the discussion is still fresh in our heads as opposed to in the morning, if other people agree with that?

CO-CHAIR ROSENTHAL: Well, again, unless we need some really substantial additional discussion, which I would suggest we have beat to death, I think the issues are very well-described and very well-defined. And I have a feeling that nobody is going to be swayed one way or the other by much further
discussion.

I would simply agree let's vote.

Right? That's the only thing left we have to do on this measure at this point for tonight, right?

CO-CHAIR STEINWALD: For tonight,

scientific --

MS. TURBYVILLE: Yes.

CO-CHAIR ROSENTHAL: Yes, to finish up scientific acceptability.

CO-CHAIR STEINWALD: Acceptability.

MS. WILBON: Bruce, a point of process? Can I just recap for you your overall vote? Remember that grid is based on your ratings for overall reliability and validity, for scientific acceptability.

So, for the overall rating for 2a, just to recall, just to jog everyone's memory, there were 8 high votes and 6 moderate votes, 4 low votes, and that was it.

And, then, for your overall rating for validity -- sorry, just a second -- you
had 4 high, 6 moderate, 7 low.

So, reliability was high to moderate and validity was moderate to low, predominantly.

It was 4 high, 6 moderate, and 7 low.

CO-CHAIR STEINWALD: So, on overall scientific acceptability, 1 yes, 2 no, and we vote again electronically, yes?

MS. TURBYVILLE: Yes.


(Whereupon, a vote was taken.)

MS. TURBYVILLE: So, we have 9 high and 10 low, and I think we will have to figure out --

CO-CHAIR ROSENTHAL: Well, it's obviously divided.

MS. TURBYVILLE: It's divided.

CO-CHAIR ROSENTHAL: And nobody is right or wrong.

MS. TURBYVILLE: Right.
CO-CHAIR STEINWALD: So, how do we proceed? Do we continue to discuss this tomorrow?

MS. TURBYVILLE: I think yes. I think it was just too close --

CO-CHAIR ROSENTHAL: I would recommend that we do the usability and feasibility conversation, despite the vote. I mean a 10-to-9 vote is a tie.

MS. TURBYVILLE: Yes.

CO-CHAIR ROSENTHAL: It's a tie.

MS. TURBYVILLE: That was your Steering Committee hat right there.

CO-CHAIR ROSENTHAL: Look, we should just do it. Okay?

MS. TURBYVILLE: Yes. It's so close. Yes.

DR. BURSTIN: And the other thing is we will prepare for you, just so you could actually take another look, we will actually prepare the votes, just so you can see it laid out, which I think it will be helpful.
CO-CHAIR STEINWALD: I think we are close to adjournment.

Before we do, all in favor of having business casual attire tomorrow? Could we have a -- no? Okay. Somebody got the memo. Lose the tie. Lose the necktie.

Any other administrivia?

CO-CHAIR ROSENTHAL: We can't leave anything in the room.

CO-CHAIR STEINWALD: Oh.

MS. WILBON: We do need to do a public comment for anyone else who is still on the phone.

CO-CHAIR ROSENTHAL: Okay. Well, let's quickly do that.

MS. TURBYVILLE: Operator, is it Nicole?

THE OPERATOR: Actually, it's Elizabeth.

But, again, it is *1 for any public comment.

(No response.)
And we have no comments.

MS. WILBON: Thank you.

Anyone in the room have any comments for the Steering Committee?

(No response.)

No? Okay.

CO-CHAIR STEINWALD: And can we leave materials in the room?

MS. WILBON: I wouldn't leave your computer, which you probably wouldn't, but --

Just a reminder, for tomorrow, we start at 8:30 and not nine o'clock. So, just a brief reminder. Breakfast starts at 8:00.

MS. TURBYVILLE: Adjourned.

Thanks, everyone.

(Whereupon, at 5:57 p.m., the Committee adjourned, to reconvene the following day, Thursday, June 30, 2011, at 8:30 a.m.)
hotspot hospital-level hospitalization
hospital's hospitals 190:8 415:3
hospitalizations 452:4
hospitalized 286:17
hospitals 204:1
124:17 244:11
246:19,21,22
249:2 252:11,12
252:16 265:9,11
265:15,16 266:6
267:16 283:17
293:13 375:9
411:8 417:21
hospital's 241:11
265:19
hospital-level
200:15 204:12
265:8
hospital-stay-plu... 208:1
hotspot 281:20
hour 68:4 88:22
128:4 216:11
217:11,12 313:15
hours 10:13 142:9
262:12 377:18
388:16
hour's 373:8
hub 203:11
huge 248:18 261:22
287:1 295:4 303:3
Huh 178:11
Human 385:7
hung 306:9
hurt 137:16,18
HWI 421:4
hypothesis 362:6
362:8
I
ICD 320:13
ICDs 355:11
ICD-10 263:14
ICD-9 73:4 236:20
237:20 239:14
242:15 317:10
icon 387:7
ID 290:19 291:8
293:5,7,8 313:9
334:9 334:22
335:1,2 337:11,18
338:2 372:11
idea 11:6 26:11
49:15 52:21 59:8
112:17 134:22
137:2 141:11
209:1 218:7
221:12 283:16
311:3 312:13
313:6 360:18
427:3 462:20
470:5
ideal 276:5
ideally 100:12
254:18
identical 280:18
323:19
identifiable 211:9
identification 68:17 72:9,21
104:11 107:11
120:13 265:18
339:14,18,20
476:21 480:17
481:12
identified 87:8 91:8
107:18 109:20
160:22 265:10
275:17 284:18,22
309:18 316:9,18
317:9 320:14
321:13 322:4
327:14,15,17
332:15 334:13
347:2 366:3 401:2
480:14,15 481:10
identifier 334:19
334:19
identifiers 124:2
266:10,15 282:16
identify 27:20
28:10,13 52:5
60:14 65:17
109:13 155:9
157:6 164:22
188:12 189:19
210:13 211:4
234:6 280:19
292:22 315:19
321:1 324:7,15
356:10 363:17
459:21 481:19
482:19
identifying 68:16
109:2 119:17
318:6 319:19
identity 371:8
idosyncratic 44:8
IDs 289:11,13
291:4,5,8 336:4
336:11 337:18
356:5 372:17
ifier 171:2
ignore 313:1
II 1:4
illness 260:3,9,12
262:19 283:10
293:22 379:20
illnesses 99:20
illustrate 384:6
387:19
illustrates 24:14
26:6 387:3
illustrating 382:19
382:22
imagine 120:8
150:12
imaging 236:15
277:15
immediate 190:14
immediately 344:9
iminently 161:5
impact 131:5 223:4
399:5
impasse 439:13
implement 367:7
implementation 50:9
implemented 48:15
52:2,4,4,10 440:4
440:18 445:16
448:17 456:11
implementer 367:9
implication 178:4
implications 364:10
imply 235:5 255:10
457:6
import 254:14
importance 4:9
5:13,15 6:8,10,22
6:24 7:21 25:11
28:17,20 32:3
53:20,21 54:2
69:18,22 71:12
74:8 76:4,12,14
76:16 78:5,16,20
79:2,3,6 80:12
80:21 86:1,9,12
87:22 88:14,22
89:7,9 159:10
192:4,6 194:3,10
201:13 276:15
279:11,12,15
280:6,8 322:19
323:2 379:2 389:5
398:10,11 399:6
400:14 403:11
424:9,11 427:11
445:13 455:14
464:21
important 21:21
22:2 32:22 33:6
33:12,13,18,22
60:10 70:8 71:10
76:18,18 77:6,13
78:6,7 79:10,11
79:11,19 80:2,2
80:18 81:4,5,13
82:11,12 85:10,19
85:21 87:16,19
99:5 101:3 105:19
192:7 194:5
208:21 212:7
223:5,14,18
245:18 248:14
269:2 273:13
276:4,12 280:12
299:3 323:9,15
361:18 362:3
399:8 403:18,22
448:18 465:3,3
471:4
impossible 350:21
impression 207:13
impressive 152:17
improve 148:18
157:3,7 165:1
improved 238:2
350:15 352:20
improvement 22:3
47:19 49:8 50:1,2
50:6 55:20
183:17 184:5,13
278:1 307:8
388:18 389:2
402:5 418:3
imputation 176:1
impute 176:13
263:11 336:3,11
336:11
imputing 175:12
176:9
inability 210:2
inaccuracies 170:19
inaccuracy 178:6
inadequate 189:6
inasmuch 42:22
incentive 170:8,11
427:21 463:2
incidentally-related 102:14
incidents 121:2
<table>
<thead>
<tr>
<th>longitudinally</th>
<th>292:11 302:17</th>
</tr>
</thead>
<tbody>
<tr>
<td>long-term</td>
<td>126:6</td>
</tr>
<tr>
<td>long-winded</td>
<td>299:13</td>
</tr>
<tr>
<td>loses</td>
<td>123:6 203:22</td>
</tr>
<tr>
<td>looks</td>
<td>26:4 50:7 76:16 87:14 122:9 126:22 327:22</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>1:22 11:20</td>
</tr>
<tr>
<td>losing</td>
<td>122:16</td>
</tr>
<tr>
<td>lost</td>
<td>222:8 304:16 304:17</td>
</tr>
<tr>
<td>Louisiana</td>
<td>410:10</td>
</tr>
<tr>
<td>love</td>
<td>454:8</td>
</tr>
<tr>
<td>loves</td>
<td>396:15</td>
</tr>
<tr>
<td>lowest</td>
<td>339:5</td>
</tr>
<tr>
<td>lowest-cost</td>
<td>247:12</td>
</tr>
<tr>
<td>lows</td>
<td>171:6,11</td>
</tr>
<tr>
<td>loving</td>
<td>526:20</td>
</tr>
<tr>
<td>low-cost</td>
<td>247:8</td>
</tr>
<tr>
<td>LTACs</td>
<td>226:6,10</td>
</tr>
<tr>
<td>luck</td>
<td>314:6</td>
</tr>
<tr>
<td>lumping</td>
<td>101:22</td>
</tr>
</tbody>
</table>

**M**

| MA | 2:14 3:3,16,24 |
| MACP | 1:18 |
| Madison | 62:22 |
| magnitude | 223:19 |
| main | 208:12 |
| mapped | 46:18 |
| manner | 214:3 |
| map | 110:10 133:10 |
| mapped | 87:4 |
| mapping | 87:22 |
| market | 117:6 |
| largest | 411:6 412:10,16 413:16 416:4,6 |
Page 531

Neal R. Gross & Co., Inc.
202-234-4433
| Page 537 | Neal R. Gross & Co., Inc. | 202-234-4433 |
| 7 | 7402:22 403:10 |
|   | 474:6 476:16,16 |
|   | 477:7 478:17 |
|   | 485:4 487:1,5 |
|   | 7th 1:10 |
|   | 70 301:16 |
|   | 73 4:5 |
|   | 74 4:6 |
|   | 75 4:8 103:21,22 |
|   | 120:7 220:21 |
|   | 308:10 309:1 |
|   | 407:8 |
| 8 | 8 441:4 446:17 |
|   | 485:4 486:19 |
|   | 8:00 490:13 |
|   | 8:30 490:12,19 |
|   | 80 133:5 155:5 |
|   | 80,000 455:22 |
|   | 800 152:14 180:12 |
|   | 82 133:6 |
|   | 85 116:17 204:19 |
|   | 220:16,20 |
| 9 | 9 3:2 472:7 475:20 |
|   | 475:22 487:14 |
|   | 9:00 1:11 |
|   | 9:12 9:2 |
|   | 90 4:9 287:15 |
|   | 443:19 444:6 |
|   | 90th 302:6 |
|   | 91 4:10,12 287:16 |
|   | 287:17 |
|   | 93 4:13 |
|   | 95th 254:22 255:11 |
|   | 98th 297:5 |

Neal R. Gross & Co., Inc.
202-234-4433
CERTIFICATE

This is to certify that the foregoing transcript

In the matter of: National Voluntary Consensus Standards

Before: NQF

Date: 06-29-11

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