THE NATIONAL QUALITY FORUM
+ + + + +
PATIENT OUTCOMES
MENTAL HEALTH STEERING COMMITTEE
+ + + + +
WEDNESDAY
APRIL 7, 2010
+ + + + +

The Steering Committee met at the National Quality Forum, Suite 600 South, 601 13th Street, N.W., Washington, D.C., at 9:30 a.m., Tricia Leddy and Jeffrey Susman, Co-Chairs, presiding.

PRESENT:

TRICIA LEDDY, MS, Co-Chair, Rhode Island Department of Health
JEFFREY SUSMAN, MD, Co-Chair, University of Cincinnati
SHEILA R. BOTTS, PharmD, BCCP, University of Kentucky College of Pharmacy
RICHARD J. GOLDBERG, MD, MS, Lifespan Corporation
WILLIAM GOLDEN, MD, University of Arkansas for Medical Sciences
ERIC GOPLERUD, MD, Department of Health Policy
MAUREEN HENNESSY, PhD, CPCC, Gardener Health Systems
DARCY JAFFE, ARNP, Harborview Medical Center
J. KAUFER, MD, FAAN, University of North Carolina at Chapel Hill
ANNE P. MANTON, PhD, Cape Cod Hospital
KATIE MASLOW, MSW, Alzheimer's Association
LUC R. PELLETIER, MSN, APRN, FAAN, Sharp HealthCare

GLEN PHILLIPS, PhD, Eli Lilly and Company
PRESENT: (CONT.)

HAROLD A. PINCUS, PhD, New York Presbyterian Healthcare System
ROBERT ROCA, MD, MBA, MPH, Sheppard Pratt Health System
JOEL STREIM, MD, University of Pennsylvania Medical Center
GEORGE J. WAN, PhD, MPH, Johnson & Johnson
CAROL WILKINS, MPP, Independent Consultant

NQF STAFF:

HEIDI BOSSLEY, MSN, MBA
IAN CORBRIDGE, RN, MPH
ASHLEY MORSELL
REVA WINKLER, MD, MPH

ALSO PRESENT:

LAURA GALBREATH, MPP, National Council for Community Behavioral Healthcare
RITA MUNLEY GALLAGHER, American Nurses Association
WILLIAM E. GOLDEN, MD University of Arkansas for Medical Sciences
VANESSA KUHN, MPH, Baltimore Substance Abuse *

DIANE MAYBERRY, MHA, RN, Minnesota Community Measurement
COLLETTE PITZEN, Minnesota Community Measurement
YNGVILD OLSEN, MD, MPH, Baltimore Substance Abuse *

*Present via telephone
C-O-N-T-E-N-T-S

Welcome ............................................. 4
   Tricia Leddy, MS (Co-Chair)
   Jeffrey Susman, MS (Co-Chair)

Project Recap ............................... 7
   Heidi Bossley, MSN, MBA, Senior Director
   Reva Winkler, MD, MPH, Program Consultant
   Ian Corbridge, RN, MPH, Program Manager

Depression Measures
   Measure OT3-002: Patient Attitudes Toward and Ratings of Care for Depression. .......... 36

   Measure OT3-011: Depression Remission at Twelve Months ...................... 63

   Measure OT3-012: Depression Remission at Six Months. ......................... 119

   Measure OT3-022: Depression Utilization of the PHQ-9 Tool. ............. 130

Readmission & Mortality Measures

   Measure OT3-001: Suicide Deaths of "At Risk" Adult Psychiatric Inpatients Within 30 Days of Discharge. ............. 189

   Measure OT3-003: 30-Day Readmissions. .......... 277

Substance Abuse, Patients Clinical Status, Recovery and Substance Abuse Treatment

   Measure OT3-010: Milestones of Recovery Scale ......................... 277
Measure OT3:013: Time from First Face-to-Face Treatment Encounter Buprenorphine Dosing. . .316

Measure OT3-017: Percent of Eligible Patients Who Transfer From a Substance Abuse Program to a Continuing Care Physician for Ongoing Buprenorphine Maintenance Therapy . . . . . . . . . . . . . . . . .340

Adjourn
CO-CHAIR LEDDY: Welcome, everyone. We are going to begin. We are going to begin with introductions. I'm Tricia Leddy, and I'm co-chair of this group.

CO-CHAIR SUSMAN: And I am Jeff Susman, your other co-chair. For those of you who I haven't met somewhere before, I'm at the University of Cincinnati and the chair of family medicine there. And I guess we'll just go around the room like this.

CO-CHAIR LEDDY: Ashley.

MS. MORSELL: I'm Ashley Morsell. I am on the NQF staff.

MR. CORBRIDGE: Good morning, Ian Corbridge, also on NQF staff working on the project.

DR. MANTON: Good morning. I am Anne Manton, and I'm a psychiatric mental health nurse practitioner at Cape Cod Hospital.
MS. JAFFE: I am Darcy Jaffe from Harvard View Medical Center.

DR. STREIM: I'm Joel Streim. I'm an internist in geriatric psychiatry at University of Pennsylvania.

DR. PHILLIPS: I am Glen Phillips. I'm a senior research scientist at Eli Lilly & Co.

MR. PELLETIER: I'm Luc Pelletier, administrative liaison at Sharp Mesa Vista Hospital.

DR. BOTTs: Sheila Botts, University of Kentucky College of Pharmacy, and clinical pharmacy specialist in the VA.

DR. KAUFER: I am Dan Kaufer, behavioral and geriatric neurologist at UNC Chapel Hill.

DR. GOLDEN: I am Bill Golden, general internist, University of Arkansas, and medical director for policy at Office of Medicaid.

DR. GOLDBERG: I'm Rich Goldberg.
I'm a psychiatrist from the great state of Rhode Island and head of a mental health or health care system, regional health care system, Lifespan Corporation.

DR. WAN: Good morning, everyone.

George Wan, senior director at Johnson & Johnson North American Pharmaceuticals.

DR. HENNESSEY: Good morning, everybody. I'm Maureen Hennessey. I'm a psychologist and health coach, and I'm with Gardener Health Systems Trauma Support Network in the University of Missouri in Kansas City.

DR. ROCA: Good morning. I'm Bob Roca. I'm a psychiatrist, and I'm also the vice president of medical affairs at Sheppard Pratt in Baltimore.

MS. WILKINS: Good morning, I'm Carol Wilkins. I'm a consultant. I do a lot of work on homelessness and mental health. And for a long time I was the director of policy and research at the Corporation for Supportive Housing.
MS. MASLOW: Sorry I'm late.

I'm Katie Maslow. I'm from Alzheimer's Association.

DR. WINKLER: Good morning, everyone. I'm Reva Winkler. I welcome you all back to work with us here at NQF. I'm the program consultant as I have been at NQF for the last nine years.

MS. BOSSLEY: Good morning, I'm Heidi Bossley, a senior director in performance measures at NQF.

MS. MAYBERRY: Diane Mayberry from Minnesota Community Measurement.

MS. PITZEN: Collete Pitzen from Minnesota Community Development.

Just to orient you where we are
with the process, and then we'll turn it over
to the NQF staff, we had two goals with this
project, one of which we are going to
concentrate on today which is the evaluation
of these candidate measures, and to decide
which ones we are going to pass along through
the process; the other that will I think come
up as we go through this, and I know some of
you are very interested in, are to identify
gaps, to look at areas that we really should
have measures, or there might be some outcomes
that we aren't assessing or measures that have
not been submitted, to identify those gaps and
to be able to document those.

Reva has told us that we will
probably want to circle around back to that,
so if we don't get to it, given the agenda
that we have, don't worry, we are conscious
that this is an important part of the process.
But if you see gaps or issues as we have the
discussions, I hope you will let us know so we
can keep that on the parking lot and make sure
that that is clearly identified.

Harold, do you want to introduce yourself, please, because we've got some new people.

DR. PINCUS: Okay, sorry I'm late, I'm Harold Pincus, I'm vice chair of psychiatry at Columbia University and director of quality and outcomes research at New York Presbyterian Hospital.

CO-CHAIR SUSMAN: So I think Tricia and I will try to do our best to keep us on time. We certainly envision the first evaluation discussion of the measure that we undertake will be a little bit longer, but we will have to keep a pretty brisk pace. I also would suggest that if on further reflection we look at one of the candidate measures and decide really it isn't an outcome but rather a process measure, that we deal with that up front, because that would be out of scope of the project. It could save us some substantial time in not having to go through
the whole process that is laid out before us
if we can say right up front, no, you know,
this really is a process measure after all.

With that, Tricia, do you want to
--

CO-CHAIR LEDDY: I think that
just following on what Jeff just said, which
is, if we do have process measures we can and
we feel that it is the only measure that has
to do with a certain subject, there isn't an
outcome measure, I think that in putting aside
the measure because it is process it will give
us potentially the opportunity and time to
identify what outcome measure we would like to
see, and therefore, use the time to not feel
bad about not having done that area, because
we can say, well, what really would be the
outcome measure, and then as in the report
there will be not only the measures that we
vote on but also a portion of the report that
will identify those specific gaps.

So if we can get very specific
about what we do want to see in an outcome
measure in a certain area, then I think that
will get us eventually to the goal of having
outcome measures in those we think we are
important rather than feeling that we have to
accept a process measure.

CO-CHAIR SUSMAN: So I think
without further ado we will turn it over to
Reva. Harold, do you have a question?

DR. PINCUS: What is the path
that may get us further to -- what is the
pathway to getting us further? Because since
we are not developing --

CO-CHAIR LEDDY: Right, it
wouldn't be our group, you are absolutely
right, Harold. So I will throw that one to
Reva.

DR. WINKLER: And I can catch
that one easily. Because it is a specific
deliverable on this contract, and the contract
is with the Department of Health and Human
Services, they have indicated that it is their
intent to take these recommendations and use
the development resources within the
Department of HHS to address those gaps. So
that is why it's particularly important and a
very specific deliverable for this project.

DR. PINCUS: So we need to devote
a significant amount of time -- what we have
is disappointing.

DR. WINKLER: Right, exactly.
And you are not alone. Mental health is not
the only sort of orphan child in this area.
We've got several topic areas in the other
parts of the project where there were no
majors either, and there are certainly some
large gaps. So we are - the initial work that
we are doing is looking at the measures we do
have and evaluating them, because they have
several months worth of process to follow with
public comment and voting and all of that. So
we need to get them going on that track. But
then we do want to put in some thoughtful time
around what would be the desirable outcome
measures that we didn't get, don't exist yet, need to be developed, and what would they look like, and to be as specific as possible.

So we will need to continue working with you as time goes on so we can develop that part of the project, but it's definitely a very important part of the project, so it's not an afterthought, it's not a sort of footnote. It really is one of the two main deliverables for this project.

DR. PINCUS: So it also occurs to me that as part of that discussion we should revisit the ones that were seen as being out of scope or into processing for ideas about where we should go.

DR. WINKLER: I think that is what Tricia was saying.

MS. MASLOW: What is going to be the process for that? Is that going to be -- are we going to have specific time on the agenda today? Or is that going to be a substantive meeting? How will that work?
DR. WINKLER: I think it will be a couple of things. Depending on how your meeting goes in terms of making progress on the agenda, if there is time I think it would be worthwhile to begin to address that, but I envision it more as follow up phone calls. Because we just need to get this work done and get it moving along, then we can take the time to do some thoughtful addressing of the gaps.

CO-CHAIR SUSMAN: And I think if something comes up, mention it, but we are not going to have time to fully work it during this process. If we get way ahead.

DR. GOLDEN: Before we get into the individual measures, will we have some time this morning to talk about some generic questions that the measure set raised? CO-CHAIR SUSMAN: I think what I would suggest we do is first allow staff to give us an orientation, and then perhaps as we work through the first measure to talk about those generic issues, because if we have something...
specific before us it will help us really
focus some of the discussion around that.

If we want to spend a few minutes
up front talking about those, I think --

DR. GOLDEN: I have a specific
issue that applies to several of the measures
that I would like to discuss.

CO-CHAIR SUSMAN: Sure.

DR. WINKLER: I think that just
in terms of project status and where we are,
just a couple of things. I know that there
was a great deal of response to our notice of
intent, thank you to all of the work that you
all did in notifying all of your contacts.
Ian spent hours talking on the phone with all
sorts of people who would have been previously
unaware of NQF and now are aware of the work
we are doing. So that was the first real
significant work for you when you were -- A
plus on that. So we did get a lot of interest
and a lot of new organizations that had not
been involved.
When it came to the actual submissions, again Ian did a lot of phone calls with people asking questions, and this issue of process versus outcome came up a lot, so there was a certain amount of filtering that happened at that point, because he would tell them, we really want the outcome measures. But again the actual formal submission process, which is not a trivial thing to do, measure developers put in the time and resources to submitting them. We still have some of the issues around process outcome, and on your phone call that we did three weeks ago or so you eliminated a few of those. That's a filtering process that seems to be ongoing so we will just have to address it as we go along.

You have really gotten yourself into the meat of the work by your initial evaluation of the measures. It's very important in the evaluation of the measures that we use, the measure evaluation criteria.
We have given you all copies of it. That was one of the directives from the CSAC was, be sure the committee members have it in their hand to refer to it. So there you go.

When we talk about the measures and their strengths and weaknesses, we really do want to couch them in terms of the criteria; it either does or doesn't meet the criteria. There is a problem with it because it doesn't address this, or it's really great because it does do this. So the criteria really are the framework around the discussion we'd like you to have, with the exception of importance to measure and report. There are no actual thresholds. So you do have to all agree that it is important to measure and report, and if you say it's not then that's it; we stop right there.

We will need you to vote on your evaluation of each of the four main criteria: importance, scientific acceptability, usability, feasibility, for each of the
measures. We are trying to give a little bit of hard data on your assessment to the subsequent audiences that are going to be reviewing it during public comment, during voting, for the CSAC, and for our board of directors. So we are trying to capture those ratings in a way that helps underpin your ultimate recommendation.

There is no numerical adding, subtracting -- you have to get a majority of them have to completely meet the criteria, or if you get half Cs, that's a good -- none of those -- there are no magic formulas. There is no math to this. They should be the things you are thinking about and considering and balancing, realizing there will be tradeoffs. We have yet to see a perfect measure. We just don't see them. There isn't anything that completely meets all the criteria every time. There are some that come close, but none that have hit them all. So it's a balancing act, but the rationale for
recommending the measure should be supported by, we feel it's very important. We feel the scientific acceptability is good enough. We feel it is usable and provides meaningful information. We believe it is feasible to do. Those are the kinds of right reasoning, even though the actual subcriteria may not be perfect for each measure.

So I think you have all had experience doing the several measures that were assigned to you in your workgroups. The purpose of that was to spread the work out. You've got, what 18 measures? Seventeen measures. And asking each of you to be intimately involved with all 17 was overwhelming. So by breaking it out we asked each worker to spend some time with a limited number of measures.

As we go through these today we will expect the members of the workgroup who are really familiar with the measure to kind of lead that discussion and help the rest of
the group understand the strengths and
weaknesses of the measures as we go through
them so at the end of the day the ratings and
recommendations reflect the input of everybody
on the steering committee.

So that is essentially what we are
up to today. We are going to go measure by
measure. We will help you through the first
couple, and there is a learning curve so it
will take a little longer. But it's very
important that we hear your issues. You all
are here representing different stakeholder
perspectives. There should be some
disagreements among you; there should be
different points of view, and we need to make
sure that those are brought to the table and
that everyone has a chance to speak them and
have them heard. That is a fundamental part
of NQF consensus process is to have all of
that diversity of input.

So that is the reason you are
here, so we really do encourage everybody to
speak up, and if you are going to say
something that disagrees with the rest, please
do it; that's what you are here for.

So I think in terms of background
that is kind of the summary of how we got here
and what we are planning on doing today. And
I think, does anyone have any questions?

Okay, Bill had a question. We'll see in a minute. Bill, did you have a
question?

DR. GOLDEN: I was going to - did you answer his issue? I was going to ask a
question about scope.

DR. WINKLER: Okay.

DR. GOLDEN: About measures in
general.

MR. CORBRIDGE: Can we just hold
on one second? We are actually trying to get
the phone lines hooked up. So we just have to
go through the process of talking with the
actual operators. And I guess while we are
waiting for that, I'd just like to follow up.
Once again thank you everyone very much for all of your participation so far and dedication to the project. A couple of housekeeping issues. This is actually not NQF's workspace here. We are actually in a law firm here. So they requested -- they are obviously having some meetings today as well -- so they requested that if any individuals do have to make a phone call, need to step out, if you actually need to make a phone call if you can go down to the main lobby. They just don't want to have people coming in and out here, and they'd like us not be out in the lobby making phone calls. So just one thing as indicated by other staff members a couple of members of the steering committee are unable to make it this morning. I know Dr. Thompson had some car issues, and Dr. Goplerud had some previous appointments, so they should be coming later on today. Maybe once we get this phone line hooked up and answer some questions, we will go over some of the
documentation you have in front of you as well as what we are projecting on the screen. So hold on just one second.

(Technical interruption)

MR. CORBRIDGE: I'm sorry, we seem to be having some issues. So if you would like to go ahead, Dr. Golden, and just ask your question, I will see if we can get this issue with the phone figured out and go from there.

DR. GOLDEN: The question for the staff in terms of just the measures themselves, the unit of measure is sort of interesting. Is the NQF still using for outcomes and process decision making or impact the provider as the unit of measure of the community? Because some of these measures were starting to go toward community units of measure rather than provider units of measure, and I was just curious where you all are?

DR. WINKLER: Well, we are actually expanding. Traditionally in the past
most of the measures that NQF has addressed or
endorsed are focused on some level of
providers, whether it's the hospital, the
individual clinician, the group, the facility,
whatever. However one of the national
priorities partnership goals and priorities
areas is around population and health, and we
have -- and I guess you weren't at the meeting
on the call, Bonnie Zell who oversees our
population health work here at NQF is helping
us move into that population realm. So the
fact that some of the measures may be more
appropriate for communities or more population
rather than provider specific is something
that NQF is quite open to entertaining.

DR. GOLDEN: But it changes how
you apply the criteria, so that's why I was
asking. And I guess the other follow up, the
other issue there, is you talk about the
usability. Some of these measures are valid,
but they are useful in the process of care
rather than evaluating the care, and I was
I just curious if you had thoughts on that as well. I think people are coming to you with tools to be used and endorsed as opposed to a measuring tool.

DR. WINKLER: Right. Well, I think underlying all of it, remember that NQF's goal in all of the quality enterprise is, we endorse measures used primarily for accountability and public reporting, so using the measure in that way, and suitability for being used in that way, is really embedded in many of the criteria, and certainly the one on usability. So the 3(a) criteria on usability is, is it useful for a variety of stakeholders in terms of actionability, and is it usable, understandable, meaningful if it's used in public reporting?

So that is really the kind of context you need to be thinking about these measures going forward.

DR. GOLDEN: And my only comment is on the usability statement in three. It
says, why they actually find them useful for decision making, but it didn't put in there for - and you need to maybe -

DR. WINKLER: Right, that's a good point. I think, Bill, that sort of up front as the overlay is the public reporting part, but you are right, embedding it specifically in the criteria statements would be a good idea.

CO-CHAIR SUSMAN: So are there any other general questions? I think, Harold, you did have a question or comment?

DR. PINCUS: In terms of the forms to be filled out, are they totally a result of - do they go through editing or someone intentioned by staff.

DR. WINKLER: No, essentially what we have done is taken the information submitted by the measure developer and embedded those in the form. Those are the unchangeable parts of the form.

The areas that have the rating,
and there are blocks for TAP comments if there
is a TAP project, or the steering committee
comments, those will be putting in your
assessment. So this is a document that grows
through the process. It starts with the
information that is submitted, then the
evaluative elements are added to it as it goes
forward through the process.

DR. PINCUS: A measure developer
unfamiliar with NQF is kind of clueless as to
what you are going for. Basically you're
stuck with what they have even though they
might have had some different measures.

DR. WINKLER: We can certainly
feed that back to the measure developer and
make the suggestions, and we do have a
mechanism by which they can edit it or change
it and revise things, to change the
information that is there in their portion of
it.

MS. BOSSLEY: And Ian spent some
time doing that already. So if we saw a big
section blank, so for example, the testing pieces, the reliability, validity. He went back and had a conversation with them to make sure that indeed that does need to stay blank because they haven't done that testing. If they haven't, that's where he marked it as not tested.

DR. PINCUS: A few more questions, one, is the absence of information on something indicate that there isn't any information or that they didn't put it in. And the second thing is, particularly with regard to the harmonization piece, how do they know what else is at NQF?

DR. WINKLER: Well, I can answer your second question first. And that is, NQF's website actually has a searchable database on it. And you can search and find out what measures NQF has endorsed. The NQF staff also does the backstop on that.

DR. PINCUS: So that does have editing by staff.
MS. BOSSLEY: So what we would do is if they included it in a separate document or in some way indicated to you that there is a comparable measure.

DR. PINCUS: Is that what's being looked at?

DR. WINKLER: Correct.

MS. BOSSLEY: I don't think there was anything. Ian, can you clarify?

MR. CORBRIDGE: If there are similar measures to the measure that was submitted to this project, it should be - I can't remember the actual page number, but at the very end of the evaluation document it indicates if there are similar measures what those measures are, providing the NQF number and some specs for that.

MS. BOSSLEY: We try to do that work for you as well, to try to help identify --

CO-CHAIR LEDDY: Are there any other questions before we launch into the
first measure?

    DR. WINKLER: Ian, did you have anything else?

    MR. CORBRIDGE: All right, if there are no more questions, I guess we'll just go over some of the documentation that is in front of you as well as some of the documentation that we will be projecting up on the screen.

    In front of you you should have an agenda for the day as well as a breakdown of the measure evaluation workgroups within that indicating what members of the steering committee were a part of that workgroup, as well as what measures for the title as well as the NQF initial tag number with that measure.

    As we've already gone over you do have a copy of NQF's measure evaluation criteria, so we just hope that you will be able to refer to that as we go through this process, and I'm sure you're probably had to use it. We provided it in digital format as
you were reviewing these measures.

Projected up on the screen we will be showing the survey of the subcriteria that members of the steering committee worked on. We tried to capture all the information that was submitted to us yesterday, and we will be projecting that up on the screen, and hopefully that will serve as just a platform to help facilitate the discussion and dialogue. And from that standpoint we will just kind of be able to dive deeper within each measure and workgroup.

We also - and we'll project it a little bit later on - once we get to the points for the voting process, NQF, we are going to be capturing the votes for each measure. We will be looking at issues of importance, scientific acceptability, usability and feasibility. So those are the four main NQF evaluation criteria. So we will project that a little bit later on when we get to that point.
For starting off each measure I'm just going to open it up or read off the number of the measure that we are going over as well as the title. I will give you a brief description, a numerator and denominator statement if that will be helpful for members. From that point we will really open it up to the workgroups to really kind of head off and further dive into that discussion. We tried to seat each workgroup next to each other so there can be conversations and dialogue amongst each other, and we will go from there.

DR. PINCUS: Are we breaking into workgroups?

MR. CORBRIDGE: No, not specifically breaking into workgroups, but as we are talking if you would like to share some information, we just wanted to make sure that you were sitting next to each other if there was information you wanted to share or pass along to each other.

Any additional questions regarding
that process? Does that seem clear to everyone, like it will work?

Heidi is there any way that we can

--

CO-CHAIR LEDDY: Do you need it bigger?

MR. CORBRIDGE: I don't know if this was - we tried to have as quick a turn around time as we could, so I emailed this out to every member yesterday, and I do have some limited hard copies, maybe I can just pass this out to the back of the room as it is difficult to see back here.

(Off the record comments)

MR. CORBRIDGE: Pass these around.

All right, are there any additional questions before we begin looking at the first measure? And so as we go through the process, when we get to the measure, if the measure developer is on the line or is here in person, if they would like to make
just a brief presentation, just talk about the
measure, the process, they are more than
welcome to if that's what the steering
commitee would like as well as later on
throughout the process if there are any
questions from the steering committee members
please feel free to ask them of the measure
developer through the dialogue or at the end
of there are questions that are raised.

If there are no more questions I
guess we can start moving forward, to keep on
time.

CO-CHAIR LEDDY: So Ian, you are
going to describe each measure first?

MR. CORBRIDGE: Correct, yes. So
we are going to go over each measure first.
I'll just read a brief description of it and
we will move forward from there.

And I don't know if the
representatives from Johns Hopkins University,
are you on the line?

(NO response)
MR. CORBRIDGE: I know they were hoping to make it. But it doesn't seem like we have anyone at this time. So we will just proceed forward with the measure that we have first on the agenda, and that is measure number two, and that is patients' attitudes towards and ratings of care, depression.

MEASURE 0T3-002: PATIENT ATTITUDES TOWARD AND RATINGS OF CARE FOR DEPRESSION

(PARC-D 30) QUESTIONNAIRE

MR. CORBRIDGE: And so that was the brief title. Just a brief description, and this is the information being projected on the screen for that measure, and that's the information for the subcriteria.

A brief description of the measure is, developers employed a comprehensive patient-centered approach, developed an instrument to measure primary care patients' attitudes towards and ratings of care for depression.

To help prioritize attitudes,
additional domains including 126 items identified previously in focus groups, we asked patients to rate the importance of each aspect of depression care on a five-point scale. Items were ranked according to a mean score, and the percentage of patients ranking the items as extremely important. The items were selected for inclusion and an instrument to measure patients' attitudes toward depression care based on importance ratings. We performed reliability and validity testing on a scale compromising our 30 most important items, and a shortened version that included 16 items. So they do go on further. Let me just read to you the numerator statement for that measure.

So the numerator statement for this measure reads, patients in primary care settings who complete a depression screener such as a patient health questionnaire PHQ-9, and score greater than or equal to five indicating a mild or moderate depression.
Additional target populations include primary care patients with clinically significant depressive symptoms, minor depression, dysthymia, major depressive disorders, in partial remission or mixed anxiety depressive conditions.

The denominator statement for that measure reads: all primary care patients. So that's just the intro for that measure. That measure resided in workgroup one, and members from workgroup one, I'm sorry, would you mind raising your hands? It's on the top of the slide, but just members from workgroup one? All right, wonderful.

So that's just a brief way to start off the measure. And we can look up on the screen, the initial results for the subcriteria for the main evaluation criterion, importance projected up there. And if the workgroup would like to add any insights on that.

CO-CHAIR LEDDY: We would like
to comment on whether this is first, on
whether this is enough toward an outcome
measure to or whether it's clearly process at
this point?

MR. CORBRIDGE: Correct, I think
that would be a wonderful idea.

CO-CHAIR LEDDY: Maybe would
anyone in the workgroup like to comment on
that?

DR. PINCUS: I actually didn't
see how it was a performance indicator at all.
It's a research tool to assess patients'
attitudes toward depression care. And it
wasn't clear to me how insomnia - what one
would expect, to monitor everything in a
client someway.

DR. GOLDBERG: I think our
summary says a lot. It looks at the patient's
outcomes. You've scored it as two minimally,
one not applicable, and one partial. I
thought it was an interesting measure. My
comment is on engagement, it had something to
do with the engagement of a patient. I didn't see it as an outcome measure, primarily.

DR. WINKLER: Well, if you recall, when we had our conversation in November, we discussed the wide variety of outcome measures, and types of outcome measures. And you all spent a lot of time expanding those fairly broad categories that did include patient experience with care, patient adherence, all of those sorts of things, as a result. So you all kind of defined outcomes that way. So the question is, does this fit?

DR. PINCUS: I can see how one could use it as an outcome measure. But as currently defined, it's not even a measure of depression care, it's a measure of depression attitudes.

(Simultaneous speaking)

DR. PINCUS: Well, but it's actually - so it's heterogeneous in that way.
(Simultaneous speaking)

DR. PINCUS: But my sense was it didn't meet the importance criteria.

CO-CHAIR SUSMAN: So it looked to me at least in the description from staff that there were sort of two components to this. One was attitudes toward and the other part was the perceptions of care itself. And that to me is problematic, because you are mixing an outcome and a process essentially, or an attitude about their depression, so I was just wondering whether this was even within scope, given that complexity. But I'd be interested in the folks who really spent a lot of time with this.

DR. HENNESSEY: I have a question. Is it true that the mission of this group is to look at measures dealing with patient engagement of care? Because if it is, this may partially address that, but as you pointed out, it looks like it's measuring two different variables, so you can have some
murkiness there. But is that --

      DR. WINKLER: You all have
defined outcomes to include patient experience
and care as an outcome of health care
delivery.

      DR. HENNESSEY: Which makes
sense to me, but whether or not this is the
measure for that because of that is the
question on the table right now.

      CO-CHAIR LEDDY: I think on the
phone call though, that's when you are
referring to, Reva, where we were fairly
broad?

      DR. WINKLER: No, not the phone
call, your meeting.

      CO-CHAIR LEDDY: I think the
discussion on the phone call at least was that
we wanted to be somewhat broad and inclusive
if there was any question because we didn't
have a lot of detail about the measure, and
that would give us more things to consider at
this meeting where we would be more strict and
DR. PINCUS: I don't think it's necessarily just whether patient engagement - for example, one of the items is, faith in God will heal my depression. I'm not sure how that is related to an engagement that you monitor for quality.

DR. GOLDEN: I don't think this is in our scope. I think if I were a provider the information for this survey would help me understand the patient, but it's not going to make a lot of reflection on my management of the patient or assessment of how I manage the patient. So I recommend that this would not be considered.

DR. GOLDBERG: I don't know if you want to go further. Though our process would be if it doesn't pass the first step -

(Simultaneous speaking)

DR. PINCUS: One other point there is that on the harmonization it goes further, it raises a sort of broader issue
about harmonization is, it wasn't mentioned
and I'm not sure whether in TAPS or ECHO could
have overlapped with some of the elements of
this as well. But it seems to me at least of
all the items that do relate to patient
perceptions of care, these overlook what the
overlap was. I consider these not with all
the others but just as a process issue.

DR. GOLDEN: You know I just
wanted to just introduce, just looking at the
measure evaluation criteria on the second page
it talks about these intermediate types of
process outcome measures. It seems like this
would fit under the patient experience or
assessment of patient experience of health
care outcomes and values. The values piece
will address that question of your faith in
God, things like that.

CO-CHAIR LEDDY: But I think
that the measure evaluation criteria is all
kinds of measures, not just outcomes. So this
is a generic tool and could be used for other
groups that are doing the process measures as well as outcome, whereas what our assignment is is to really stick to outcome measures, and I think the Donabedian definition that they gave us at the first meeting was really good. It says, outcome refers to changes, either desirable or undesirable, in individuals and populations, that are attributed to health care, and even down the paragraph it says that an outcome would be something that the patient is seeking care for, like improvement in function, that sort of thing.

So if we stick to - I think that is really what they want us to focus on as far as outcomes, because there are other groups that are going to be looking at process, I assume.

DR. GOLDEN: A comment on George's comment. The difference I think though here is that on the values piece, I think that we often are assessing the respect of the values in the process of care rather
than the values themselves. So this tool assesses what those values are as opposed to how the health care system dealt with those values. And I think that's a difference in terms of how the measures deploy.

CO-CHAIR SUSMAN: It would seem to me that feedback to this measure developer might be that there are indeed some important elements of the experience of care that perhaps a submeasures within this could be used as a valid measure of patient experience, but there are other elements that are clearly outside patient experience and led the committee to say this wasn't a useful outcome measure.

But one can imagine many of these sub-elements they talk about - health care providers' interpersonal skills, their perception of treatment and effectiveness might be very important measures given our broad definition of outcomes. But the admixture of other things like intrinsic
spirituality probably made us less excited about this measure.

DR. PINCUS: It is not designed to sort of pull out individual items. As broad domains of potential interest, yes. But as a measure, no.

CO-CHAIR LEDDY: So is the next procedure that we vote, or have we achieved consensus?

DR. WINKLER: It's sounding like we do need to vote on the importance to measure and report, because if it doesn't pass then we are done with this and we can move on to the next one.

DR. PINCUS: So is it a majority?

DR. WINKLER: Typically a majority.

DR. PINCUS: A simple majority?

MR. CORBRIDGE: Chris, I guess before we get to that vote we do need to make sure we open up for public comment if there is anyone on the phone line or anyone here who
would like to comment on the measure under discussion.

(No response)

MR. CORBRIDGE: So NQF staff, I don't know, Heidi, if you are able to - there is just a show online, up on the screen, the measure voting tool. So this is what staff have on their screens. So we are just going to capture throughout the process the information and dialogue that is discussed here as well as the votes for each. So we will keep that. And so for this measure, if we are just getting to importance, we will just capture the importance vote, and then say that it was tabled due to not meeting importance.

Yes.

DR. ROCA: Is this an issue of importance or scope?

DR. WINKLER: The two kind of have a not a sharp edge between them. You can eliminate it on scope if you are saying that
it's not an outcome measure. On the other hand what I heard more was that maybe it's within the scope of the mental health outcomes, but that this isn't a performance measure that is important to measure and report for public reporting that will provide meaningful information to audiences.

DR. HENNESSEY: You know my dilemma in this is from the importance perspective I do think in terms of patient engagement attitudes are quite important. From what I'm hearing from this group that's really looked at this measure, though, it sounds like the psychometric properties of it are not well delineated. That's one issue.

DR. PINCUS: Does it measure performance? It's unclear whether doing something, what the results would be that would be a good result.

DR. KAUFER: When I look at - there are seven main domains that these items cover. And I look at these, and these just
strike me as being independent variables or covariants, potential covariants, than they are dependent variables.

DR. MANTON: I am wondering, in terms of the process, if this is - if we vote to not accept this, what happens to it? I think there are some good elements to it. So will there be feedback to the developer? It almost feels like they could create two tools from it, one just dealing with the outcomes, and then one dealing with the patient issues. And so I'm wondering if that is the kind of thing that happens if it's voted down, or is it just, sorry, but we are not accepting it?

DR. WINKLER: No, actually two things happen. We do let - directly advise the measure developers of the feedback from the steering committee. But it's also your discussions included in the report, and when the measures go out for public comment, the information is available and we actually encourage people to comment on measures that
were not recommended. So there are
opportunities for this to have an ongoing
discussion about the usefulness of the
measure, providing that feedback. So it
doesn't just drop, no.

CO-CHAIR LEDDY: So is there a
certain way that we can vote? Or can we just
entertain a motion from the workgroup about
this measure and we'll all just vote on it?
Is that acceptable, Reva?

DR. WINKLER: What I need is a
vote from all of you, does this meet the
importance criteria, yes or no?

DR. GOLDBERG: The way I can say
that is - if you look, the relationship to
outcomes is so low that that is the important
category, in importance - tied together, so on
that basis -

DR. GOLDEN: The question I had is
on the importance measure. I have read
through the criteria. I could not tell if the
topic was the important issue or whether the
measure - it was very uncertain as I was
filling out the questionnaire.

DR. WINKLER: In this particular
case the importance is addressing the topic,
all right? So is this an important topic to
measure? Is there a variation in care? Is
this the topic that is being measured, have
relationship to outcomes? You start, when you
move into the scientific acceptability
criteria is when you are talking about this
measure specified with this numerator.

DR. GOLDEN: What is the topic?
Is the topic depression? Or is the topic
attitudes toward depression?

DR. WINKLER: Well, that's the
question I think for you all to consider.

DR. HENNESSEY: The question
down the line is whether or not this is
important.

DR. GOLDBERG: Aren't there seven
domains?

DR. HENNESSEY: I'm reading over
your shoulder here.

DR. GOLDBERG: Are you going to have to have us vote on every one of these elements for this meeting?

DR. WINKLER: The four elements.

CO-CHAIR SUSMAN: But one is the entrance point to the rest, correct?

DR. PINCUS: Threshold.

CO-CHAIR SUSMAN: Threshold,

thank you.

MS. BOSSLEY: There is perhaps a way to maybe handle these, if we are going to go through a lot of these I think. So for the ones that truly would be a process measure, I think you should determine if they are in or out of scope. Probably are going to say they are out of scope. You won't do any voting. They won't appear in a report. They won't go further. Any feedback will go back to the measure developer, so that they know what you thought. And that's it.

But for the ones that fit within,
and this one I would say kind of fits in within looking at how you define an outcome, didn't go far enough, and that's part of it, I would recommend we do have at least a vote on importance. Because then it goes out in the report, it's included in the final document, and that information is put out to the public. And you can include research recommendations of where you think this measure didn't go but we need to go next. But I think this one is one of those kind of squishy ones that it would be good to include out in the public - you know, out in the public and member comments. Does that seem to make sense?

DR. PINCUS: Is this a motion? Is this how you proceed? What rules are followed?

MS. BOSSLEY: But I'm asking our chairs too, does that seem like a reasonable approach?

CO-CHAIR LEDDY: So it sounds
like we have a choice of whether to vote on
importance or determine it be in or out of
scope. So would any of the workgroup members
like to recommend one or the other that we
consider, either that we vote on importance -
determine first whether this is in or out of
scope as an outcome measure.

DR. GOLDEN: One more question
here for Reva. There is under importance, you
have three elements. There is no global vote
on importance. So are you asking us to vote
on the global?

DR. WINKLER: Yes, that is what
we will be asking you to do.

DR. GOLDEN: So the impact could
be high but the other - okay.

DR. GOLDBERG: So we have a
measure here that because we decided
engagement was within scope, maybe within
scope, but because this particular measure's
relation to outcomes is so low, that its
importance, bundled score of importance, is
going to be very low. It's within scope but of such low importance that we are not going to proceed to the additional measures.

DR. KAUFER: Is that a motion?

DR. STREIM: As a general procedure, just to get us through all these measures we are reviewing, what I would like to propose is that we first consider the scope question on all of these as a first cut, and then if it is within scope then we look at importance to measure. And I think that might move it more quickly.

With respect to this particular measure we are looking at, well, actually maybe I'll come back to that. Harold, did you have a comment on the process?

DR. PINCUS: I agree that we are going to do that, at least from my thinking. I hate to be picky about this. But we need to have a fairly specific definition of what scope is, and when we talk about measure focus, what that means that we are determining
the importance of. Is it the topic of depression which is basically what the evidence that they've marshaled showing that depression is a big problem and that there is bad care. Or is it the focus being the measurement of attitudes and engagement of care as demonstrated by this measure?

MS. BOSSLEY: Right, so if you look at the measure criteria, the extent to which the specific measure focuses is important.

DR. PINCUS: What does measure focus mean?

MS. BOSSLEY: So it would be the patient attitudes toward and ratings of. Literally it gets down to that granularity.

DR. PINCUS: So it's not depression.

MS. BOSSLEY: It's not depression.

DR. PINCUS: Okay, that's helpful.
MS. BOSSLEY: So it's getting at the aspect of care that we are really trying to measure here, is that important.

DR. PINCUS: And the problem is that this is an and rather than an or. That patient engagement, yes; attitudes towards, no.

DR. STREIM: I think another comment about scope as it relates to this particular measure, if scope actually for outcome measures, and really does depend on the goal of - the goals of treatment. So when you are looking at the importance to measure an outcome, you have to have some sense of what the goal of that treatment is, otherwise we don't know what we are talking about.

So in this example, let's say in the course of treatment for depression perhaps a patient becomes - has a change in their attitude, and values treatment for depression more or less as a result of their own experience during the course of treatment,
that is a kind of outcome. But we don't
really as a field, we don't have an
established set of goals about whether we
should be getting our patients to love
treatment for depression or hate it. We do
care about things like engagement, but I think
the way this particular set of - this measure
with its 126 various independent variables is
not linked to a widely accepted goal of
treatment, mental health treatment. So
therefore it may be interesting, but I'm not
sure what health care consumers in general
would say if they could vote on what kind of
attitudinal changes we would hope for. That
is kind of far afield of where we are with
outcomes right now.

DR. HENNESSEY: I have a question
for the people who are really looking at this
- a measure says that they are developing -
that they are looking at treatment
effectiveness, treatment problems, patient
understanding about treatment, health care
providers, interpersonal skills. Do they demonstrate, do the developers demonstrate that there is evidence that what they are measuring has an impact on engagement and outcome?

DR. PINCUS: No, and that is the problem.

DR. HENNESSEY: Thank you.

DR. PINCUS: At least within here.

Under the criteria as a process if it is linked to outcomes then it is appropriate. But there is no data here that says that.

DR. HENNESSEY: Thank you.

CO-CHAIR SUSMAN: And just one final comment. If you look at the elements in the numerator, there are things like: faith in God will heal my depression. Prayer alone can heal depression. Thanking God helps depression to get better. Asking God for forgiveness will help heal my depression. And while they may be important elements, they aren't outcomes, and I don't suspect that I
can influence those effectively during the course of treatment. They are intrinsic spirituality elements.

So for me again it gets back to the motion I think on the table here which is, I think, we've got some elements of engagement which are very important, but we also have some intrinsic elements that I don't see directly related to outcomes. So I think we really should circle back to Richard's motion.

MS. BOSSLEY: So your motion is to vote on these, correct?

DR. PINCUS: Yes.

DR. WINKLER: So it is a yes-no vote. So essentially we will ask you, how many of you agree that it meets the importance criteria?

(A show of hands)

MS. BOSSLEY: Any abstentions?

DR. WINKLER: Eighteen nos.

MR. CORBRIDGE: Dr. Thompson, are you joining us on the phone?
CO-CHAIR LEDDY: Okay, so we are done with our first measure. How long was that?

Our objective for us would be a little more just getting through the process.

Eric, did you introduce yourself?

DR. GOPLERUD: Yes. I just arrived, Eric Goplerud, I'm a research professor at George Washington University, and I primarily work on substance abuse issues, though I have also done mental health performance measurement work. And this being NQF, I have no conflicts to declare.

CO-CHAIR LEDDY: So we are going to move on to our second measure to consider, and Ian is going to take us through the basics.

MR. CORBRIDGE: Thank you. So we are moving on to measure number 11. This was submitted by Minnesota Community Measurements. The measure developers have actually joined us today. So we may want to open it up to them to see if they would like to talk about the
measures briefly, or if at the end we can have
the dialogue with the measure developers as
well.

MEASURE OT3-011: DEPRESSION REMISSION AT
TWELVE MONTHS

MR. CORBRIDGE: So moving along to
the measures presented up on the screen,
measure number 11, depression remission at 12
months, so just a brief description of the
measure. Adult patients aged 18 or older with
major depression or dysthymia, and an initial
PHQ-9 score less than nine to demonstrate
remission at 12 months defined as a PHQ-9
score less than five. This measure applies to
both patients with newly diagnosed and
existing depression whose current PHQ-9 score
indicates a need for treatment.

The patient's health questionnaire,
PHQ-9, is a widely accepted standardized tool.
All rights reserved. This measure
additional promotes ongoing contact between
the patient and provider as patients do not
have follow up PHQ-9 scores at 12 months, plus
or minus 30 days are also included in the
denominator.

So just a brief description of the
numerator statement. It reads: adults aged 18
and older with a diagnosis of major
depression, dysthymia, and initial PHQ-9 score
greater than 9, to achieve remission at 12
months as demonstrated by 12 months plus or
minus 30 days a PHQ-9 score less than five.

The denominator statement reads,
adults aged 18 or older with diagnosis of
major depression or dysthymia, and an initial
PHQ-9 score greater than nine.

That's just the initial specs from
that measure, and that is once again measure
workgroup number one.

CO-CHAIR LEDDY: So do we want to
invite the measure developers to present
before we consider --

MR. CORBRIDGE: Yes, if that is
agreeable to the workgroup, if you'd like just
a brief, five minutes, come up and present
that, if that would help move the discussion
forward.

DR. GOPLERUD: I know you folks
have come from Minnesota, and we want to say
hi to them and all of that. But I'm wondering
if there are questions it might make sense to
ask them. Whereas I'm not sure that it may be
in some ways the converse, I may be trying to
read too much, maybe the converse of the first
measure, in that there may not be a whole lot
of question about it, and so if what they are
doing in some ways is say preaching to the
choir, it's wonderful to preach but it may not
be necessary. So I kind of don't want to take
15 minutes of our time having them present
things where there really isn't a whole lot of
controversy.

CO-CHAIR LEDDY: That is a good
point. And this is intended to be
interactive, I think.

DR. GOPLERUD: So I would kind of
recommend that we at least have a preliminary
discussion of the measure and decide if we
really need a pitch on it.

CO-CHAIR LEDDY: Okay.
MR. CORBRIDGE: I think it has been
run both ways at different committees. So
it's really up to the judgment of what the
workgroup would like to see. So if you feel
it would be more informative as I guess you
indicate Dr. Goplerud to have that discussion
afterwards, or ask questions as needed, then
we can proceed with that, that would be more
helpful. If more clarification is needed at
the end, then we can proceed that way.

DR. GOLDEN: I do have a question
for them in the beginning. In the beginning
it said this measurement tool is widely
accepted, quote unquote. So the question is:
what does that mean? And what major
specialist societies have endorsed it for its
use as a standard of care?

MS. PITZEN: I guess I just wanted
to say in our state it's a widely accepted tool that many practitioners are using. We have 233 clinics submitting data to us currently.

DR. GOLDEN: Have any national medical societies endorsed this as a standard of care?

DR. GOPLERUD: The American Psychiatric Council, the PHQ-9, and they have done collaborative studies on - I'm not sure that they said that is, but they have used it in a major research --

DR. GOLDEN: Are they saying that every patient should be having this done as a standardized tool?

DR. GOPLERUD: On this specific measure, I don't think so.

CO-CHAIR SUSMAN: No, but as a tool for measuring outcomes.

DR. GOLDEN: But this is important, because if this is a performance measure that we endorse, it becomes a standard of care. So
I'm asking is this considered a standard of care to use a standardized tool in practice like this?

CO-CHAIR LEDDY: Can the NQF staff address that?

DR. GOLDEN: You're basically requiring people, an insurance company to say, NQF has endorsed a measure saying anybody with this diagnosis should have this tool being used.

MR. CORBRIDGE: No, no, that's not what it says. These NQF measures are really up to individual entities to adopt the measure if they would like to at their facility. So an NQF endorsed measure doesn't mean that it is put out there and then everyone has to abide by that and measure that.

DR. GOLDEN: I disagree with you. Having dealt with this, if an NQF measure comes along, okay, then you are going to see Medicaid and you are going to see insurance companies say this is a national standard, and
that we believe that anybody with this
diagnosis should have this tool done for
reporting.

CO-CHAIR SUSMAN: You know to me
one of the salient questions, just to frame up
is, should we be tying measurement in this
area to a PHQ or is there a more general need
to measure remission? And it might not
necessarily have to be a PHQ. By doing a PHQ
you are narrowing the measurement focus, and
I would think also not endorsing, you are
recommending the use of a single tool. I
think the tool itself is great.

DR. PINCUS: You are setting
yourselves up so that you are between a rock
and a hard place. On the one hand if you want
to endorse something you have to have a
certain level of evidence and you are not
going to get the evidence if you have
something that is generic that you can't
capture the performance standards, especially
when you are talking about outcomes. So
ultimately if you want to meet this criteria
of having sufficient evidence and
documentation of the implementation it's going
to have to be a specific tool. If you are
leaving it up to whatever people want to use
as a rating system, it will never get the
evidence necessary.

CO-CHAIR SUSMAN: I certainly agree
with that, but I'm thinking why PHQ. I mean
one could choose a CSD, where there's plenty
of psychometric data about CSD.

DR. PINCUS: It has been proposed.
And there is to my mind there is more than
sufficient evidence to recommend it. If you
want to go to medication developers and say,
gee, why don't you modify your measures to use
any one of these six different options, you
could do that, but it'd raise a lot of
questions. Not all of them have been tested
the same way in the same populations and so
forth, and you wind up getting picky about all
these things.
It seems to me, I mean my own view is that this is an exemplary measure of what we are talking about, and it doesn't - I don't believe that NQF endorsing a measure requires that everybody does it. It's simply an option for insurers or even local clinics to say we want to measure --

DR. GOLDEN: I would be very cautious about that assumption, very cautious. Hey, I'm an old board member of NQF; I've been doing this for years. And I can tell you that an NQF endorsement of a measure would essentially say to a number of decision makers that this is considered to be an accepted national standard that we expect providers to adhere to.

DR. GOPLERUD: Let me suggest two analog situations. One is on the alcohol screening brief intervention CPT measure in which it specifies the use of a standardized instrument such as the AUDIT, the ASSIST or the DAST. So it says, for example, but it
basically puts the thumb on the weighting
scale so it uses these measures.

The second is, if you take a look at
say the diabetes NCQA measures, they don't
specify what blood pressure cuff you have to
use; they say you have to monitor blood
pressure. They don't say what lipid test you
use, what strip you use or what assay you use.
But they do specify what the number is. What
we could do with a measure like this is to
say, endorse it or other standardized metric
demonstrating 50 percent reduction or
something along those lines.

CO-CHAIR SUSMAN: Just for
clarification from the NQF staff, as I
understand our goal, our task if you will is
to deal with the measures before us. And that
we have been given the PHQ, and that is sort
of - and this is the measure where the
psychometrics have all been worked out on, and
to get to Harold's comments that really this
is what we have to deal with.
I agree, and that was sort of where I was coming from in my general remarks, but at the end of the day we have to deal with this single measure that is before us.

DR. PINCUS: I think the reality is, that if we throw out something that specific then we might as well go home, because there is nothing that is generic that will meet the criteria.

DR. KAUFER: I see this as a harmonization issue. I think we need to, if the data exists for something, I think just the wording can be softened to say that this is an example of an appropriate standard, and that certainly other candidate measures if they show evidence supporting that as an outcome, could be equally well qualified. But we have an instrument where we have the data in hand, I don't see any problem with moving forward.

DR. STREIM: I think the kind of statement you are talking about really has to
do with how these measures are viewed and used. And I think our job today is to endorse measures or not. I think if you have concerns about whether a measure by being endorsed will implicitly be regarded by legal entities and insurers as a national standard of care, I actually - first of all I don't think that is a bad thing necessarily, but I think in terms of what other measures could be used, I believe Medicare individuals, private insurers, and health care systems, are still quite free to use Hamilton depression rating scales, or other, Beck rating scales, to have - with defined parameters for remission, just as you could with a PHQ. But I think the question before us, as I understand it is, does this measure meet muster and I think that is all we have to answer. I appreciate what you were saying before about what the implications are. I think it's a good thing if we actually have a measure that is endorsed that looks at remission at 12 months as an
important outcome.

DR. GOLDEN: To follow up I would agree. I think that the measurement tool may be valid and useful, the measure that we might want to have is some sort of standardized way of assessing outcome. But the way the measure is written probably would not pass muster to be - because it really does define the method of how that assessment should be done.

DR. STREIM: But you have to have a measure to have a measure.

DR. GOLDEN: I understand. But having wandered through this world and forest, there will be many many entities that will say that this is the way to do it, just go do it. And you will then essentially create a standard of care.

CO-CHAIR LEDDY: I think that what the disagreement is if we endorse one scale that that becomes a standard of care for all care, that there may be other scales that could be endorsed as Eric said. So if we
endorse this scale, perhaps we could have Minnesota talk a little bit about this particular scale, than that doesn't mean that every single provider or insurer or government program has to require this scale be done.

DR. GOLDEN: Yes, but in the context of how this world is working, people are looking for measures. You now have an NQF endorsed measure of a scale, and there will be many entities that will take that measure and say, this is a simple - this is done - NQF endorses it. Everybody should do this. So you have locked into that scale. It's basically done.

CO-CHAIR SUSMAN: At the end of the day we are going to go through a process, we're making a global judgment about whether the world is a better place here and all the criteria are going to be met. And I can imagine we could come to a decision on the basis of the psychometrics and all the data presented here that the PHQ is a reasonable
tool, that there are good psychometrics, yada yada, and so the world would be a better place if we measured depression initially and measured that patients achieve remission. That's important, it has impact. Yada yada.

First Harold.

DR. PINCUS: A couple of points. I really appreciate that. One, if this was approved, it could potentially open the door for other groups to come in and say, okay, we've got a tool, we've got a tool, and that is not a bad thing.

Number two is, I think if everybody did PHQs I think that would be fine. I mean the comparison is, we don't have a measure that says, you must measure patient perceptions of care. But CAHPS is endorsed so everybody has to use it.

DR. GOLDEN: No, it's now a requirement. CAHPS is now a requirement. I mean if you want to be in Medicare --

DR. PINCUS: So what is the
problem? What is the problem with PHQ-9 being a requirement?

DR. GOLDEN: I'm just pointing out that you are endorsing a single scale that would become the standard of care.

DR. PINCUS: What is the problem?

DR. GOLDEN: Well, that's for the discussion.

DR. STREIM: It is true that we would be endorsing a single measure that has embedded in it a single tool that allows us to do the measurement, but it is not exclusive. I mean I think that is why this is okay to do. It's not saying - well, Harold has already said that other people can come forward and say there are other ways to measure remission. And all we are doing when we endorse is saying that we vetted this, we believe it has validity, it has utility, et cetera, and the results will be interpretable. That's all we're really saying. I understand your point that it may be pushing the field in a certain
direction to have -- the availability of an endorsed measure does move the field ahead.

DR. GOLDEN: But the measure says to use this tool, it doesn't say, a tool such as. If you are going to use the measure at all. Nobody is obligated to use this measure.

DR. PINCUS: I don't see where the issue you raise is embedded within the criteria.

MS. BOSSLEY: This I think can go down in the scientific acceptability discussion, and perhaps feasibility. But it's definitely there. So can I suggest because we kind of skipped, allow me to sort of give a little background of why they selected the survey, why it is measured the way it is. I mean I think they could try to give it for all three because it's pretty much the same thing. And then let's have an importance discussion, have you vote on that, and then move down through - because I think you are going to address these issues when you get into the
different criteria for this.

MS. PITZEN: My name is Collette Pitzen. I'm a staff member at Minnesota Community Measurement. And these measures were developed in concert with ICSI, around the Diamond Project improving depression care across Minnesota. A lot of the reasons why the tool was selected is that it does have validity and reliability. A lot of recent articles are coming out, even in the psychiatric community, that this can be used in a psychiatric setting. It's easy to administer and score, and the patients can understand it. And I just wanted to share that some of the discussion I'm hearing here is actually playing out in our state. For quite some time PHQ-9 has been used in the primary care setting, and not hearing a lot of gruff about that. But initially some of our behavioral health providers were expressing some of those same sentiments. I get emails that it is insulting for me as a psychiatrist
to use this tool. I just wanted to share some
comments of some replies to that. A
psychologist who is leading up this effort in
the male health systems, but I was not
completely on board at first too, I will have
to admit. However after using the tool for
many months I find it an essential part of my
work with depression. My two favorite stories
consist of, one, a patient who stated, I still
feel depressed, but after showing her trend in
history, PHQ-9 scores, she was able to track
her progress and recognized her treatment
gains.

And secondly, the patient who
endorsed suicidal ideation in the PHQ-9 but
denied it with primary care and then with me,
but opened up about it after going through the
PHQ-9. I hope this helps encourage use of
this measure. This was actually a suicidal
patient that she would have missed.

In having all these discussions,
it's interesting, it didn't come up, oh we
should use the HAM-D or we should use the Beck. It's like why - they didn't want to be measured. They weren't applying measurement on a routine basis. And I've seen a huge acceptance over the last year and a half, and many of our behavioral providers are coming on board. This is still a voluntary measure for a certain amount of time. Our state has endorsed this going forward though.

Any other questions?

DR. STREIM: So when you say it's endorsed but voluntary, that is saying that it is not required by the state for reimbursement purposes?

MS. MAYBERRY: It's just a matter of time. In 2011 the provider groups are going to have to all report this measure, and it will be used in a quality incentive program for the state. It's voluntary now in terms of there is a provider coalition in town that does have a payment for performance program built around this measure, as well as all of
the health plans in the state are moving
towards payment for performance on this
measure.

DR. GOLDBERG: This measure has
some momentum. Now the issue of NQF
endorsement I think there are so many people
looking to mandate outcomes measurements for
depression that if they look in the NQF book
and they find one that is endorsed by NQF that
is likely to push this momentum forward. And
it's up to the other measure people to get a
measure adopted and endorsed by NQF. I know
there are other measures out there. This
measure is pretty good; not great, it's got
problems. But it's pretty good. And there
are other measures that are just as good,
maybe better. But they didn't submit them to
us. So the people that didn't submit them, I
think if this becomes an endorsed measure, if
it's going to further the momentum of this
measure, we're going to see it even more
widely, because people will say, well, we're
looking for something. Wait a minute, here is one that is NQF endorsed, let's use that one. There may be no stopping them after that. It may become like the MMSE.

DR. HENNESSEY: Can I ask how did you arrive at this particular measure?

MS. MAYBERRY: You know I think it was that primary care is our initial audience, and this is a tool used in Minnesota widely in primary care.

DR. HENNESSEY: Thank you.

CO-CHAIR SUSMAN: As a primary care clinician and mental health researcher, this is widely disseminated. It has clear face validity to people; it takes a very quick time to administer; it's easy to incorporate if one is so inclined into one's routine. I mean we should go down and start considering the points. And let's get down to the business here, because I think we are really getting into some of the weeds that will come out as we go through the criteria.
DR. GOLDEN: Again, the question though as we go through this, and this is something for NQF staff, people say, okay, other people could come forward with a measure, the windows of opportunity for further measurement tools to come forward to be endorsed is fairly narrow. It's not like this is a continuous process. So as we go through this the question before us is endorsement of a standardized measurement process versus the endorsement of a standardized measurement tool specifically. And I think there is a nuance there, and I fully - as opposed to the issue of not being measured at all. And I just don't think it's that easy for the iterative process if suddenly, if we endorse one measure, to say, oh yeah, there are five or six other things that you could use as an acceptable alternative.

CO-CHAIR SUSMAN: But the reality is, we have the measure before us. This is
what was submitted, and I think we need to go through the process. I hear what you are saying, Bill, and I agree. On the other hand this is our task for the day.

CO-CHAIR LEDDY: So could we start with importance. And would anybody from the workgroup that reviewed this measure like to comment on this - on the importance issues, impact, gap, and relation to outcome?

DR. PINCUS: From my point of view this is clearly a major problem, for importance. Actually there is some data that if you look at people currently under treatment, using Medicare and Medicaid datasets and you do PHQs on them, a large proportion of them are still highly symptomatic and are not in remission. So that is - there is clearly a gap. It's embedded actually into a quality improvement process in terms of how Minnesota is doing it so that it is actually in the course of care that one does this, so it's not just a measure, it's
actually a tool for monitoring treatment. And
it's one of the best performing measures of
outcome. So it clearly meets the importance
criteria.

CO-CHAIR LEDDY: Any comments on
the importance of this measure?

DR. GOLDBERG: Bill's point,
though, the importance - are we voting on the
importance of - look at the title: depression
remission at 12 months. Measuring that is
important. That is what we are talking about.
And we are not even mentioning any particular
way of doing it. Just that it is important to
measure depression remission at 12 months.
That's it. I would say yes. It's very
important.

(Simultaneous speaking)

DR. GOLDBERG: That's what we are
voting on. We don't have to worry about how
to do it.

DR. PINCUS: And the STAR*D part
clearly endorses the fact that if people don't
achieve remission that there is subsequent significant problems in failure to achieve remission.

CO-CHAIR LEDDY: So the vote on importance.

DR. WINKLER: How many say yes?

(Show of hands)

DR. WINKLER: All right, does anybody say no? Or abstain?

CO-CHAIR LEDDY: Okay, so the next thing that we consider, and then vote on, is scientific acceptability. Of the measure properties. Now you are getting into numerator, denominator, exclusion, all of that.

DR. GOLDEN: I have a question for the developers. I believe - am I correct that the denominator includes MDD and dysthymia? That's a pretty diverse audience, so tell me about dysthymia being included with MDD.

MS. PITZEN: The decision was made early on that this was a population that their
care could be improved. We did exclude 311, depression not otherwise specified, from the denominator.

DR. GOLDEN: And is there—in terms of consistency of application in the coding, you have to code for this to be included, is that the deal?

MS. PITZEN: Correct.

CO-CHAIR LEDDY: Any other questions or comments?

Did the workgroup want to talk about your votes?

DR. PINCUS: One other question. So you are defining remission as coming below a threshold rather than 50 percent kind of thing. So that is one reason why it would apply to dysthymia as well. So it reduces the heterogeneity because it is below a threshold.

DR. GOLDBERG: Can you tell us how risk adjustment applies to this measure? That seems to be the one weakness.

MS. PITZEN: We actually convened a
workgroup, a technical advisory workgroup, that met March 22nd to start looking at the risk adjustment methods for these measures, and initially determined that we need to work on getting our response rates a little better. And I would speak more about the six-month measure. We have a good full set of data. We are getting ready to publicly report the 12-month data. But going forward with severity and risk adjustment we selected the severity at the initial PHQ-9 score to be used for risk adjustment in the future. We also did consider other comorbidities like diabetes, acute MI, double depression, chemical dependency, substance abuse, and those will be future considerations in our risk adjustment model.

DR. HENNESSEY: I have a question, are there any populations for which people are concerned this may not be a valid concern at this time?

MS. PITZEN: Pretty much as far as
the measure goes we are only including ICD-9
codes 296.2, 296.3 and dysthymia, 300.4. So
four that the instrument is valid in those
areas, and that the measurement is
appropriate.

DR. HENNESSEY: How about from a
demographic perspective, culture, gender, so
on?

MS. PITZEN: Going back to the risk
adjustment question, we did do some analysis
and literature search about the socioeconomic
impact. So for diabetes and vascular measures
we are risk adjusting based on insurance
product as a step towards that, but the
decision of the workgroup was that that was
not - that once patients who identified to
receive care that there were very little
difference based on type of product. The
differences were more in terms of access. So
for this measure that was kind of set aside as
a potential risk adjuster.

DR. GOPLERUD: Is it applied to
children?

MS. PITZEN: Eighteen and older.

DR. GOPLERUD: Is it available in other languages?

DR. PINCUS: And what about this in the geriatric population?

DR. STREIM: There is actually data on its performance and actually Deb Saliba at RAND, a group of people did a national validity field study using it for MDS 3.0 which has been adopted by Medicare, will be implemented next fall. So actually the PHQ will be used in all 16,000 nursing homes across the country.

DR. PINCUS: So that is a national standard?

DR. STREIM: It is. So the horse is already out of the barn.

DR. GOLDEN: I have no problem with it being a national standard as long as it is being accepted as a national standard. And that was my first question: who else has
endorsed the measure.

    CO-CHAIR LEDDY: So any other comments about scientific acceptability?

    DR. WINKLER: I just have one question to clarify, the denominator statement includes those with those diagnoses and a PHQ-9. What about patients who haven't had the PHQ-9 done? They wouldn't be included, right?

    DR. PINCUS: Right. That is captured in the third measure. You wouldn't be able to measure pre and post unless you had that.

    DR. STREIM: A question for NQF staff on endorsements when things like risk adjustment are still being developed. I understand the stewards are supposed to update these periodically, but at the point at which it is endorsed, at one cross-section in time, is it endorsed with caveats or explanations or comments regarding the lack of risk adjustment may limit the interpretation in certain
settings?

DR. WINKLER: The discussion around that appears in the report, but doesn't necessarily get tagged to the measure like in the database. However, I think there is a general understanding that measures have life cycles and they evolve and they need to evolve. So we do review them for maintenance review every three years, or on an ad hoc basis as needed if something changes or becomes dramatically obvious that it needs a sooner look.

DR. STREIM: If flaws are discovered in later validity testing, can a measure be un-endorsed or revoked?

DR. WINKLER: Yes, that would be the purpose of an ad hoc review, is if in use is usually where we are hearing the feedback is somebody has tried to do it and something - it did not work for any number of reasons, and they tell us about it, then we would do an ad hoc review to reevaluate that to determine
whether that needs to go away.

    MS. BOSSLEY:    I think the key
question for all of you is, do you feel
comfortable that this measure without risk
adjustment is appropriate to be put out for
public reporting right now. I think that is
your question, and that is what you all need
to grapple with.

    DR. PINCUS:   So, two questions.
One is, when a measure is endorsed, is it
endorsed with instructions to do risk
adjustment, or is it endorsed with
instructions saying, here is one way of doing
- what is the relationship of the endorsement
to the risk adjustment procedure?

    DR. WINKLER:   Well, the endorsement
is the measure as specified as it was
submitted. Now in the course of time until
the next maintenance review on an endorsed
measure there may be annual updates. Measure
developers have different schedules. It may
be every six months, who knows.
DR. PINCUS: If we endorse it and there is a - not just for this one, for any of them - and there is a statement in here about risk adjustment, but as I read it it's more like it's advisory than it is this measure requires it. So what is the meaning of that in terms of endorsement? Are we endorsing the measure with the associated risk adjustment procedure? Or are we endorsing the measure with the option of a risk adjustment procedure?

MS. BOSSLEY: This measure before you, you would be endorsing without any risk adjustment because there is no model include and there is no specification, they haven't tested it, so they are in the process of doing that now. So this is where it gets fun again. There are three criteria right now for time limited; this was just approved by the board in December. It needs to be - there is no other measure within the NQF portfolio which I think there isn't a measure within
addressing this. There needs to be a need for it, so either a legislative mandate, that type of thing, I think that one we'd have to think through.

The last one though is that the measure isn't complex, and it doesn't require risk adjustment, isn't an outcome measure, isn't a composite measure, and that's where I think this is hard to apply time limited to because it is an outcome measure and it is complex and you are talking about risk adjustment. So I think that is where it gets a little difficult to say within one year you need to come back to us and tell us whether or not it should have been risk adjusted.

DR. PINCUS: If we wait until all the risk adjustment issues are solved for these measures, we - it's going to be three years.

DR. WINKLER: At this point, the measure you are evaluating is not risk adjusted. It is looking like they are
considering it and thinking about it, and
maybe another iteration in a couple of years
will be modified and we can look at that at
that point in time. But today's issue is the
way it is.

DR. GOLDBERG: Any competent user
group, if they are going to use this as a
comparison across settings, is going to bring
up risk adjustment immediately.

MS. JAFFE: And actually we are
involved in a similar project in Washington,
and the risk adjustment issue is a big problem
right now. It's not the use of the tool so
much as determining what the score should be.

DR. GOLDEN: To follow up on I
guess with the developers, how is this
performed with comorbidities such as stroke or
heart attack or substance abuse? Has that
been an issue?

MS. PITZEN: The comorbidities were
considered by the group looking at risk
adjustment, and they will consider them in the
future, not for the first go round.

DR. GOLDEN: No, I guess my question was, is there any track record in notes or what have you about do those comorbidities affect the response rate over time and the score?

MS. PITZEN: I do have some literature that talks about that, Unutzer and Katon. So yes, and that has been discussed within our workgroups. Right now we are not excluding patients based on risk comorbidities; they are included.

DR. GOLDEN: But it does - does it affect the score over time?

MS. PITZEN: I think that it can. I guess I don't have any hard evidence to give you today.

CO-CHAIR SUSMAN: I think the group that Wayne Katon and the group up in Washington has done, there's been a lot around comorbidities, and they have used tools like PHQ and patients with diabetes, asthma, and
multiple comorbidities, and the importance of monitoring to remission and the use of the PHQ in doing so has been pretty well validated.

DR. PINCUS: Yes, the fact that it is threshold kind of - so it reduces that issue. I mean the fact that there is fairly good evidence that the threshold as suggested is - failure to achieve that is associated with negative outcomes.

CO-CHAIR LEDDY: Robert is next.

DR. ROCA: This may be a usability question, but I'm wondering how one handles the fact that over the course of 12 months somebody is likely to have passed through the hands of several caregivers, especially if the initial ascertainment occurs in an in-patient setting. How is it determined whose care is being evaluated over the course of 12 months? How is that being handled?

MS. PITZEN: I can answer that, and it's kind of a technical question. Groups submit data to us, actually at a visit level
detail. So every contact that the patient has gets submitted as a record with their clinic. And we are attributing it to the location where the patient first met that diagnostic criteria. But then all of the information within that medical group then comes forward for that patient, so we have all the scores and can see their history.

DR. ROCA: So for instance if someone is in the hospital and the hospital is reporting PHQ scores, somebody may very well have a very high PHQ score at that point because they are in the hospital. Twelve months later they may or may not have stayed in treatment with who knows which provider down the road. Is the hospital then responsible for that outcome?

MS. PITZEN: I can answer that question for you. It's an ambulatory care based measure, so the identification of patients is starting in the ambulatory care center. However we do have some systems who
have an integrated hospital and clinical record, and they are submitting those patients' PHQ-9 scores as well, but we are not going after inpatients with depression.

DR. ROCA: So what we are looking at here is an ambulatory process?

MS. PITZEN: Correct.

DR. STREIM: So the score from the index episode would be whatever is available from the current provider.

MS. PITZEN: That is correct.

DR. WINKLER: George had a question?

DR. WAN: Just a general observation. When looking at Minnesota's submission they actually summarized their results of 17,000 patients with data from 123 clinics. And I was amazed to see the average scores. So they had the scores of 4.6 percent from a population based level. So that seems to me very low in this setting, so then the question would be, I understand from an
assessment point of view this will help, once you assess that and identify that gap, you want to have target improvement interventional programs to achieve a much higher rate. But I'm just very surprised to see that very low rate.

MS. PITZEN: Can I make a comment on that? I think I mentioned earlier that part of our problem in this initial go round is that groups are not getting that six-month PHQ-9 score or that 12-month PHQ-9 score as much as we'd like. In the Diamond project they are hitting that compliance rate at about 60 percent. In the full population, general public, usual care, we are at about 20 percent.

If I look at just the patients that we do have a PHQ-9 score on, and I will have to give you six-month data, we are at about 24.6 percent are achieving remission. But we don't want to promote - set that forward, because that is usual care. Or we are only
going to measure the patients that we can contact. That is not going to change.

DR. PINCUS: It's not so surprising if you look at the existing process measures, you know, depending on which measures you look at, they are in the sort of 20 to 45 percent range. And this is outcomes which are much harder to achieve.

CO-CHAIR LEDDY: So are we ready to vote on usability for this measure? Sorry, scientific acceptability.

DR. WINKLER: The voting for this one is along the same categories of completely meeting, partially meeting, minimally or not at all.

So how many of you think that the measure specs and information meets all the criteria completely?

(Show of hands)

DR. WINKLER: Partially?

(Show of hands)

DR. WINKLER: Okay, how many
minimally. That's a zero.

DR. GOLDEN: I am partial.

DR. WINKLER: Got you as partial.

(Off the record comments)

MR. CORBRIDGE: So 18 partial, is the denominator.

DR. GOLDEN: I am partial, but I would like to make a comment about the
reliability just for your own notes. If you took 300 of these patients and you put them
trough Clinic A and you put them through Clinic B, the ones that take the test would
probably have similar results. However, Clinic A and Clinic B may code grossly
differently, so you may have very different numbers of patients receiving the test, so
there is a reliability issue about coding, and entry into the assessment process.

DR. PINCUS: To the extent to which they use - not everyone is specified.

DR. GOLDEN: Or the fact that many primary care practices don't code depression
or dysthymia. I certainly don't, because it's a payment problem, and it's also a stigma problem. So it's a coding avoidance issue.

CO-CHAIR LEDDY: Okay, with that comment we are ready to move onto a discussion of usability of the measure. So that includes is it understandable, harmonization issue, does it add added value. Would anybody from the workgroup like to comment on their votes on that or how they found the measure?

Any discussion or questions?

DR. PINCUS: Yes, I have not sure what partial means with regard to harmonization.

DR. WINKLER: The ratings are, does it meet the criteria as laid out in your measure evaluation criteria. So it completely meets them all, partially meets them all, minimally meets them all, that kind of spread out scale. So harmonization I think, I think in this particular case the harmonization that might be applicable would be the capturing of
the patients with depression compared to other measures of depression.

CO-CHAIR LEDDY: Or the various settings it's used in. The definition of harmonization says, could this measure be used not just in an outpatient setting but also inpatient or nursing home.

DR. PINCUS: It's not just could be used, no, it's a question of whether it's related to measures that are already endorsed by NQF in other settings.

CO-CHAIR LEDDY: Yes, yes, you are absolutely right.

DR. PINCUS: So looking at other depression measures at NQF they utilize very similar criteria.

CO-CHAIR SUSMAN: There is an effective continuation phase measure that we've come up with.

DR. PINCUS: I think that - I didn't look at the specific details, but my sense was, they were well harmonized.
Somebody may want to look at the specifics of that. But that's why I didn't understand the "partial" in harmonization.

CO-CHAIR SUSMAN: I think the longest is the six-month continuation phase. But this is getting at longer term remission.

DR. PINCUS: The inclusionary suite and criteria seems pretty similar.

CO-CHAIR LEDDY: Any other discussion on usability? Or are we ready to entertain a vote?

DR. WINKLER: Okay, so who all believes it meets the usability criteria completely?

(Show of hands)

DR. WINKLER: Seven.

Partially?

(Show of hands)

DR. WINKLER: Okay, nine.

Minimal?

One.

Not at all?
Thank you.

MS. BOSSLEY: We are missing one?

DR. WINKLER: Luc is out.

MS. BOSSLEY: That's it.

DR. WINKLER: Okay, a flexible denominator.

CO-CHAIR LEDDY: Okay,

feasibility.

DR. GOLDEN: What is the status of this as an electronic tool to query. Is it a single score? I haven't used it.

CO-CHAIR SUSMAN: Yes. A lot of PHRs now bake it in.

DR. GOLDEN: So it'd be sort of like putting in the cardiac - the New York State, New York Heart Association risk for heart failure.

CO-CHAIR LEDDY: Yep.

DR. GOLDBERG: Sort of a widespread ad hoc option of this says something about its feasibility. People are finding it feasible.

DR. PINCUS: I had a question about
the criterion of data generally is a byproduct of care processes. Is it - what do you mean by that? Is it a byproduct of how care should be, or how they are?

DR. WINKLER: Let's put it this way. I can tell you what we meant it isn't - we do not mean - and that is where someone has to go in and abstract the blood pressure recording from a chart in order to generate the data to go do the performance measure. So in this case the fact that you were doing the PHQ-9 as part of the care of the patient and it's in your records, if it's in your electronic records so that the end result number is readily extractable electronically -

DR. PINCUS: Is that a separate criterion, electronic source? I thought just in terms of 4(a) it - the - that it is sort of a byproduct in the sense that if you are providing care irrespective of where it is located that you are doing it. So that if you
are doing blood pressures and typically
reporting it is a byproduct of care, then it
would be there whether it was electronic or
not. But it seems to me if you are treating
somebody with depression and you are
monitoring their response to treatment, this
would be a natural byproduct of care.

DR. WINKLER: Correct.

DR. ROCA: It might be, but you may
not do that scale routinely, though. Wouldn't
you have to do the scale routinely in your
regular practice?

(Simultaneous speaking)

DR. GOLDEN: I think the question
is, as currently constituted. You can provide
these - a glucose measure is a byproduct of
care. If he came for depression this score
may or may not be a byproduct of care at this
point in time.

DR. PINCUS: I am not sure I
understand the distinction, what makes glucose
a byproduct of care as a pressure measure or
not.

DR. GOLDEN: Only because not everyone's doing it.

DR. PINCUS: But there should be.

DR. GOLDEN: But that's the point.

DR. ROCA: But it is certainly not a standard of care. To use the scale. A lot of us would say it should be but it isn't. So in that case --

DR. PINCUS: That is why I was kind of getting at the sense of, what do you mean when you designed this thing as a byproduct of care?

MS. BOSSLEY: The goal is that you are not putting forward a measure that requires this additional data collection or going to somewhere else --

DR. PINCUS: That is irrelevant to the care you are providing.

MS. BOSSLEY: Right, so we are asking you to rate just this measure.

DR. PINCUS: Clearly this measure
is not irrelevant to the care being provided. It ought to constitute a key feature of your decision making with regard to the care you are providing.

MS. BOSSLEY: The goal is to not have any measures out that require a huge additional piece of data unless it is absolutely critical. I don't know that this measure is a good example of that.

(Simultaneous speaking)

MS. BOSSLEY: We are trying to look at the burden of data collection and the feasibility -

CO-CHAIR SUSMAN: If this took an hour to administer it'd be a very high burden. In point of fact it's much much shorter.

DR. GOPLERUD: Two pieces. One is that there are CPT II codes that could be used for this, so it's built in and those were adopted two years ago. The other is that it's baked into the VA/DoD electronic medical record, and it's the PHQ-9.
CO-CHAIR LEDDY: So are we ready to vote on -- oh no, Robert, I'm sorry.

DR. ROCA: I was just going to say that I completely agree this is a reasonable thing to do and feasible, and what we ought to be doing, and it really depends on what this criterion means. Because clearly if you are treating diabetes there is nobody who treats diabetes without getting a glucose clearly. But there are - most clinicians I dare say treat depression without using a scale, so it is going to be something extra to do, and I can agree that it ought to be done, that we ought to be doing it, it ought to be the standard, but it would require something additional for clinicians than they are already routinely doing. And I thought that was Harold's question, but maybe it wasn't.

DR. PINCUS: Well, my question was, what do you mean by a byproduct of care. It seems to me like I said it's certainly not something that is irrelevant to the decision
making process of the clinician providing care. It's very relevant to that. So in that sense it is a byproduct of care.

DR. WINKLER: And certainly in its most simple form, this is about burden of data collection to do the measure.

DR. PINCUS: It is burden versus benefit too, or critical benefit, not performance measurement benefit.

CO-CHAIR LEDDY: Are we ready to vote on feasibility?

DR. WINKLER: How many think it meets the feasibility criteria completely?

(Show of hands)

Fourteen is what I get.

How about partially?

(Show of hands)

Four.

DR. WINKLER: All right. So then the final vote of the day.

(Simultaneous speaking)

DR. WINKLER: Is to recommend that
it go forward for endorsement or not.

CO-CHAIR LEDDY: So it is just a yes or no. This is a yes or no question.

MR. CORBRIDGE: Before we do that, we would like to open it up for public comment. Are the lines open? Anyone on the line want to comment?

(Telephone dialing)

DR. GOPLERUD: We are voting on the 12-month measure. We will do this again for the six-month measure?

CO-CHAIR LEDDY: Correct.

(Off the record comments)

CO-CHAIR LEDDY: Did you have a comment?

MS. GALBREATH: I just wanted to say, at the national council we do a lot of work in terms of working with primary care on PHQ-9 and doing this kind of screening measuring, so we are very supportive of this measure. We think there are questions regarding implementation in terms of primary
care versus community and nursing home
patients to some of the things that are down
the road, but we are very supportive of this
measure.

DR. HENNESSEY: What do you see as
the challenges for -

DR. WINKLER: Can you use the
microphone?

DR. HENNESSEY: Oh, sorry about
that. Where is a microphone?

What do you see as the major
challenge for community mental health centers
moving forward?

MS. GALBREATH: We are working,
there are community mental health centers that
are working to use the PHQ-9 as a tool as was
explained in terms of using that as a
beginning place for further assessment. But
I think the cultural shifts for the
professionals, the time, data, how they list
PHQ-9 in an electronic medical records, if
centers are at that point. So some of the key
issues in terms of primary care in terms of measurement and it means a piece of the puzzle to start the conversation.

DR. PHILLIPS: I also have a question. So the PHQ-9 I understand, but the concept of remission at 12 months, could you comment on that?

MS. GALBREATH: I have more of a policy background than clinical. But I imagine a lot of our centers are doing measurement of best practice. I'm not really sure in terms of the measure. I think that that would be supportive of that.

CO-CHAIR LEDDY: All right, are we ready to vote?

DR. WINKLER: How many vote to recommend this measure?

(Show of hands)


CO-CHAIR LEDDY: Now, hopefully
the next measure will be a little speedier, because it is at least similar. So Ian, are you going to introduce the next measure?

MEASURE OT3-012: DEPRESSION REMISSION AT SIX MONTHS

MR. CORBRIDGE: So we are on measure #12, entitled depression remission at six months. So still from Minnesota Community Measurement.

Just a brief description of the measure. Once again, adult patients aged 18 or older, major depression or dysthymia. Initial PHQ-9 score greater than nine, who demonstrate remission at six months defined as a PHQ-9 score of less than five.

This measure applies to both patients with newly diagnosed and existing depression whose current PHQ-9 score indicates a need for treatment. The patient health questionnaire is a tool widely accepted, just once again similar constructs as the last measure that we read over, and once again the
numerator and denominator statement are the same from the last measure that we discussed.

DR. STREIM: Actually I just realized, the word current PHQ score, in that second sentence, implies current at what point in time?

MS. PITZEN: If I can address that, it's the process where — I mean you have a starting point for measurement collection, and it's the initial, the first PHQ-9 score that is coming in that also we have the confirming diagnosis that they do have major depression or dysthymia.

So it's not newly diagnosed, the very first PHQ-9 ever given. It's the very first PHQ-9 when you are starting your measurement process, going forward from that.

This is a longitudinal measure, so patients can come into the population whenever they are identified, so it's not like in this last year, it's not like a snapshot; it's whenever they are meeting the criteria for
that measurement then they come into the population.

DR. GOLDBERG: What is the last one, other than six for 12.

MS. PITZEN: There is absolutely no difference technically, population, and anything. The only difference is six months and 12 months. We have a lot of data on six months, and we'll actually be publishing 12-month data in June of this year.

DR. GOLDBERG: So there is more data on this?

MS. PITZEN: There is more data.

CO-CHAIR LEDDY: Eric.

DR. GOPLERUD: This may come in the area of harmonization, but the NCQA measures are typically a measurement within the year, and so six months is an unusual length of time.

DR. WINKLER: Eric, just one thing. When you talk about a measurement year like NCQA uses, they are talking about their data
collection stance at that time. The actual specification of a measure may have other time frames, because that is what the clinical situation asks for.

   DR. GOPLERUD: But if you look at their asthma measures, their diabetes measures, they - continuous care - has there been a measure within a one-year interval after --

   DR. WINKLER: Right, but most of those measures are usually a point in time, something happened, yes or no, within the measurement year, as opposed to here we've got a change measure, and so the timeframe of change is more about the measure than the measurement program. You can put whatever parameters you want to your window of data capture. So I think that is where there is a difference.

   DR. GOLDBERG: Why six instead of four, five or seven.

   MS. JAFFE: I also asked why six,
and if you are doing 12 why also six.

MR. CORBRIDGE: I guess I would say the reverse of that. There was strong evidence in studies in the literature that six months was one of the cut points for measurement and also 12 months. If I talk about the importance of the two, the six month measure is where the most of our efforts are being focused.

CO-CHAIR SUSMAN: I think you tie it back to the data about length of continuation phase treatment, and the data are not precise that it's exactly six or eight or five or nine. You can take that cut where you want, but it's a very reasonable cut based on best evidence.

DR. STREIM: I think one thing to consider is that there is an emerging literature on stepped care for people who don't respond to the first line treatment. And if you look at time, expected time to improvement or remission, response or
remission, I'm not sure we have really good
studies of that for stepped care. Even the
impact study and the prospect study didn't
really look at it that way.

So I think that in some ways it's
almost arbitrary to say let's take a look at
six months because it's an awfully long time
to be suffering but I'm not sure we have a
scientific rationale in terms of time to
improvement, sort of as a survival analysis,
that would guide us to what is a reasonable
time interval for expecting remission.

MS. PITZEN: Initially with a lot
of our providers they were like, well why
don't you take an earlier score? And I think
oftentimes that that is just not enough to say
that that patient is better or in remission if
you are going to take a score at one month or
three months.

DR. STREIM: I think given that
treatment studies have clearly shown that
people can continue to improve on treatment up
to 12 weeks, as probably less than three months wouldn't make sense.

CO-CHAIR SUSMAN: And there are other NQF-endorsed measures that look at the 12 week milestone. So I look at this as a family of measures that we are trying to develop for the use of improving care of depression. And there is a certain arbitrariness here, and there are patients who are going to fall out and will need further steps here perhaps to get to remission. But given where we are and the state of the art, I think overall this makes a lot of sense.

DR. WINKLER: One comment, I would ask you, they are very similar measures; the timeframe is different. Do we need both measures, or conversely, if you want to see these measures widely used, should you expect to use both measures?

CO-CHAIR SUSMAN: I guess I see this as not making that decision for people but giving people a set of options where there
is sufficient rigor, where there is sufficient importance and so on. And some organizations might choose to focus on initial 12 week of therapy and choose the NQF measure in that family. Others might choose six months because of issues of tracking and getting patients back into the longer course of therapy, while others are really going to be pushing for full-year follow up. So I don't see this as an either/or or in some way specifying. I see it giving more tools to the field to help improve care. That is my own personal belief on it.

DR. PHILLIPS: I think also if I were a provider, the shorter timeframe I would want because it's more likelihood I'm still seeing this person, whereas at a year who knows. They could have gone through three other centers by that time.

DR. WINKLER: Is there any information about the lack of follow up for 12 months versus six months in terms of what
experience you've had with the measure?

    MS. PITZEN: It's about the same.

About 20 percent in achieving that follow up
PHQ-9 score at 12 months, and the remission
rates are similar as well; a little bit
better.

    CO-CHAIR LEDDY: So are we ready
to look at importance and we've just had a
pretty long discussion about really the
importance of this measure at six months
measurement. And the scores of the group were
pretty consistent. So --

    DR. GOLDEN: Just to comment. It
would seem to me in doing comparison is the
six month measure more important than the 12
month measure, and I could argue the answer is
yes.

    MS. BOSSLEY: Maybe the best thing
to do is to vote. You've got three measures
to discuss. You've got another one coming up.
And then go back and revisit.

    DR. PINCUS: Are we supposed to be
there, or are each one standing on its own?
Or is it a nested thing?

DR. WINKLER: It's a two-step kind of thing. Each measure needs to be evaluated on its own, but at the end of the day you want to look at your group and say, does this make sense as a group?

CO-CHAIR LEDDY: Okay, so let's vote on importance. How many people think it meets the completely definition for importance?

(Show of hands)

DR. WINKLER: Are three any nos?

CO-CHAIR LEDDY: You mean any not at all?

DR. WINKLER: Let's go back so we're consistent straight across.

DR. WINKLER: Completely, going back to completely?

(Show of hands)

DR. WINKLER: Any not at all's, minimally or partially? Oh, we have 18 people. Okay. Scientific acceptability?
Completely, I see none.


Usability, completely? One, two, three, four, five, six, seven, eight.

Partial? One, two, three, four, five, six, seven, eight, nine.

And is there someone with a minimal amount - okay.

Feasible, completely? Twelve.

Partials. Six. That's it.

And to recommend the measure or not.

CO-CHAIR LEDDY: So now to recommend or not recommend. So all that would recommend this measure?

(Show of hands)

DR. WINKLER: That's seventeen.


We are ready to move on to the fourth in this group which is also submitted by Minnesota, right?
MEASURE OT3-022: DEPRESSION UTILIZATION OF THE PHQ-9 TOOL

MR. CORBRIDGE: Correct, yes, so we will be moving on to Measure #22, as Trish indicated, also submitted by Minnesota. The title of the measure is, Depression Utilization of the PHQ-9 Tool.

All right, so just a brief description of the measure, very much similar to a degree with what we have been talking about. Adult patients aged 18 or older with a diagnosis of major depression or dysthymia. ICD-9 - go over the ICD-9 codes who have PHQ-9 tools administered at least once during a four-month measurement period. The patient PHQ-9 tool is widely accepted, which we have gone over.

A little bit further down, the process measure is related to they outcome measure of depression remission at six months and depression remission at 12 months. This measure was selected by stakeholders for
public reporting to promote the implementation
of processes within a provider's office to
ensure that the patient is being assessed on
a routine basis with a standardized tool that
supports the outcome measure for depression.

Looking at the numerator statement
for the measure, would be adult patients aged
18 and older with a diagnosis of major
depression or dysthymia. They provide the
codes who have a PHQ-9 score administered at
least once during the four-month measurement
period. The denominator statement reads as
follows: Adult patients aged 18 and older with
a diagnosis of major depression or dysthymia
and they provide the codes there.

So that is just a brief overview of
the measure.

CO-CHAIR SUSMAN: Just to clarify,
it could be just one initial measurement with
the PHQ? This does not imply response,
remission, is that correct?

MS. PITZEN: Yes, correct. It's a
process measure, and it's applied to a whole population with that diagnosis. It doesn't matter what their PHQ-9 score is. Did the patient have administered at least one time in the last four months, and there is the implication that they were in for a visit in that timeframe, did they have a PHQ-9 test administered or not?

DR. GOLDBERG: If you were following at six and 12 months, you had to have a measure at the beginning?

DR. WINKLER: If you didn't have the test done you weren't captured in the measure.

CO-CHAIR SUSMAN: That was the entrance criteria.

DR. HENNESSEY: So this is, as I understand this then, this would be a uniform administration of the test regardless of the presenting problem to the PCP's office?

MS. PITZEN: Let me clarify: it's for patients that have major depression or dysthymia.
DR. GOLDBERG: Yes, it's not screening.

DR. WINKLER: Bill.

DR. GOLDEN: A question on the operation of this measure. You have a patient being seen by a psychiatrist for major depression and managing the depression. The patient sees a PCP for their urinary tract infection or their bronchitis. The question is, it's not necessarily coded for the visit. Is there an expectation that the PCP administers this? Because the patient does carry a diagnosis of depression? Or does that have to be coded at the visit?

MS. PITZEN: It has to be coded at the visit, but it is related to that patient. So if that patient is being seen in primary care for a variety of reasons and they also have ICD-9 codes that support the depression diagnosis, the expectation is that they have a PHQ-9 also.

DR. GOLDEN: But if the depression
codes are in a separate office with the
psychiatrist as opposed to the primary care
office.

MS. PITZEN: I can answer that.

Technically we only have the ability to
capture information at the level of the
medical group, and when I talk medical group
that can be a broad health care system that
has a common patient identifier. Even a
chart, we have some clinics that have paper-
based charts that are participating. But you
can't know what you don't know. So in a
separate psychiatry office seeing that patient
we don't have a way to put that data together.

DR. GOLDEN: So you would expect the
psychiatrist to report but not the office that
didn't code?

MS. PITZEN: No, if both of those
offices are coding major depression for that
patient I would expect them both.

DR. GOLDEN: I understand, but if
only one is reporting major depression and the
other one is not, you would be expecting the one who's reporting it.

MS. PITZEN: Correct.

MS. JAFFE: I have a question about the scope of this one. It sounds like a process as opposed to an outcome, and maybe we need to talk about that first?

DR. STREIM: Agreed. I had the same determination on first pass. So can you suggest any way in which this might be construed as an outcome measure, indirectly related to measuring outcomes?

MS. PITZEN: Part of the reason why we put this measure forward, our groups initially were publicly reporting the six month remission measure, and our first data results, of course, were dismal, and a decision was made immediately that we also need to - we also have a set of 10 measures that we need to get this out in a transparent way because it is going to lead us to our outcome.
Currently, groups are at about 70 percent overall for many thousands of patients for having at least one PHQ-9. We still have a ways to go to get our six month and 12 month response.

CO-CHAIR LEDDY: So this sounds like clearly a process measure.

DR. MANTON: The other thing, if I'm correct, it's post-diagnosis. So the person would already have had to have been diagnosed. So in some ways it's measuring - I'm assuming that if they had been diagnosed they're being treated. So it's in some ways a kind of - kind of quantifying that, too, in terms of where they are. And it also is getting back to the earlier discussion, it really is pushing that particular tool, as opposed to others.

DR. WINKLER: In terms of your question on process outcome, one of the reasons I asked the question about how patients who didn't have a PHQ-9 done were
handled in the remission measures, there are
several different approaches to measurement
for dealing with getting the whole thing
started in the first place.

One of the things you could do is
pair this with one of your outcome measures.
To make - so that you've got the process
measure that says, yeah, you do it, and we'll
figure out to get a number on the
participation - or the use of the tool is, and
then you pair it with the measure that is the
remission measure, which is the true outcome
measure. But the two go hand in hand.

It's tied to it, exactly. You can -
one of your recommendations could be to tie
the two together, which would sort of take
care of your scope issue if you'd like. You
can tie all three.

MS. BOSSLEY: And what that means
is, anytime anyone went to use one, they
actually need to use all three and publicly
report all three measures together. We're
throwing it out there.

(Laughter)

DR. PINCUS: My view, while on the face of it, it would seem sort of by itself out of scope as a process measure, the reality is, we've sort of enlarged the domain slightly when we put out the call, and looking across many of the other measures that are submitted that are process-like, this is actually one of the better ones. And so I would come down on the side of including it, because I think it's actually typical. At least it allows people to have a way to demonstrate that they are actually looking at outcomes.

DR. HENNESSEY: I have a question for clarification. We talked about six months, we've talked about 12 months. Now I see here they are talking about administering it at least once during the four month measurement period. That seems a little out of synch. Am I missing something here?

MS. PITZEN: I can try to answer
that. It is a little bit arbitrary. We are having groups submit to us three times a year in four-month segments. And part of the questions, as they submit their outcome, their denominator data to us is, how many patients are you seeing in your clinic? How many have the diagnosis of major depression or dysthymia? And how many of those patients received the PHQ-9? It is a counting-type measure. The four months his just how we happen to have it.

DR. HENNESSEY: So the denominator is the patients seen in that four month period --

MS. PITZEN: Correct.

DR. HENNESSEY: So it could be any increment?

MS. PITZEN: Right. We had a historical catch up period of actually three quarters, and it's very easy to achieve on PHQ-9 in three quarters. So the time frame is a little bit arbitrary. If the group said oh
we are going to look in 12 months did you receive a PHQ-9 your rates are probably going to be much higher.

DR. GOLDBERG: There are a couple of issues here. Now I hear you say it's a counting measure, I'm more concerned about not including it in the scope. You could start counting a lot of things. But I am concerned about the other two that we voted on, yes, that unless we link the other two with some initial measure, the other two are going to be a problem.

CO-CHAIR SUSMAN: But the other two do have an initial PHQ embedded in and then measuring their effort, is that correct?

MS. BOSSLEY: It is correct, but what you will not capture, the ones who do not have a PHQ-9. It won't capture those patients in the other two measures.

CO-CHAIR SUSMAN: I mean, you know if you look at the existing NQF measure on acute phase or practitioner contacts, it's usually
your typical 12 week, number of visits.

Ideally you would tie the PHQ within that period and you'd have some harmonization here that makes sense from a process point of view. As it stands now, as a simple counting measure, I agree with Harold, and I'm okay with including this. It, in many ways, is not at all an outcome measure per se.

CO-CHAIR LEDDY: Isn't it informing you of the validity of the denominator of the other two measures, so it is linked.

MS. PITZEN: We did start publicly reporting this information, and the groups that are at 20 percent or below, they aren't very happy, because they know that their efforts to embed this process in their care haven't been too successful so far.

CO-CHAIR SUSMAN: Well, if you are going to take this a step further you don't have any idea about all the patients that were not recognized and therefore did not have a
PHQ, so it just depends on how far --

   DR. PINCUS: I think standing on itself, it is one of the better process measures, one of the better process measures. And it's one of the better process measures that have been submitted to us as a quasi-outcome measures. It certainly is justified in terms of being linked to the other two measures, although I think we should make sure we separate them. Because the others let you know who they didn't — how many people they didn't get to.

   DR. WINKLER: Bill.

   DR. GOLDEN: I am confused. To me, this measure becomes irrelevant with the other two being passed.

   MS. BOSSLEY: The only way that we can do it, which is why NTQA does it, anytime their PHQ score does not exist, it counts against them in the remission measures. That is the only other, I think, way you could do it without this measure. And capture
everything. So if you didn't have a PHQ-9 score it would be the same as using diabetes as an example or if they had an A1C test done but you didn't have the level, that counts against them in meeting the performance of that measure.

DR. GOLDEN: I just assumed that would be --

MS. BOSSLEY: But that is not the case here, correct?

DR. WINKLER: Yes, I mean one of the ways to get around the remission measures is to never do a PHQ-9. And that's what this measure is trying to --

CO-CHAIR SUSMAN: If you don't diagnose depression you don't have to do any of this. Right.

MS. BOSSLEY: So you guys have a couple options. You can always request, develop, or consider some changes to the measures and have conditions on the recommendations asking for that type of change
on remission. Or you can accept this as paired with the --

DR. GOLDEN: That weakens the integrity of the other two measures, but that's all right.

MS. BOSSLEY: You have a few options before you,

CO-CHAIR SUSMAN: I guess I don't see why we would link this to the other two measures, since the other two measures embed an initial PHQ in there. I see this going after a different population, a different set of issues, and I basically agree with what Harold has been saying, but I think it's clearly process right now.

DR. MANTON: The other two have a PHQ score that people would be entered into. This has none. So is the assumption that if their PHQ score wasn't nine, it was less than nine, that they wouldn't be part of this follow up, the six month/12 month?

MS. PITZEN: Correct.
DR. STREIM: I would argue that linking these is essential because it goes to the issue of usability of the outcome measures. We have enough problems with lacking risk adjustment, but at least if you can look at the measures, the outcome measures we just endorsed, and make a determination about the denominator, and whether you are actually getting at a substantial part of the population with depression or you are missing most of them. This will allow you to interpret what you have captured in your outcomes measures, and I think that it really is anything we can do to help improve interpretability of a publicly reported measure is a good thing.

CO-CHAIR SUSMAN: But Joel, I am not following. If I understand this, if you are going to rely on the initial measures, doing an initial PHQ and then a follow up to demonstrate remission - pardon me?

DR. PINCUS: It doesn't require that.
CO-CHAIR SUSMAN: Yes, it does.
The last two measures did.

(Simultaneous speaking)

DR. PINCUS: Again, you get into the denominator by having had it. It's not based upon the initial score; is that correct?

DR. GOLDEN: Right. All you have to do is look at put that at the end of the six months.

DR. STREIM: No. You have to have a PHQ to be in that denominator.

So for this measure that is under consideration before us right now, you don't have to have a PHQ to be in the denominator. All you need is an ICD-9 diagnosis of depression. So it's a wider - it's potentially a larger denominator, and what this really tells you is, if only 20 percent of people are getting a PHQ, then when you look at your other out comes, the true outcome measures, you are really only capturing 20 percent of people who have an ICD-9 diagnosis,
and that is still not the whole universe of
depressed people, but it's getting at a larger
denominator.

DR. PINCUS: If you are looking for
people who are currently depressed by PHQ
measure, the best place to look for that is
people who have a current diagnosis of
depression by ICD-9 diagnosis.

MS. JAFFE: So I am a little
confused. This standing by itself, not linked
to the other outcomes, just the fact that you
are just collecting this information once
every four months; that's all that's required,
right? So I guess I'm a little bit of: so
what, I mean if it stands by itself.

CO-CHAIR SUSMAN: I see this as a
process improvement measure. It's to get the
adoption of PHQ out into user care in
evaluating patients with a diagnosis of
depression, and I think it's wonderful for
that reason. But I still don't see this as an
outcome measure, and I still - I mean I
1 understand what you are saying.

   DR. STREIM: It's not an outcome measure, but it helps you incorporate the other outcome measures, and so it becomes an important part of the toolkit where the end user is going online and looking at a publicly reported measure and wants to know who are these people in the denominator. It doesn't answer all those questions, but it helps you along to know whether you are only capturing a small proportion - that is what I'm arguing that it should be approved not because it is in scope, but because it adds to the usability of the other two measures.

   DR. GOLDBERG: That is the only reason I see to support it. Standing by itself. So why have the complication of another one? Why not simply change the others to say, your first measure is, how many had a baseline, rather than having this other thing floating around out there.

   DR. PHILLIPS: Then you are
radically changing the measure because you
have a new denominator.

DR. GOLDBERG: Right, but when you
start up eventually it's the same thing. I
mean, essentially you are changing the
denominator. If you link this, you are
changing the denominator.

DR. STREIM: No, I think what
you're suggesting would require a whole new
measure, set of outcome measures where the
denominator is ICD-9 diagnosis, and that is --

DR. PHILLIPS: But that is what this
requires, right?

DR. STREIM: No. No, this does
require it for this measure.

DR. PHILLIPS: Right. So it's not
different. If you are linking it you are
doing the same thing.

DR. WINKLER: Right. There are
multiple approaches to get to the same thing.
One of the reasons people like to keep them
separate is because they become more
actionable. If all you have is a low score on the outcome measure, you don't know without being able to break it down how many just never had the test in the first place versus how many had - did not, you know, change over the timeframe, whereas if you break them down. But we've seen both kinds of measures.

DR. PHILLIPS: Then this isn't just process.

CO-CHAIR LEDDY: Why couldn't you just change the first two measures, to measure the first two measures, but then using the current database use the same numerators for the first two measures and come up with some other measures that use the ICD diagnosis, ICD-9 diagnosis population as the denominator. Wouldn't you get to the same thing? Too confusing?

DR. STREIM: But nobody has done that, and nobody has submitted a measure like that, so we don't have an option to work with that right now, somebody unless next year
somebody or the year after does that.

CO-CHAIR SUSMAN: The other concern I have about this is that patients have a diagnosis of depression, and at least in primary care, it is not uncommon to carry that diagnosis forward for a long time. So if I documented a PHQ and the person's remission, then the question becomes, well, how frequently should I surveil patients with treated depression for recurrence? And frankly the data are not, I don't think, very robust. So we are adding a substantial burden since depression is an extremely common diagnosis in primary care. Now, we could argue whether that would be on the whole a good thing or not, but the question I would say is, gee, is that burden, which is getting down here a little bit. And I see both Harold and Bob.

DR. PINCUS: One question: What is the current U.S. Preventive Health Services Task Force recommendations with regard to
depression screening?

(Simultaneous speaking)

CO-CHAIR SUSMAN: Once a year.

DR. PINCUS: Is it once a year? So that's in the general primary care population, and this is likely to be an enriched source of people, it actually is good evidence, and it's an enriched source of people who currently have depression symptoms above threshold. So one could easily say that certainly once a year would be a reasonable amount to do that, certainly if somebody is still carrying a depression diagnosis.

CO-CHAIR SUSMAN: This is a four month, not a year measure.

(Simultaneous speaking)

DR. PINCUS: I'm just trying to say -

DR. ROCA: Can I make a comment here? This is a very interesting discussion, and I don't suppose we're following Roberts Rules of Order, but I feel an urge to call the question right now. Because I think some of
us are going to think this is a process
measure and shouldn't - isn't within scope.
I think some people would think it ought to be
within scope. I think we are just going to
have to at some point vote on it, because I am
not sure we are going to come to consensus.

DR. STREIM: Just one other
question about - or clarification, the four-
month measurement period that you refer to
here, that begins in someone who is first seen
in a health system and gets a diagnosis, an
ICD-9 diagnosis of depression?

MS. PITZEN: Correct. They would
have to have a visit with that diagnosis in
that timeframe that you are measuring.

DR. STREIM: So it is possible to
have somebody who has been depressed for 20
years, but what would define the measurement
period is - it has to start with the
availability of an electronic record that has
an ICD-9 code in it, correct?

MS. PITZEN: Correct.
MS. JAFFE: You wouldn't - building off what Joel was saying, if they are not scheduled to come in every four months, you wouldn't have them come in simply the screen, would you?

MS. PITZEN: No, they actually couldn't be counted, because the identification of those patients are, you have to have a visit with a diagnosis of major depression or dysthymia in that time frame. If you don't have a visit during that timeframe you are not even in the denominator.

CO-CHAIR SUSMAN: I think we ought to take a vote. I want to make a comment though. As much as I like to improve care, and you said this is a good process measure, this is in preparation for the vote, of all the process measures this is a good process measure. It would help improve care. I don't think that's what we are here for. I think we are here to identify measures that are outcome measures, that's why I think we need to have
a vote.

CO-CHAIR LEDDY: Okay, how about if we call - okay, you have a question or comment.

DR. HENNESSEY: Is there a way we can vote on this as linked?

CO-CHAIR LEDDY: Maybe Reva could explain - I thought that what we were going to do is vote whether it's in or out of scope as an outcome measure, and then or now I want Reva to explain more clearly what you mean by a process measure is linked. Where does that vote go?

(Laughter)

DR. WINKLER: These are cumulatively, and that's why sometimes we can put ourselves in a box. But one of the alternatives if you are concerned about it not being an outcome measure and out of scope but you still feel there is something valuable about it and you would like to maintain it in some way is, you do have the option of linking
it or pairing it is what we say so that you
would have the paired process measure paired
with, say, the six month outcome measure such
that if you did one you did both, the two
travel together. They are really two parts of
the whole recommendation. And so that is
always an option. And that is a way of
getting around, you have a dangling process
measure. But for those of you who feel it has
value to the outcome measures, this is a way
of using it.

DR. PINCUS: Separate votes?

DR. HENNESSEY: So, okay, I'm just
trying to clarify. I'm on The Price is Right,
I'm on the TV show, I've got Door # 1 saying
doesn't meet scope, not important. Door #2
says, does meet scope, important, and we can
go down the complete partial.

DR. WINKLER: We got a bunch of
doors. We've got the Winchester Mystery
House, actually.

DR. HENNESSEY: So we got more
doors.

DR. WINKLER: Well you are talking about two measures at a time is what is going on. So I think the question probably first off is, is there strong enough feeling by the majority that this measure is out of scope under all potential eventualities, linked or not linked, separate, or whatever. So should we just take all potential eventualities, linked or not linked, separate, or whatever? So should we just take it off the board altogether because it is just out of scope for the project.

Do that, and then we can do the ones that follow. Does everybody get that?

DR. ROCA: Can I just - I would vote that it is out of scope, but if there were an option saying that if you were going to use either of the other two then you are also having to report this, then the other two are the primary measures and this is just sort of a hanger on and I would vote for that.
(Simultaneous speaking)

DR. WINKLER: Yes, you have three votes. So can we vote three options?

DR. STREIM: So if the initial vote is on in or out of scope, up or down, that doesn't preclude further votes. It's not like the Senate where discussion is ended, you will never hear about this again, right?

DR. WINKLER: Right.

DR. STREIM: Okay, thank you.

DR. WINKLER: Let's try it in kind of two steps. The first one is in or out of scope. So if you vote that it is out of scope it does not come back; it's gone, goodbye, keep that in mind.

CO-CHAIR LEDDY: So this is not in or out of scope as an outcome measure. No, this is in or out of scope of whether you ever want to hear about it again. That's what you're saying, Reva. That's different.

DR. WINKLER: If you say it's out of scope it's because it's a stand alone process
measure you feel does not have any role in the outcomes work you are doing. Is that fair?

(Simultaneous speaking)

CO-CHAIR LEDDY: So what I'm putting up here, does this make sense, out of scope, in scope, and then in scope would be - as a stand alone. I think you would definitely have to break it down.

MS. MASLOW: What if we vote on what we want first?

DR. WINKLER: I'm hearing we want something totally different.

MS. MASLOW: So what if we vote on that instead of making us make an illogical statement.

DR. WINKLER: Okay, what do you want?

MS. MASLOW: We want it to be tied to one of the other measures, and it is in scope in that context.

MS. BOSSLEY: So we can switch it, so if for some reason it doesn't pass as
paired with one of them, then we'll go back to
the out of scope. So I think that is what you
are getting at right? Does that make sense?

DR. KAUFER: We have already

endorsed this.

DR. WINKLER: We have?

DR. KAUFER: Well, logically we
have by approving the other two outcome
measures, we have tacitly approved this
measure as part of - as part of that outcome
measure.

DR. WINKLER: No. There is a four
months window.

(Simultaneous speaking)

CO-CHAIR LEDDY: I think the group
is saying that they - we don't want to say
this is an outcome measure, because it would
be silly to say that. But we would like to
consider it as a hanger on, but clearly
process. Is that what we are saying, because
it will help the other outcome measures.

(Simultaneous speaking)
CO-CHAIR LEDDY: Joel?

DR. STREIM: I will just restate. I believe that this is a process measure by itself. As a stand alone, it is not an outcome measure. However I think it's important to measure because it helps improve and enhance the usability and interpretability of the two other measures we just voted to endorse.

CO-CHAIR LEDDY: So how about if we have a motion, and we vote. That is very well stated, and why don't we say whether we agree with that statement or not, and that is what we will be voting on. Is that okay, Reva?

DR. WINKLER: Yes.

CO-CHAIR LEDDY: No?

DR. ROCA: But does this mean, is this voting to say that this would be a stand alone measure? Or that it would have to be - because Joel, what you implied is that it was not really an independent measure or a stand alone measure.
DR. STREIM: I don't think it meets the criteria as a stand alone outcome measure. It certainly could be a stand alone process measure, but that is out of the scope of this committee's scope definitions from last November. So maybe, I don't know if we need to disaggregate those statements and vote on them separately or you want to do the package. That is really the chair's prerogative.

DR. GOPLERUD: I'd like to suggest, based on what we did last November, developing an incredibly broad definition of outcomes, which included population health, the social determinants of health, you know, we basically voted on climate change as health outcomes.

DR. STREIM: As health outcomes, though, not as processes of care. Not processes of care.

DR. GOPLERUD: Okay, but given the incredible breadth that you all accepted, or we all accepted as being outcomes, why not just define that we like this measure and know
that it is a process measure, and say that we endorse it anyway?

DR. STREIM: Well, because I think we have a process here that allows us to endorse this as a linked measure that enhances the usability and interpretability of the other two outcome measures we endorsed. I know I'm being redundant, but I think that is really the legitimate reason for this committee's - within the scope of what this committee really did lay out last fall.

CO-CHAIR LEDDY: So that latter little bit shorter statement, can we vote on that? That was very good. Would anybody like it repeated?

DR. PINCUS: I missed it.

CO-CHAIR LEDDY: Can you repeat that latter statement, Joel?

DR. STREIM: You want the latter, not the former. Well, the former was the aggregate statement, let me do that, and then if you want a shortened version I will try and
reiterate. As a stand alone measure this really is not an outcome measure, it's a process measure, so technically out of scope. However, I think it is a measure that enhances the usability and interpretability of the other outcome measures we just endorsed, and therefore, I believe it should be endorsed as a linked measure to each of the other two.

CO-CHAIR LEDDY: Are there any questions about Joel's statement?

DR. GOLDEN: The comment is that I think we have before us that we have endorsed a concept, the concept of the measurement of status through this tool. The problem we have is, I think is the measures themselves could be made stronger, and we are now cleaning up imperfect measures that unfortunately that is not the rules of the game. But I think that we are taking measures from a community that I think, if we had more time to work with, there would be a better numerator and a better denominator.
CO-CHAIR SUSMAN: So just a point of clarification from the measure developer. If I had depression diagnosed at time zero, and let's say I come in at five months, and I have depression diagnosed at five months, and there wasn't a PHQ in the first five month interval —

MS. PITZEN: You weren't seen in the office.

CO-CHAIR SUSMAN: Well, let's say I was seen in the office.

MS. PITZEN: If you were seen in the office in that first five-month interval —

DR. WINKLER: Could you use your mike, please.

MS. PITZEN: If you were seen in the office in that first five-month interval, had the ICD-9 codes applied to one of your visits and then if you had a PHQ-9, that would be counted. But if you were not seen in the office during that time with the depression diagnosis you would not be in the denominator
for this process measure.

CO-CHAIR SUSMAN: That seems pretty much garbage in garbage out in the sense that it is implying that there is a follow up and then there is rediagnosis. I understand from a community adoption spread of diffusion of the technology if you will why this is being used. I still am worried about this as an accountability measure, even when linked to the other two. I also wonder then, to just take my question one step further, then I'll let the vote occur, is if I had that first five months, and let's say I didn't come in, and then let's say at the six month I get another diagnosis of depression, it starts over again, or are you excluded? Or what happens?

MS. PITZEN: Let me see if I can try and explain without being too confusing. The denominator is different for the remission measures and this process measure.

CO-CHAIR SUSMAN: Right.
MS. PITZEN: So going back to the remission measures, if you are diagnosed with major depression or dysthymia and your score is ten or above, you are in.

CO-CHAIR SUSMAN: Right.

MS. PITZEN: And if you never see your provider again over the next seven months, because we do allow a plus or minus, grace window, then you fail.

CO-CHAIR SUSMAN: Right, got it.

But now for this current measure --

MS. PITZEN: Right, for this current measure it doesn't matter what your PHQ-9 scores are, you are in the denominator if you have depression or dysthymia.

CO-CHAIR SUSMAN: And is that a denominator that lasts just four months?

MS. PITZEN: Four months. Right.

CO-CHAIR SUSMAN: So if I came in at time zero and had the diagnosis, you would have one to get the PHQ within the four month period. If I came in at five months time
frame with depression that would be a new
episode of measurement.

MS. PITZEN: Correct.

CO-CHAIR SUSMAN: If I came in at
eight months or nine months it would be yet
another episode of measurement; is that
correct?

CO-CHAIR SUSMAN: Correct.

CO-CHAIR SUSMAN: Okay.

DR. STREIM: Can I also comment on
the issue of unintended consequences, which
will always be our concern here. If this is
endorsed linked to the two other outcome
measures and it is not endorsed to be used as
a stand alone process measure, then there
wouldn't even be a situation where someone
would get dinged for not doing a PHQ in the
first four months, because - let me finish -
because it would only be used in conjunction
with the outcome measures, and - that we just
recommended for endorsement, and therefore to
get in those denominators you have to have a
PHQ. So nobody is going to get dinged for not having a PHQ as a result of endorsing this as linked.

DR. PINCUS: But linking it does not require that they have the same denominator, correct?

DR. STREIM: No, not at all. All I'm saying is, I'm just addressing the concern or potential concern that people may have that if we endorse this in any way that failure to have a PHQ, in particular that tool on the chart, is going to result in health care provider or system getting dinged. That won't happen the way I last stated it in the proposal to endorse.

DR. WINKLER: Just as a clarification, when we talk about linking them, what we are doing is saying that when these are implemented the expectation is that they will be used together so that you will get a report of the results of this measure and the results of the outcome measure.
It's not a composite, it's just that the two travel together. So it's not a cafeteria; you don't get to choose one and not the other. We're saying do them both.

DR. PINCUS: The current sort of set of the three depression measures that you have endorsed are there? They are? So this must be a reasonable thing. So just one point about this being—could it also be done, could it be also as a separate measure, too?

Could it be linked and also separate?

CO-CHAIR LEDDY: Not by our group; we don't do process.

DR. PINCUS: Well, no, in that case, as I looked at the list of measures, only four processes—definite outcome measures on our list. I'm just saying that when we actually sent out a call, we enlarged the notion of outcomes.

(Simultaneous speaking)

CO-CHAIR LEDDY: We redefined outcomes sort of broadly?
DR. PINCUS: Right, so what I'm saying is, that is a question I have is if this is - you know if we are taking a very strict - if now we are taking a very strict notion of what is outcome versus process --

DR. WINKLER: I would hope you are internally consistent in your notion of outcomes.

DR. PINCUS: My view is that this is one of the better process measures that actually has pretty good evidence linking it to outcomes so that that is why - so from my point of view, I think that as an outcomes-related process measure, whatever you want to call this sort of enlarged Venn diagram, it has significant value. But also I think it helps to interpret those other two measures, because you get a sense of what they didn't capture.

CO-CHAIR LEDDY: So you would like to amend the statement that Joel made that where you said that it would be useful in
coordination with these measures to interpret
the other outcome measures, and it sounds like
Harold is saying that it also should be
considered as a separate, as a stand alone
vote. So we could --

   DR. PINCUS: You need two votes,
and a stand alone vote, that is correct.

   CO-CHAIR LEDDY: Right.

   (Simultaneous speaking)

   DR. PINCUS: And the rationale for
that is that I think there may be
organizations that choose to only do the
remission measures, and it would be important
for them to have that information linked if
that is what they are going to do so they can
interpret them better. And on the other hand
there may be organizations that don't want to
use the remission measures but want to have a
sort of outcome-related process measure.

   DR. STREIM: I could be convinced
that it should be recommended as a stand
alone. I could be convinced, but I have a
question based on Reva's last clarification about harmonization, whether we could even link these because if I could wrap my brain around this part, it looks like if you link them and they are traveling together and you have to do them all, if you have an ICG - no I guess I've answered the question, it doesn't matter.

MR. PELLETIER: The four months, how did you decide that? That's when you kind of report things in your organization?

MS. PITZEN: Correct, it aligns with the data submission.

MR. PELLETIER: Right, so I don't think we should be getting hung up on four months because it's the way they are reporting compliance with getting a PHQ for someone with three diagnoses. That's all that is. You can do that in two months; you can do that in eight months. You can do that yearly, you can do that every two years.

DR. BOTTS: I think the idea is
that what you are getting is a cross section
of how many people are doing measurement based
care. So it just gives you a figure of how
frequently are we getting those, and that is
important in terms of interpreting the
outcome. As a process measure, even as a
stand alone, it's not necessarily tied to, you
are getting a clinical assessment that is
applied temporally with the initiation or
management of treatment. It just says, you
have been seen, you have an active diagnosis,
and we have assessed you with this tool. You
could be eight months out; you could two weeks
out; you don't know in that process. So even
as a process measure I would say that it needs
work. As an add on to our outcomes, I think
it makes a lot of sense.

CO-CHAIR LEDDY: So why don't we
take a vote, then. Joel put on the table
about the add on that this is the add-on to
help interpret the first two that we
recommended.
DR. PINCUS: So this is a paired or linked measure? Is that correct?

DR. STREIM: And just again to be really clear, by doing that, and I am still struggling with the unintended consequence thing, it means when they are performed they will all be performed together, meaning all three?

DR. WINKLER: You've got again more options. Which ones are you linking? Are you going to link the process measure with both outcomes as a triad or link the process measure with each outcome independently?

DR. STREIM: But even if you do it with each of them independently, it means that everyone with an ICD-9 diagnosis will be included in the denominator at a minimum.

DR. WINKLER: At the first measure.

DR. STREIM: Right, and then the second measures would be applied to those, but that is where the harmonization problem is; you couldn't do it unless you had a PHQ score.
DR. WINKLER: Exactly.

DR. STREIM: So that is the harmonization issue; it doesn't matter.

DR. WINKLER: It doesn't matter. That isn't so much harmonization. The numerator of the first one --

DR. STREIM: It doesn't preclude you from doing that.

(Simultaneous speaking)

CO-CHAIR LEDDY: Does everyone agree, then? So the recommendation that we are going to vote on, yes or no, is going to be Joel's statement with the linking with Richard's caveat about linking independently, and - did you have another caveat Rich? That's it. Okay. So we are going to vote yes or no. How many vote yes to recommend that?

(Show of hands)

DR. WINKLER: Fourteen.

CO-CHAIR LEDDY: And how many vote no?

DR. WINKLER: One.
CO-CHAIR LEDDY: And how many vote, abstain?

DR. WINKLER: None.

MR. CORBRIDGE: Eric is out of the room.

DR. WINKLER: Eric is out of the room and Carol is out of the room.

CO-CHAIR LEDDY: Okay, and Harold's back, so he voted. So this is whether we would like to recommend this as a process measure or as - as a stand alone measure. As a recommended measure.

MS. BOSSLEY: You would be recommending this measure in the NQF portfolio that would be used by itself by anyone and everyone as long as they report it.

CO-CHAIR LEDDY: Within our scope.

(Simultaneous speaking)

MS. MASLOW: So this is recommending it as an outcome measure?

(Simultaneous speaking)

DR. WINKLER: One of the issues
around scope is it helps us limit what we - we could bring you guys 200 measures to play with if we didn't put some boundaries around what we wanted to talk about. It also provides the field when we ask for the call for measures, and submissions, to tell what we want to consider. So that is all the scope does. Once they go through the process, these could end up in the portfolio to be used.

DR. HENNESSEY: Sheila, you had a comment about this measure from a process perspective.

DR. BOTTS: Well, my comments were related, I think what this measure, this process to me just says, are we using measurement-based care or not. Are you getting that tool? It doesn't tell you about the meaningfulness of when you are doing the assessment or how that might relate to treatment decisions. Just that when you see a patient with a diagnosis of depression using a measurement based tool to assess. And so
that is probably acceptable as one process measure. I would like to see other process measures that said you would have this within \( X \) time from the initial diagnosis or the initiation of treatment. But this at least says, are you doing it, and I think that is an important measurement, but we could go a step further in terms of where it falls in treatment.

DR. PINCUS: Or we could actually say that when we get into what our recommendations are for further development.

CO-CHAIR LEDDY: But that is not right now. Are we ready to vote on this measure recommending it as an independent measure by this board? All in favor?

(Show of hands)

DR. WINKLER: Six.

CO-CHAIR LEDDY: Should we do it again?

MR. CORBRIDGE: Seven.

CO-CHAIR LEDDY: Okay, and then -
or opposed to recommending this as an
independent measure?

(Show of hands)

DR. WINKLER: Seven, it is a push.

Did everyone vote?

CO-CHAIR LEDDY: Oh, I'm sorry, I
didn't ask for abstentions. One abstains.

You want to change your vote?

DR. GOPLERUD: Yes, for independent.

CO-CHAIR LEDDY: So it's eight and
six then, eight, six and one.

(Off the record comments)

CO-CHAIR LEDDY: So this is - do we
have anything else to do before lunch, Ian?

MR. CORBRIDGE: No, at this point
this concludes the first section of workgroup
number one. So at this point in time we had
planned -

(Simultaneous speaking)

MS. BOSSLEY: We need to know if
you feel - again I think well you are actually
evaluating it both ways, stand alone and
linked. Does it meet the importance criteria?
Does it meet scientific acceptability,
usability, feasibility? You have now
determined it would be used alone and linked.
So as a measure itself.

CO-CHAIR LEDDY: Okay, so are we ready to vote? We've had a lot of discussion on this measure. Can we vote on importance?

DR. WINKLER: Does anybody think it's not important?

Okay, great. What is the next one?

Scientific acceptability. Does anyone think it completely meets the criteria?

Partially meets the criteria? One, two, three.

MS. MASLOW: Did you assume completely?

CO-CHAIR LEDDY: I saw no one vote. Did you want to vote completely Katie?

DR. WINKLER: Shall we start over?

MS. MASLOW: I will vote partially.
(Laughter)
DR. WINKLER: Twelve.

How many minimally? I saw a couple of no votes. Did you vote?

MR. PELLETIER: I didn't vote.

DR. WINKLER: How many abstain?

MR. PELLETIER: You know what it is? When you develop a measure you want people to do something, okay. You then collect your data, but the implicit is that they are doing it. That they are going to do this, that what you have asked them to do they are going to do, so that is going to be part of the measure. It shouldn't be this add-on later that says, oh let's check if they are doing it the way we want them to be doing it. So that's where this is very - someone said it before, we are fixing a measure that is not perfect.

DR. PINCUS: I don't agree with that notion that you are fixing it. It just gives a broader perspective. For the denominator of the two remission measures,
it's a good measure for looking at remissions, but what you don't know is with the population that the organization is dealing with, you don't know the extent to which the - you are getting information about the broader depressed population.

MR. PELLETIER: But don't you always want to know that?

DR. PINCUS: No.

MR. PELLETIER: I think you do. I disagree.

DR. PINCUS: I would say that for the vast majority of NQF-endorsed measures they are very specific to the very specific denominator, and they don't give you a broad perspective.

DR. WINKLER: We need to just sort of finish this out.

CO-CHAIR LEDDY: So the next one to vote on for this measure is - we voted on scientific acceptability. Okay, usability?

MR. PELLETIER: And this is the
paired vote?

MS. BOSSLEY: No, this is the process measure. We are evaluating this measure on its own. Not linked.

MR. PELLETIER: No, either way.

DR. WINKLER: It's either usable or it's not.

MS. BOSSLEY: I think because you have determined that you feel this measure could be used alone, you need to evaluate this measure on its own face value, on whether it meets the criteria or not.

MR. PELLETIER: I don't think that was understood when you had the last two votes.

MS. BOSSLEY: Well, that is what I am wondering, was that understood or not?

CO-CHAIR LEDDY: Okay, so let's go back and redo importance as an independent measure. Importance is the first. Importance to measure and report, completely.

DR. WINKLER: Anyone disagree?
That's almost easier.

CO-CHAIR LEDDY:  As this measure, evaluating it without thinking about the other two. On its own face value, does it meet the importance criteria, completely, partially, minimally, or not at all?

(Simultaneous speaking)

DR. PINCUS:  The thing that is disarming is that this is so far superior to every existing NQF depression measure that it is not even funny.

CO-CHAIR LEDDY:  So does it completely meet the importance in your mind?

MS. MASLOW:  Assuming it is a process measure.

CO-CHAIR LEDDY:  It is a process measure.

(Simultaneous speaking)

CO-CHAIR LEDDY:  Okay, so how many are completely?

(Show of hands)

I have 13.  Okay, how many are
1 partially?

   Two.

3 MS. BOSSLEY: Any others? I think we've got minimum.

4 CO-CHAIR LEDDY: Minimally. So the next category is scientific acceptability.

5 How many vote completely?

6 (Show of hands)

7 CO-CHAIR LEDDY: How many vote partially?

8 MS. CORBRIDGE: I have 13.

9 MS. BOSSLEY: Late hand. 14.

10 CO-CHAIR LEDDY: Okay, any minimally?

11 Any abstentions? Okay.

12 Next category is usability. How many vote completely?

13 MR. CORBRIDGE: Got seven.

14 CO-CHAIR LEDDY: How many vote partially?

15 MR. CORBRIDGE: Six.

16 CO-CHAIR LEDDY: Is that everybody?

17 Any minimally? And any abstentions or not at
alls?

MS. BOSSLEY: We are missing someone.

CO-CHAIR LEDDY: Okay, let's do completely again. We are missing someone in one category.

MS. BOSSLEY: Eight, nine of eight, okay we are good.

CO-CHAIR LEDDY: Now we are on to feasibility. So how many people would like to vote that this is completely on the feasibility measurement?

MS. BOSSLEY: Ten.

CO-CHAIR LEDDY: How many partially?

MR. CORBRIDGE: Four.

CO-CHAIR LEDDY: And how many minimally? Two? And any abstentions? No?

Okay, now we have to vote on - oh we did it backwards. So we already recommended - and do we have anything else to do before lunch? Are you going to tell us about lunch,
Ian?

MR. CORBRIDGE: I guess at that point we do conclude with that section. We have lunch right out here for the Steering Committee Members. We are hoping if we can do it quickly, I know we are a little bit over schedule, so if you don't mind take a half hour or 15-minute break to have lunch, make some phone calls, and if you would come back and start on the major process here again, that would be wonderful.

(Whereupon at 12:42 p.m. the proceeding in the above-entitled matter went off the record to return on the record at 1:15 p.m.)

CO-CHAIR SUSMAN: So we are going to go ahead and get started. I appreciate everybody's good participation during the last session, and I will try to facilitate this with the able assistance of Tricia and the rest of the NQF staff.

So we are going to do readmission
and mortality. This is suicide deaths, and then a bunch of readmission criteria.

READMISSION & MORTALITY MEASURES

CO-CHAIR SUSMAN: The group is Ann, Darcy, Joel, and Glenn. And I guess you are somewhat grouped over on the end here. So we will look forward to your thoughts about each of these. Just to review the process, we'll first decide whether it's in or out of scope, make sure that we are doing this as an outcome and not process measure; and then go through the drill which, I think, everybody has probably caught on to by now.

So the first measure I have up is the suicide deaths, at-risk adult psychiatric inpatients within 30 days of discharge.

MEASURE OT3-001: SUICIDE DEATHS OF "AT RISK" ADULT PSYCHIATRIC INPATIENTS WITHIN 30 DAYS OF DISCHARGE

CO-CHAIR SUSMAN: And would you like to give us the brief overview?

MR. CORBRIDGE: Sure. So as Jeff
started out, we have the title, which is
"Suicide Deaths of At-Risk Adult Psychiatric
Inpatients Within 30 Days of Discharge". The
description for this measure is rate of
suicide deaths within 30 days of discharge
from the inpatient psychiatric setting, adult
patients aged 18 and older, rated as "at
risk."

The numerator statement reads as
follows: suicide deaths of at-risk adult
patients within 30 days of discharge. The
denominator statement reads, adult inpatient
discharge with a pre-discharge suicide
assessment that affirms any of the at-risk
inclusion criteria and do not meet the
exclusion criteria.

And the information from that
measure, the subcriteria, is posted up there.
So from our group any concerns that this isn't
an outcome measure?

It is a terminal outcome -- I think
it's probably an outcome that matters to
patients. So I think we are all in agreement there. Why don't we talk about importance? I'll look to the group for some initial comments.

DR. STREIM: High impact.

CO-CHAIR SUSMAN: Everybody agrees this is a high impact outcome, probably self-evident.

DR. PINCUS: So the incidence of suicide post-hospitalization.

CO-CHAIR SUSMAN: So the question is, what's the incidence of suicide post-hospitalization? Is this an important issue, one that's prevalent?

DR. PINCUS: It's obviously important from the point of view of, it's a catastrophic event. But if a hospital has one of these every year, how stable is something like that?

DR. STREIM: We know that compared to other kinds of health outcomes this is a low frequency event. But most of
our suicidology colleagues would probably say
that it's one of the hardest things to study
in terms of knowing what incidence rates are
reliably. I don't know that that adds
anything.

DR. PHILLIPS: I think that gets to
a point too, that if you look - our importance
ratings are very different from the rest of
our ratings of this measure, and it's that I
think - it's readily apparent that tracking
suicide is important, but we have lots of
questions about usability and feasibility of
this measure.

DR. GOLDBERG: Is this a Joint
Commission report?

DR. PHILLIPS: I don't know.

CO-CHAIR SUSMAN: So the question
is, is this a reportable joint commission -
does anybody know?

MS. JAFFE: No, it's not.

The reportable events are suicides
that happen during hospitalization.
CO-CHAIR SUSMAN: Thank you.

DR. GOLDEN: So the question in terms of the importance of this measure on the issue, I noticed, like, the last one you had to have had a suicide risk assessment process, with about six or seven things, does that limits the utility of this as opposed to just saying hey, anybody who committed suicide after discharge from psychiatric hospitalization.

DR. STREIM: Do we address that in scientific --

DR. GOLDEN: I guess my question for you, since I'm not doing inpatient psychiatric care, are these criteria used commonly, or are they not particularly - this happens to be somebody's list?

DR. STREIM: I'm not aware of anybody who is using post-discharge suicide to measure quality at this point, but I'm not a suicidologist.

DR. GOLDEN: I'm talking about risk
DR. STREIM: I was just saying I think we have addressed that under scientific acceptability, right?

MS. JAFFE: I think one of the issues about, is this an important thing to measure or not is, I think nobody will disagree that measuring suicide is important, but measuring it 30 days after discharge is another question. And I'm not convinced that it's all that important to measure at 30 days out. Number one, because it hardly ever happens, so it's not clear what we'd be measuring, but there are just so many things that can happen within 30 days after discharge from a hospital. It's not clear to me that this is the important thing to measure about suicide.

CO-CHAIR SUSMAN: So part of the discussion we are starting to get into it sounds like, perhaps, is the scientific acceptability sort of issues, and maybe
usability issues.

DR. STREIM: Well, I think even if
we just stick with the three, impact, gap, and
relationship to outcome items, maybe just do
this systematically as we've laid out the
process. In terms of the gaps, one of the
things we are looking for is disparities
across population groups, variability across
provider groups, and I'm not, again, a
suicidologist, but I couldn't find anything
published on post-discharge suicide rates
across health systems, anything that does
anything comparing performance, whether there
are health systems that do that internally I
don't know. I didn't look at that myself as
part of my review. I don't know if colleagues
did. But those of you who are health system
administrators, maybe, can comment on that.

DR. ROCA: We certainly, and I'll
try to get some specificity here, but there is
a reporting practice, if not a reporting
requirement, for suicides that occur within a
certain time period after discharge, and it
may be 72 hours, I can't recall exactly, and
I'll try to get that number, but certainly 30
days is outside that window. And of course
you don't always know if a suicide has
occurred within 30 days, there are certain
practical problems with ascertainment. And it
certainly is a rare event fortunately, but
it's obviously a high impact outcome that we
would all strive to avoid.

DR. GOLDBERG: On this issue of 72
hours versus 30 days partly is an artifact of
we have balkanized our health care system to
inpatient, outpatient, and diverse care, and
what we are really interested in I think is
how people do over an episode of care of their
illness. And at some point it may be that
suicide is 30 days after inpatient, the
inpatient phase of the episode of their
illness, would be an important outcome. So I
have that feeling which makes me think it's
important. I don't know if our system is
quite ready for that. What our system is ready for is some – maybe not this, but engagement and follow-up treatment, which a number of people are trying to get at, either by communicating discharge plans or outpatient appointment being made and kept, that's our system creeping towards taking care of the person across the episode of their illness. So what we are doing is make sure at least you tell somebody that they left the hospital, and you get a report to them, and they get a follow-up appointment, and you give them medication. But that's not this measure, so as important and striking as this is I have questions of whether this is the right time for this measure.

DR. STREIM: Well, one of the things that is not specified at least in the materials we had access to from the measure developer here -- is the measure developer here on the phone, do you know? Sometimes we can ask for a clarification.
MR. CORBRIDGE: It's Psychiatric Solutions, and they are not here. I haven't heard them on the phone.

DR. STREIM: One of the questions is, if we are measuring the quality of an inpatient stay, which is when the patient is identified as being at risk in the way this measure is proposed, then looking at the 30-day period after the hospital stay depends -- you know, the outcomes depend heavily on the transitions in care, what part of the system is the patient being cared for. And again, that goes to the scientific acceptability which we haven't even gotten to yet.

DR. WINKLER: Just for context, because this is sounding like a very similar discussion, over the last couple of years NQF has in other topic areas, notably around AMIs and pneumonias and heart failures, moved in the direction of 30-day post-hospitalization mortalities readmission. So the idea that transition of care, that the hospital has a
role to play in sort of setting and assisting
the trajectory of this patient to a successful
transition into the outpatient world it's
challenging, the data collection can be quite
difficult. But that is a direction that
measurement is moving in at a fairly rapid
clip, so we are certainly seeing in the main
outcomes, historically a lot of the measures
are, the data can be coming from both
inpatient and outpatient, coordination between
those two different settings of care is very
very much trying to get at this whole episode
of care.

So don't, I really would caution you
against, don't let that stop, because you are
going to find that this idea of that follow-up
after hospitalization is really of significant
importance in measurement that we are seeing
now.

DR. HENNESSEY: So mortality,
within 30 days of hospitalization discharge,
is becoming more prevalent within NQF
especially.

CO-CHAIR SUSMAN: So I am hearing that everybody acknowledges that suicide is a high impact condition, that while there is probably a gap in overall care, the gap demonstrated here isn't really very well articulated, and the relationship to outcomes obviously is there. So are we ready to vote on importance here? Are there any new concepts or questions?

So how many people would say that we have completely met the importance? Raise your hands please.

(Show of hands)

CO-CHAIR SUSMAN: How about partially?

(Show of hands)

CO-CHAIR SUSMAN: Okay, so we will move on. The next part, and I think we already started to talk about this a bit, was scientific acceptability. Let me ask the group if you can shed some light on this
further. You will see there are lots of comments up there.

DR. MANTON: The denominator statement I thought was complete. A lot of what was there was to be determined, which is, I think, why that whole section really is blank. Just about every measure, reliability, validity, said it was to be determined, to be determined, to be determined. So we really don't have anything to go by.

CO-CHAIR SUSMAN: Who is the measure developer?

MR. CORBRIDGE: It is Psychiatric Solutions, Inc. And I guess because they are not here, I have discussed it with them, so I'll just kind of help inform that conversation. They submitted under the intent call for measures for this project, and after having a discussion with them they realized that their original measure didn't really target the outcomes project. It was more process oriented. After that conversation
they went back and restructured their measure,
and this is I guess that second draft, and
they are currently, right now, testing that
measure, but that is why there is kind of a
lack of that information is because they are
now going through that process. The numerator
for this measure is suicide deaths of at-risk
adult patients within 30 days of discharge.

DR. STREIM: The devil is in the
details. If you look at there are six factors
that define at-risk.

DR. PINCUS: Do you look at death
certificates? Is it mortality reports, or
what's the --

MS. JAFFE: They do talk about that
in feasibility, but they expect that you would
try to contact these people.

DR. PINCUS: It is hard to do.

MS. JAFFE: That was one of the
comments. And if you don't contact them they
are not included.

DR. HENNESSEY: How do you
determine at-risk? How is that determined?

CO-CHAIR SUSMAN:  It looks like there is a sixth criteria, patient verbalizing despair and anxiety, admitted for suicidal or self-injurious behavior, history of post-discharge suicide attempts, complete discharge safety plan, admitted with significant suicidal ideation, on suicide precautions, yada yada yada.

DR. STREIM:  The yada yada ya is what matters here.  So the last thing in the list - I think it's the last one - is actually that the patient has had a suicide - a discharge safety plan.  Now that basically undermines in terms of the face validity of the measure it basically undermines the whole intent.  If you have already done the safety plan and responsible discharge planning, a la what Reva was referring to, and make sure they are connected to follow up care and monitored properly, that should move us in the direction of suicide prevention of the thing.  But if
you exclude, systematically exclude anybody who has not had a safety plan, then you have excluded from your denominator the universe of people who are truly at the most severe risk. So I see a structural problem that really undermines face validity. So that is my biggest concern.

CO-CHAIR SUSMAN: Any of the six - it isn't all six.

DR. HENNESSEY: Pre-discharge suicide assessment that affirms any of the following at-risk categories.

CO-CHAIR SUSMAN: So they might not have had the --

DR. STREIM: But the point is still that they built in an exclusion essentially.

CO-CHAIR SUSMAN: Is there other comments about scientific acceptability or questions from the group as a whole?

DR. PHILLIPS: One of the things that we talked about earlier is that they don't - they essentially have no plan for risk
adjustment. And there are certainly many things that can affect this, case mix being the one that most readily came to my mind. And the fact that there is essentially no plan to do that is a little concerning for this as a measure.

DR. STREIM: The fact that they actually indicated that that wasn't applicable here was really - I mean, to have a measure steward look at a measure like this one and say, we don't need to worry about risk adjustment is a concern. It's a concern about the acceptability, scientific acceptability of the measure, but it's also a concern going forward about the stewardship.

CO-CHAIR SUSMAN: Those points are good. Other points from the group? Or questions from the committee?

DR. MANTON: Just overall I don't see how you can make a determination on this section, because there is so much that isn't done.
(Simultaneous speaking)

DR. HENNESSEY: What is troubling about this is that this is a very very important issue but the way it is hammered out is highly lacking, and when we talk about topics to put on a parking lot, this would certainly fit that.

CO-CHAIR SUSMAN: So I will ask that Ian or staff capture this as one of our important parking lot gaps.

Are we ready to vote? Okay, so how many believe the scientific acceptability is completely?

(Show of hands)

DR. WINKLER: Zero.

CO-CHAIR SUSMAN: Partially.

(Show of hands)

DR. WINKLER: Zero.

CO-CHAIR SUSMAN: Minimally.

(Show of hands)

DR. WINKLER: Eight. I got eight.

CO-CHAIR SUSMAN: And how about not
at all?

(Show of hands)

DR. WINKLER: Ten.

CO-CHAIR SUSMAN: All right, our addition is correct.

DR. PINCUS: And I'm saying, how do we think about -- this wasn't submitted as a population-based measure, but does it require that there be -- that they submit it in some ways? I can imagine this as a population-based measure.

DR. WINKLER: And that might be something that you would want to couch in the recommendation of the measures needed that haven't come through. But we are certainly not excluding population-based measures, because particularly these low-incidence measures, patient safety measures, they are difficult to handle because they're low frequency, so there are issues around that. But if perhaps you are talking about, thinking about the integration of mental health
services in your community, perhaps a
population-based measure would be more
appropriate to capture, especially some of
these low-frequency things. So we can put
that as part of the recommendation.

CO-CHAIR SUSMAN: If we could sort
of flag that. So that is additional cars in
the parking lot. Let's talk about usability.
I think we had some implications about
usability from your prior comments.

MS. JAFFE: I think there are a
couple of things about usability. Number one
is, so much of it is not done, it's hard to
know how it would be used if it were done. I
think the expectation that patients are
contacted at 30 days and after three attempts
you don't try any more puts a lot of questions
into its usability.

DR. STREIM: As with all
suicidology, as I was saying before,
ascertainment for the numerator is the most
challenging thing in that whole field, and
this doesn't really propose a method for getting at that and a remedy. Not that it would be easy, but it is not even attempted here.

CO-CHAIR SUSMAN: Glen, any additional comments?

DR. PHILLIPS: No, I'm fine.

DR. MANTON: All of 3A is to be determined. Or not applicable.

CO-CHAIR SUSMAN: Any thoughts on harmonization here?

DR. WINKLER: I don't think there are really any other measures that harmonization really applies to.

DR. STREIM: You raised the point, Reva, about measures from other fields where they look at post-discharge mortality, and I don't know whether any of those would be relevant, but --

DR. WINKLER: The 30 days, I think, is arbitrary for those, but at least they have all picked 30 days. I can see where you might
argue a different timeframe, if you have -- do we know that the suicide rate post-discharge is, going on a time line, where is the peak in incidences or not, and frame your measure based on data to say what the appropriate interval for surveillance is. So I don't know that you should be wedded to 30 days, but I think it might be nice to see what the data might show would be a good interval.

MR. PELLETIER: I'm pretty sure that at least in hospitals and under the joint commission that if someone suicides within three days of discharge that is a sentinel event. And just for context, suicide risk assessment is something that the Joint Commission is focusing on. It's a new national safety goal both in psychiatric settings and in non-psychiatric settings, so people are really at this point putting together their risk assessments, and those of course are not standardized at all.

DR. HENNESSEY: And looking at
Google I am seeing a lot of one-year posts popping up.

CO-CHAIR SUSMAN: I think again there is a sentiment that this is headed in the right direction but perhaps not ready for prime time. Other comments about usability? Are we ready to vote about usability?

Okay, how many completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially?

(Show of hands)

CO-CHAIR SUSMAN: Minimally?

MR. CORBRIDGE: Five.

DR. WINKLER: I can't tell.

CO-CHAIR SUSMAN: Can we please, minimally?

DR. WINKLER: Five.

CO-CHAIR SUSMAN: Okay, not at all?

(Show of hands)

MR. CORBRIDGE: Twelve.

DR. WINKLER: Yes. Did we lose somebody?
CO-CHAIR SUSMAN: Oh, okay, Eric is out.

Let's go to feasibility. I think we have already alluded to some of the feasibility issues here. Group, thoughts further?

DR. PHILLIPS: Getting this data from most facilities I think would be impossible. So being from the Midwest, large state hospitals that serve half a state, how are they ever going to track this across those patients when they send them back out to the community? I mean, it's unusable, I think, for many of the facilities.

DR. MANTON: I guess the only thing would be, because I think the phone contact is unlikely to work and I don't know if they have a lot of time to do it. They could look at death registries or something like that. But I think that would probably be about the only way they could do it.

DR. STREIM: I think we can say
it's not a byproduct of care.

CO-CHAIR SUSMAN: Other thoughts around the exclusions, inaccuracies, implementation? Was there any data?

Okay, I am hearing a theme here.

Any other comments before we vote?

CO-CHAIR LEDDY: It seems like on death registries it wouldn't be that hard to do. Like in Medicaid, that's how we take our enrollment accurately is using death registries, and most states find it pretty easy to do.

MS. JAFFE: Actually we have looked at death registries and looked at suicide. It is not that easy to do because it doesn't always come across as a suicide.

CO-CHAIR LEDDY: Right, okay.

CO-CHAIR SUSMAN: Okay, so let's take a vote then on feasibility, then.

Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially?
(Show of hands)

CO-CHAIR SUSMAN: Minimally?

(Show of hands)

CO-CHAIR SUSMAN: Not at all?

(Show of hands)

MR. CORBRIDGE: Seventeen.

CO-CHAIR SUSMAN: So we are going to vote to recommend this for adoption. All those in favor of recommending this measure for adoption please say yes, raise by hand.

(Show of hands)

CO-CHAIR SUSMAN: Thank you. And how many nos?

(Show of hands)

CO-CHAIR SUSMAN: Anybody abstaining?

Okay, so all nos. All right, thank you.

Okay, so we are moving on to 30-day readmissions. I'll give people a chance to get to this.

DR. GOLDEN: Let me ask a question,
before you do that. You have several readmission measures, and before we do each one you may want to prioritize which one you want to do, do you want to do all of them? Or do you want to decide seven versus 30? That might save you some time and energy.

CO-CHAIR SUSMAN: Let me ask the group who actually considered these. We do indeed have three readmission measures, 30-day, seven-day, 48 hours.

DR. PHILLIPS: They're essentially identical proposals with different timeframes, and they're all as poorly put together.

CO-CHAIR SUSMAN: So I'm hearing a telegraph about where we might be headed with these, but is there any merit to discussing the timeframe up front in your mind, or will that just keep us from an inevitable decision?

DR. STREIM: No, I think probably not. If we just go through the first one I think that will get us through the next two quickly.
CO-CHAIR SUSMAN: Okay, I'm going to then --

DR. GOLDBERG: Well, I'd like to say, the seven-day one, we're being asked to report on that by somebody. All our payers are asking us to report on seven-day readmissions, and feeding that back to us and giving us regional norms comparing how we are doing.

DR. STREIM: I think that it's an important issue in terms of what timeframe would you look at, but the problem here lies with the measure itself and the way it's been proposed, and so if we want to just address what was submitted we will be more efficient. I think it's not that the timeframe is irrelevant; it's very relevant. But in terms of what is going to probably kill these it's other issues.

CO-CHAIR SUSMAN: So I would assume that this is indeed an outcome measure worthy of our attention. Why don't we turn to then
importance, and get the thoughts of the group.

This is the 30 days of discharge. Do you want to provide us the overview, Ian?

MEASURE OT3-003: 30-DAY READMISSIONS

MR. CORBRIDGE: Yes, just to bring people up to where we are. So we are looking at number three, 30-day readmission. This was submitted by Presby Shadyside. Description as stands, percent of patients readmitted within 30 days of discharge reported as percent of discharge for an inpatient psychiatric hospital or unit. The patient is admitted to the hospital within 30 days after being discharged from an earlier hospital stay.

The numerator statement reads as:
total number of patients readmitted within 30 days of discharge. The denominator statement:
total number of hospital discharges.

DR. HENNESSEY: So we are not looking at a patient who discharges and then readmits at another facility? Is that correct?
DR. PHILLIPS: Correct.

DR. PINCUS: Are there existing NQF measures on readmission that generic? Or are they all condition-specific?

DR. WINKLER: They are condition-specific in terms of capturing the denominator. They are all causes of readmission but they are for patients with an AMI, for patients with history of heart failure, whatever.

DR. PINCUS: And I guess, this comes up in the context of harmonization, but I think just going into this, is there a typical or standardized way by which those numerators and denominators are defined? And to what extent?

MS. BOSSLEY: These are the same measure developers, so I would assume so. We'd have to go back and look, to be sure.

DR. WINKLER: Most --

DR. PINCUS: I don't think so, that we've had it, for AMI. This is UPMC.
MS. BOSSLEY:    For the other ones that are endorsed, though, it's all the same developer.

DR. PINCUS:   For AMI?

DR. WINKLER:   For AMI and -- no. Not the same as for here, but the same ones, the ones that are endorsed, are all the same. So they are all specified very similarly.

DR. PINCUS:   OK, so we know the extent to which this one is like those?

DR. WINKLER:   I don't think we've done that in that great detail yet.

DR. PINCUS:   It ought to be from the point of view of general hospitals.

CO-CHAIR SUSMAN:   So I am hearing some interest, at least as a parking lot issue, to provide that sort of feedback. Okay, any other questions about the specification of this measure itself, or understanding the measure? Yes, George?

DR. WAN:   I know that there was a summary in the packet of materials, but I just
want to have that discussion on how this
particular measure compares with others, in
particular the NCQA, was it the HEDIS
measures, right? They have, they assess
readmissions after the 30-day window as well.

DR. HENNESSEY: Do they still do
that? Or did they stop doing that? I thought
that was archived. My impression was that
they determined that it did not have validity,
from a patient outcomes perspective, and so
they had archived it.

CO-CHAIR SUSMAN: So there is a
question of fact here, and there is a thought
that this might be an archived measure for
NCQA.

CO-CHAIR LEDDY: That is what is
so different about this one. There's no
database, you can't -- like, I've looked at
30-day readmission from a public reporting
point of view, and the issue is, if you are a
payer, such as Medicare, on Medicare Compare,
they have 30-day readmission. And you could
link it to diagnosis, if you wanted to, let's say. But that is only for Medicare patients because they have the claims database. Payers can do this, because they have their own claims database. So they can link it and they could say, for psychiatric as the primary or secondary diagnosis on the discharge. But for the whole population there is no database. The required hospital discharge databases in each state that are aggregated at the national level do not have unique identifiers, so a hospital can't see who is admitted to another hospital. There is no database.

DR. GOLDEN: But wait a minute, though. If Blue Cross of Alabama said we are going to, for our Blue Cross patients measure this, would that be okay?

CO-CHAIR LEDDY: Yes.

DR. GOLDEN: So then this is an acceptable measure scientifically?

CO-CHAIR LEDDY: This is across all populations, isn't it? All discharges?
DR. GOLDEN: We are talking now, let's go back to the earlier measures, this would be implemented by one payer, or by one enterprise. This would be fine, and you could do it.

CO-CHAIR LEDDY: Okay, then you could do it. You could do it by payer, or by provider.

DR. HENNESSEY: Yes, I think the big issue is that the way this is written right now, if you are a payer, or rather, if you are a provider, you are not counting someone who gets admitted to another facility. As a payer --

DR. GOLDEN: But somebody else will get you the data. They can count it for you.

DR. STREIM: The back story is --

DR. HENNESSEY: That requires a level of coordination.

DR. GOLDEN: No, they'll send you the reports, easily, that's an accountability measure, that's what it's all about.
DR. STREIM: Actually, it was informative to read further on down, in the submission, the reason they actually give for the fact that they don't - they thought risk adjustment here is not applicable, and the reason they thought that was because they only see this as a health resource utilization measure. So they use it - that is how this health system uses this information within system, and that is how they are coming at the measure.

DR. PHILLIPS: And so I think part of what -- the discussion I think is, we're drifting between, the idea of measuring this is probably a good idea. Measuring it the way they do, not. And so that is what I'm more saying is, if we stick to the proposal, even under the reason they don't defend it well. If you didn't know anything and you read this, you would say, oh, we shouldn't do this.

CO-CHAIR SUSMAN: I'm going to take Bill's comment, and then I'm going to get us
back to focusing first of all on importance, and going through. I think the comments that are coming out certainly are going to be important to consider as we work at this measure.

WPI REPRESENTATIVE: Are we still talking about importance, or where are we at?

CO-CHAIR SUSMAN: Well, I'm going to bring us back to importance, the focus. We had started out rather broad across the field. But I think it is all going to be relevant to our discussion in coming to a conclusion about the focus.

DR. GOLDEN: I will make my comments later.

CO-CHAIR SUSMAN: Okay. So let's start with importance. The impact, it looks like people felt were fairly completely -- is there comment from the person who said minimally, or some revised thought about that?

How about a gap?

DR. GOLDEN: That was my question.
You know we talked about 30 versus 7, and all this, but I'll ask the psychiatric practitioners here, is there an issue if somebody gets rapidly readmitted after a hospitalization that they may have been discharged either too soon or they had inadequate care or something?

DR. STREIM: Sure. I think that's what makes it highly important to measure, and that's highly relevant in that way.

DR. GOLDEN: So there could be differences between providers?

DR. STREIM: Right, but as this measure was submitted from a single health system, they haven't addressed comparability across health systems or providers, so there is no - they haven't really helped us look at that gap. We don't know how much variability there is, so we don't have that from the submission anyway.

DR. GOLDEN: But as a practitioner you would assume or you would say there would
be differences or potential differences between providers?

DR. PINCUS: Absolutely, I know something about it, it's --

DR. GOLDEN: All right.

DR. PINCUS: Actually, now you're talking about the development of this measure, this was developed as kind of a pilot program to incentivize reducing readmissions. And so that that is actually how this evolved. You know, reducing readmissions within their system, because they also, they have a closely affiliated payer as well as a health provider.

DR. HENNESSEY: I find this to be a somewhat troubling metric because of the timeframe which is only 30 days, and also because one can only relate the measure if you are being readmitted into your facility. I will tell you as a payor, I have actually developed a metric like this in the past, but it was measuring community tenure, and it was presence in the community and it was over a
one-year period of time, which to me is far more meaningful than what this is.

DR. GOLDEN: I'm sorry, but that's just not what the measure is. The measure does not measure you within your facility. If you get readmitted, you're readmitted. And that would not be necessarily facility-specific.

DR. STREIM: From my read of the submission it looks like the rationale for this, it was Pittsburgh that developed the measure was to be able to monitor the rate of service utilization and think about improvements in care to reduce that rate. But it was really a measure of the rate of service utilization, and therefore there was not a lot of interest in doing validation studies and other things that might not apply in that sense. But Harold was probably there when it happened.

DR. PINCUS: Just to say something about, you know, it depends on the focus for
NQF in terms of how this gets used. So if you are talking about having a measure out there that is sort of a handy-dandy easy-to-use measure for a facility, an inpatient facility, to assess itself, using its own data set, without having to rely on external sources of data, this could be a measure that might have some utility. On the other hand, it's not as good as the measure that would capture all admissions across, for an individual patient.

CO-CHAIR SUSMAN: So for quality improvement purposes, is that what you --

DR. GOLDEN: I am sorry, I'm looking at the numerator, it says, people readmitted. It doesn't say readmitted to the same hospital.

MS. BOSSLEY: Also if you look at that also, underneath it says, transferred to another hospital or setting for specific care who then returns would not count as a readmission.

DR. GOLDEN: Correct.
MS. BOSSLEY: So anyone transferred from another one and then comes back to a facility doesn't count.

DR. GOLDEN: That is just a transfer.

MS. BOSSLEY: There are no other exclusions, and it's not clear where they pull the data source from, it's management data. I think we'd have to go back and ask them to clarify what source of data it's from.

DR. ROCA: And this may be partly, and other people may know the Pittsburgh situation better than I do, but I think that is a very large system, and they may have a pretty good handle on who has been readmitted in that whole market, just through the Pittsburgh system. Joe, do you know, or have you looked at this, did they look at clinician-level readmission rates? Because I'm thinking this may have been --

DR. STREIM: That is not proposed as part of the measure at the individual
provider level. Whether they did that on the side isn't clear, but in terms of this proposal that we received it is not addressed.

CO-CHAIR SUSMAN: So let's focus on importance. I think again we've looked at a bunch of related issues, relationship to outcomes, gap, impact. Any further comments in that arena or relevant questions to those?

DR. PHILLIPS: So, again, part of the gap is a good example of one of my problems with this proposal, in that they don't bother to cite the literature around this that is out there. You know there are differences between, and there is a literature around that, that different providers, different places, have these kinds of differences, and they simply don't cite it. It's a very incomplete proposal.

DR. STREIM: And that may reflect the burden of the NQF process on would-be stewards, and they wanted to get the quick and dirty submission in in the timeframe. But I
think it doesn't mean that, again, that there
is not evidence of variability that makes this
an important thing to measure. I think one
question again for NQF staff is when we vote
on importance to measure we have to
distinguish, are we voting on the concept of
the importance to measure readmission rates,
or are we voting on the importance to use this
particular measure to get at it. Because if
the latter - no, not the latter.

DR. WINKLER: It's the former, it's
the concept of a 30-day readmission for
patients.

DR. STREIM: It's not about the
method. Okay. Because in this particular
case I think as we get further along here,
since I think we will see it's probably
important to measure, is that there is no
provision to measure readmission outside of
this health system, so if somebody goes to the
community hospital that is not part of the
health system three days after discharge, that
is not captured. So it's only capturing within-system utilization.

DR. GOLDBERG: Wouldn't it come up as a later issue, if Reva says? We're really voting, if it's importance, about the generic concept.

DR. STREIM: Right, and Harold's point, I think, is a good one, that even if it has utility for an individual payor and an individual health system, just because it doesn't generalize to the rest of the world, the health system - well, we don't have a health system at large - but if we did the failure to generalize to all hospitals, all payors doesn't mean it's not a useful measure that could be adopted by an individual hospital or health system for their own purpose.

DR. WINKLER: However, remember one of the basics for NQF in endorsement of measures is sort of an overlying criteria that these measures are suitable for public
reporting and accountability, and they are not simply for quality improvement, internal quality improvement kind of thing. And there are lots and lots of those measures, which is pretty much what Harold was describing. That's not what we are looking for. We are looking for something a little more than that.

DR. HENNESSEY: As a general comment, there are a number of these measures that are just that, they are probably good for a system from a QI perspective but whether or not they can really generalize over national exposure is very questionable.

DR. STREIM: Move to call the question.

DR. PINCUS: Just a clarification. When we decide about impact gaps, is it based on what they put into their proposal, or is based on what we know?

DR. WINKLER: Both.

DR. PINCUS: Okay.
DR. WINKLER: Both. I mean that's why -- we don't have a bunch of pediatricians sitting here looking at these measures.

CO-CHAIR SUSMAN: So I am generally hearing a sense that this is important, that there is a gap, that it may not have been documented as well, there are some questions about suited this particular measure might be that are going to come up perhaps under the other metrics that we are going to work at.

Is there anything new to discuss on this topic? Why don't we go ahead and vote? Importance, completely?

(Show of hands)

DR. WINKLER: Ten.

CO-CHAIR SUSMAN: Okay, 10.

Partially?

(Show of hands)

DR. WINKLER: Seven.

CO-CHAIR SUSMAN: Minimally.

(Show of hands)

CO-CHAIR SUSMAN: Not at all.
(Show of hands)

CO-CHAIR SUSMAN: Somebody out?

Okay, completely again, please. I'm sorry. Completely.

MR. CORBRIDGE: I got 12 now.

DR. WINKLER: I got 12 too.


So 12 and six it is, that's 18.

Let's move forward. You don't want to learn much about this process.

Okay, scientific acceptability, I've heard a lot of qualms in this realm, perhaps -- we're on this measure now. It's not the global importance, not the concept, it's this measure.

CO-CHAIR LEDDY: So for a health system, a 30-day readmission rate. This is just within a health system. Because otherwise 30-day readmission rate is really used a lot.

DR. STREIM: I think that is the
main limitation, and again it's not that – it only is designed to measure utilization rates within the health system.

DR. PHILLIPS: It says it later in the proposal. It very clearly says, a gap is we missed admissions to other hospitals within the proposal.

CO-CHAIR SUSMAN: I think we are trying to redo the measure for them. And I don't think we have the time and resources to do that.

DR. PINCUS: I just want to clarify exactly what's in there, because there's a discrepancy --

DR. STREIM: But just to summarize a few of the other points about the scientific properties and acceptability, the measure developers stated explicitly that there was no need for a validation, again, because they are using it to determine a rate of service utilization, and the second thing is really they didn't think risk adjustment was
necessary beyond — they said we sometimes, depending on our internal needs, adjust for age, gender, zip code and diagnosis, but there is nothing about disease severity, case mix, et cetera. So there is -- the kinds of risk adjustment that you would want for a public measure to make it really interpretable isn't part of this internally used measure. Those were the main points I would make about the science.

CO-CHAIR SUSMAN: Any questions about the science or additional comments from the group?

(No response)

Hearing none, let's go ahead and vote on scientific acceptability. Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

WPI REPRESENTATIVE: There is substantial evidence that this is a good measure but not as they define it.
DR. WINKLER: No, no. Scientific acceptability applies to this measure, as specified, as written, in this piece of paper.

DR. STREIM: Unlike importance which is the concept.

CO-CHAIR SUSMAN: Okay, partially again, please, just to make sure we have the count. Please raise your hands high.

(Show of hands)

MR. CORBRIDGE: Four.

CO-CHAIR SUSMAN: Okay, minimally.

(Show of hands)

CO-CHAIR SUSMAN: None at all?

Okay, one. Good, thank you.

So we are okay with that, let's move on. This is usability. It looks like the spread here in understandable harmonization and added value. Comments from the group?

Questions from the committee?

DR. STREIM: I guess we should make some comments here. Well, it's all written up there, but for those who haven't been able to
read the small font as it's projected, I thought one of the main concerns was the understandability or meaningfulness of the actual measures was pretty much anecdotal. What they do is have monthly meetings and focus groups which can be useful for these sorts of things. But it was really more our own experiences, it works for us. And, again, there was nothing to convince me that this was going to generalize to the wider group of healthcare providers, whether others would find it useful as defined. And I think if I were speaking for my own health system I would be concerned about the lack of risk adjustment in there.

DR. PHILLIPS: Right, and about the risk adjustment, the other measures that they cite actually do use risk adjustment, so it's not really lining up with the way some of the other things are being measured.

CO-CHAIR SUSMAN: Okay, so from an added value perspective I'm hearing maybe that
there doesn't seem to be as much added value as suggested by the ratings.

MS. JAFFE: Well, I think, at least when I scored it, it added value to the system, but I don't know if it's added value for the world. I think that, when I was reading it, it was very clear that they have a process that works well for their system, but to me they hadn't put a lot of thought into beyond their system and how this outcome could impact beyond their borders.

CO-CHAIR SUSMAN: So at least the definition says, review of existing endorsed measures, measure sets demonstrate the measure provides a distinctive or additive value to existing NQF-endorsed measures.

DR. PHILLIPS: And because there is not one for this population, I would say it is added value. But not --

DR. STREIM: That was my rationale for rating it completely, because if you measure anything related to readmission it's
better than nothing, but if you can't
interpret it maybe it's not.

CO-CHAIR LEDDY: How about if this
measure was available for - in the same format
for each of the health systems in a large
area? In a region, let's say, or a state.
Then will it have value?

DR. STREIM: I would say yes.

CO-CHAIR LEDDY: In that the only
thing it would be missing is people going from
one to the other, which when I looked at it
for medical and psychiatric together it's
about 20 percent.

DR. MANTON: Usability comes into
that. I'm not sure they could really do
that.

MS. JAFFE: Are you saying that
they'd get together and they'd kind of compare
who got admitted?

CO-CHAIR LEDDY: No, no, no, I'm
saying that's impossible. I'm saying that if
you have four health systems in a large
region, each of the health systems did this for themselves, then even though they were all missing that, say, 20 percent that are going across, you are measuring apples-to-apples readmission rates to their own facilities, and since readmission rates are going to be really the up and coming thing in health care reform with accountable care organizations, et cetera, and it is already measured for Medicare populations, that that could, I would say, make it usable, if you did it hospital by hospital or health system by health system, so that they are comparing themselves to each other, and the noise of people going to different places is just, they just can't deal with it, so you exclude it for all the measurements.

DR. PHILLIPS: But all of those hospitals would run some form of risk adjustment, because they are all going to be serving different populations, and this does not account for that at all. So I would say
the way they've done it, no. I mean, you would still have to account for that.

CO-CHAIR LEDDY: Hospitals have risk adjustments.

DR. PHILLIPS: This measure doesn't.

CO-CHAIR LEDDY: No, this measure doesn't, you're right.

DR. GOLDEN: I would say - I would put this in the parking lot, but you're still focusing on the system. There are already measures in place for readmissions for after pneumonia and heart attacks.

But it doesn't matter if it's not through your institution. It's in the institution, and they collect the data, and they can do that for Blue -- any insurer could track the readmission rates. So it doesn't matter. My academic center would be in a little bit of difficulty because a lot of their discharges get readmitted elsewhere in the community and that is going to count against them. So that
is still fair game.

   DR. STREIM: So based on what Bill is saying is I think you have a better measure coming out of a payor for something like this than -- I think payors are in the best position to get at this.

   CO-CHAIR SUSMAN: So I'm hearing some consistency of thought here. Are there any other additional comments on usability?

Let's go ahead and vote.

   Completely?

   (Show of hands)

   CO-CHAIR SUSMAN: Partially?

   (Show of hands)

   MR. CORBRIDGE: I got nine.

   CO-CHAIR SUSMAN: How about minimally?

   (Show of hands)

   CO-CHAIR SUSMAN: Okay, that should be it. Let's go down to feasibility. It looks like a relatively high feasibility score. Comments from the group, and then
what's in the minuscule type.

MS. JAFFE: I think that we need -- feasibility, when I was thinking about it is, feasibility for a particular system to do it for themselves, and it's not feasibility as we've sort of talked about it through the course of this conversation.

DR. HENNESSEY: Are they defining readmission as readmission to a psychiatric unit, or can it be readmission to the hospital at large?

MS. JAFFE: You know, they don't really say that in their submittal, but this is a psychiatric hospital, that's all they do, so that was one of my comments too. When they were talking about -- I made the assumption that it was psychiatric.

DR. MANTON: There are places earlier that they refer to psychiatric patients, I forget which category it is.

CO-CHAIR SUSMAN: Psychiatric hospital or psychiatric patients?
DR. MANTON: I just can't remember which one. It might have been under number one, but there was some place that they indicated it was psychiatric patients and psychiatric readmissions.

CO-CHAIR SUSMAN: So at least in summary a psychiatric hospital or unit.

DR. HENNESSEY: So concretely, I'm a suicidal patient, I leave the hospital, I then inflict a gunshot wound and I'm now in ICU for my gunshot wound, it wouldn't be reflected.

CO-CHAIR LEDDY: According to the summary it would be reflected, because it's discharges from the psychiatric hospital or unit and the patient is readmitted to the hospital. It doesn't say to the unit, at least in the summary. But I was not on the workgroup. Maybe it specifies it more.

DR. PINCUS: I just think it's worth pointing out to put this into context that the current NQF approved readmission
measure for other conditions is all cause
readmission. So that if you treated somebody
with an acute MI and then, you know, two weeks
later they get hit by a bus and come to the
hospital then that gets counted.

MS. JAFFE: And actually looking
back on my comments, in the denominator and
numerator, it just says, all patients, so that
was one of my questions. It didn't say
psychiatric patients or what they were talking
about.

CO-CHAIR SUSMAN: So it sounds like
there are some issues perhaps of the title of
the measure and maybe the specifications maybe
not quite lining up. Other feasibility,
though, reflections?

DR. PINCUS: Caution is only if you
are a system, in this?

CO-CHAIR SUSMAN: It will be what
it is.

DR. WINKLER: It doesn't sound like
you are going to recommend it, so I don't
think we need to worry yet about that.

DR. PINCUS: There is a kind of inverse relationship between feasibility and some of the other criteria. Because this actually is very feasible if you are doing it all within your own database.

DR. MANTON: That is what I was thinking, the data is there, it's accessible.

DR. GOLDBERG: But, for people on that workgroup, did they specify that this was a measure for a health care system? They didn't propose this to be more broadly used?

MS. JAFFE: They talked about straight from the hospital and readmission back to the hospital.

DR. MANTON: But for instance, when it talks about use in public reporting initiative it talks about, within our multi-system -- multi-hospital system this measure will blah blah blah. I mean, throughout, they tend to make references to within their system.
DR. STREIM: They made it clear, that - they made it clear that all this was designed and used in their system, tested in their system, they didn't really address how it would translate into other --

DR. WINKLER: Well, they did, they actually did. There is a section, question on level of measurement or analysis. It's right at the end of the specifications section. And they said facility or agency or multi-site corporate chain. So they really are talking about something that's -- But it's not individual providers.

CO-CHAIR LEDDY: Because that is the data they have.

DR. WINKLER: Right, correct. But not individual providers sort of thing.

CO-CHAIR LEDDY: Because that is the data they have.

DR. GOLDBERG: So a facility-only issue has feasibility problems.

DR. PINCUS: I find I am confused
by this discussion, and I think part of the
problem is, is this truly intended to be only
all-cause admissions to your facility? Or is
this clinicians' readmissions across whatever
we find for the broader database?

So it seems to me if it's only
within your facility then it's - the
feasibility is high, but the utility is lower.
On the other hand, if it's all sources, all
places of readmission, then it's feasible for
a payer but not for a facility.

DR. PHILLIPS: So if I may in
Section 4(d)(1) they specifically say, also
important to note the possibility that some
patients are or would be readmitted to a
different hospital and/or facility. As a
result the figures for a given
facility/operation would come with the caveat
that it may not be the true total figure for
the facility.

DR. PINCUS: That is something
worth noting. But when they specify the
numerator and denominator, who do they -

     DR. PHILLIPS: They don't talk
about it, and I noted it that it was
specifically an issue that they didn't talk
about it.

     DR. PINCUS: Is there a way that we
can interact with them to know exactly what
they are talking about?

     CO-CHAIR SUSMAN: I think what we
have here today is the data they provided is
from a health system or hospital perspective,
in a single entity, and we have to really vote
on what we have before us. I'm sure Ian and
staff did the best they could to clarify the
issues here and I think we should judge it on
what's been submitted.

     DR. PINCUS: One thing that we did
with the medication management measures
steering committee is that we were
disappointed in a lot of what we got, I think
I mentioned this at the last meeting. And so
what we did was, we sort of did not approve
things or had sort of a - did not approve
things, but pending further discussions, might
approve it if the measure developer was
willing to make some changes. And is that
something that we can do now? So if they
clarify that the intention is that they would
have it be applicable for a payer.

CO-CHAIR LEDDY: They couldn't
maintain it. It would have to be a different
submitter.

DR. PINCUS: Why?

CO-CHAIR LEDDY: To me this is
completely logical, what's happening. This is
a health system. If a health system wants to
do internal monitoring of themselves on how
they are doing.

DR. PINCUS: No, no, I'm saying
that a health system can propose anything they
want. I mean a health system -

CO-CHAIR LEDDY: But they have to
be able to do what NQF wants them to do,
right?
DR. PINCUS: Right, if I have my own little corporation I can propose anything I want, and if I'm willing to do whatever the stewardship requires -

CO-CHAIR LEDDY: Maintain the measure. They can't maintain the measure because they don't have the data.

DR. PINCUS: Well, how do you know? You can't say they don't, because in fact they do. They own a major payer.

CO-CHAIR SUSMAN: Okay, Reva.

DR. WINKLER: Yes, certainly there are times when discussions with the measure steward, there are suggestions that a steering committee will make, that they are amenable to making changes, that your approval is conditional on them making that change. So that is possible. However I would caution you, one, with outcome measures, that's hard to do; you don't turn those on a dime, so you don't tweak around the edges very readily on outcome measures as you might on certain
process measures. And two, the degree - one of the reasons our measure developers have been provided to participate, and I don't known if the fact that they are not on the phone is causing us a problem because they are not participating.

CO-CHAIR SUSMAN: Oh, nobody is on the phone?

DR. WINKLER: Anne?

MR. CORBRIDGE: I will ask.

(Simultaneous speaking)

MR. CORBRIDGE: So I guess we will ask again if one of the measure developers is on the phone? Because I know I had talked to them and they were planning on it. I know we have had some -

(Re-establishing telephone connection)

CO-CHAIR SUSMAN: Okay, so I think we are actually on your measure currently, which is a readmission measure, and I think there are some questions that people might
have. Let me ask the group if there are some specific questions for the measure developer.

DR. PINCUS: I thought we had a question about the specific of the numerator with regard to whether the readmission had to be at the specific facility or is it from any facility within some sort of range of location.

CO-CHAIR SUSMAN: Did you hear that?

WPI REPRESENTATIVE: That is a good question, because that is internally based on what we are measuring ourselves. They are only able to see people who are readmitted to our facility because that's the data we have. And I'm expecting that that is what we are proposing as well. However on a much higher level if it's possible to see readmission across systems, that would be ideal.

CO-CHAIR SUSMAN: Thank you. Other follow-up?

DR. PINCUS: What exactly are you
proposing?

CO-CHAIR SUSMAN: The question is, what are you proposing?

DR. PINCUS: The question is, what are you proposing? Is it at a single hospital or health system or is it at a broader level?

WPI REPRESENTATIVE: I think in this case, it's the hospital or system.

CO-CHAIR SUSMAN: Okay, thank you.

DR. MANTON: And are the readmissions just psychiatric readmissions or any readmissions?

WPI REPRESENTATIVE: Psychiatric readmissions.

DR. STREIM: And is that determined from a hospital administrative database or do you have a payer database that you use for that?

WPI REPRESENTATIVE: Hospital administrative database.

CO-CHAIR SUSMAN: Okay, so I think we have better clarity about the measure and
the intent from the measure developer. Are there any other questions from the committee about this measure for the measure developer?

DR. STREIM: Yes, do you have access to a payer database to track readmissions and if so, do you see a way that you could use this measure more widely beyond your own system? Or to be able to test it beyond your own system?

WPI REPRESENTATIVE: Can you repeat that?

DR. STREIM: You said that you have obtained this data from your own hospital administrative database. What I'm asking is, do you have access to a payer database where you could get the same readmission information, not only for your own institution, but for other perhaps regional institutions, so that you could test this measure more widely?

WPI REPRESENTATIVE: Currently we do not have that information available to us
readily, and we are not measuring the exact level of readmission rate; we are currently just measuring the readmission within our system.

(Simultaneous speaking)

WPI REPRESENTATIVE: It might be a possibility if the payers are willing to pass that information along. This would have to go across multiple payers as well, so that is a future measure. Currently this is just within the hospital system.

CO-CHAIR SUSMAN: Okay, thank you very much. Let's turn back, then, to feasibility and see if there is any further comments. And if not, why don't we go ahead and vote. On feasibility completely.

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

DR. WINKLER: Nine.

CO-CHAIR SUSMAN: Minimally.

(Show of hands)
DR. WINKLER: Five.

CO-CHAIR SUSMAN: Not at all.

(Show of hands)

CO-CHAIR SUSMAN: And that gives us 15. Eric is gone.

MR. CORBRIDGE: I got 11 on the partially.

CO-CHAIR SUSMAN: Okay, so we've got the count correct. And let's move forward. Any final questions that the committee has for the measure developer or any final comments the measure developer would like to make prior to our vote? Or public comments?

(No response)

CO-CHAIR SUSMAN: Hearing none, let's go ahead and vote.

All those who would vote yes for the recommendation, please raise your hand.

(Show of hands)

CO-CHAIR SUSMAN: All those who vote no, please same sign.
(Show of hands)

CO-CHAIR SUSMAN: So the vote is 17 nos, zero yes. Thank you very much.

So let's go on to the next set, which I think will probably go a little bit quicker, given our conversation. And now we are at the seven-day readmission measure. Was this also submitted by Western?

MR. CORBRIDGE: Correct.

CO-CHAIR SUSMAN: Any additional comments you would like to provide from Western Psych? Please, Richard.

DR. GOLDBERG: As long as they are on the phone I'd like to hear their thoughts about the risk-adjustment efforts they made and why or why not they made those comments.

CO-CHAIR SUSMAN: Hello, folks at Western Psych. Are you still on? She hung up after the vote. She was down, suicidal. Have we done a care plan with her?

(Laughter)

CO-CHAIR SUSMAN: Okay, Dr.
Goldberg has a question for you.

Dr. Goldberg: Could you comment on what kind of thinking you did about risk or severity adjustment in relation to this measure and what you included in it, or what you didn't include?

WPI Representative: Currently we have - we are vetting various risk adjustment criteria. We are looking basically at severity by unit of - within the hospital, our different age groups. So we have not completed the risk adjustment process. We are doing it by trade-off currently.

Co-Chair Susman: Okay, so I hear that there is some risk adjustment activity in process, thank you. From the group that reviewed this, are there additional new comments or let's focus first on importance?

Dr. Streim: Actually, it would be helpful to me since I'm not an expert on all-cause readmissions and I know NQF has had experience with these, what is the current
thinking about the - this whole issue of risk adjustment for causality?

DR. WINKLER: I thought you were going to ask a different question.

DR. STREIM: You can answer the other one first.

DR. WINKLER: Okay, the concept around all-cause - because this discussion has been ongoing - a couple of things. The idea that you look at a patient's episode of care and services from their perspective, regardless of why a patient might be there, especially with multiple comorbidities and other things going on, that, to focus in on whatever is the primary reason for diagnosis and exclude all other things and let the diabetes become problematic and not be attended to during the course - or their depression not be attended during the course of their stay for heart failure or whatever else is not appropriate, and certainly a way we want to move to. So the idea is you really
do want to look at all aspects of a patient's care, and that any lack of attention to some of these other comorbidities might be the reason for their readmission, and that is a fair sort of thing.

Also what we've started having conversations about is when you start looking at a list of what is or isn't related, to the primary readmission, it becomes very different to sort them into black and white buckets. You might think that a patient is being discharged, and then you know has a car accident. But what if they had an arrhythmia episode as a result of a heart problem that causes them to be in the accident. So you can start having a real difficult time parsing those out. And so the all-cause - and realizing that that all-cause applies across the board to everyone, so there is going to be - you will never hit zero readmissions, but the idea is to reduce them to as low as possible. So that is the current sort of
dynamics of the discussion around the all-cause readmission concept.

DR. STREIM: I will ask my second question after.

DR. ROCA: To what extent, since these measures have been out there for awhile, have we actually found that hospitals or systems have been able to reduce their readmission rates?

DR. WINKLER: Considering it's one of the biggest focuses for quality improvement you are seeing a lot of particularly forward-thinking hospitals, but a lot of systems really trying to come up with some innovative ways of doing patient follow up, of facilitating that care transition, asking what is it that is important about it, to keep them from bouncing back into the hospital. So it actually is a huge focus right now and I think you are going to see in the literature reports that are demonstrating a whole variety of approaches that may be appropriate, which is
why then people say the outcome measure is
really the most useful tool, because however
you got there is fine as long as the
readmission itself is reduced. So that is
kind of the -

DR. ROCA: And are those data
appearing already? Have readmissions been
measured for awhile in this way?

DR. WINKLER: I don't think they've
been measured all that long. I think the
readmission rate has only been up for a year
maybe. So within the last year. So I don't
think we've got lots of longitudinal data yet,
but Medicare is the big push for this. But I
think we will shortly in a couple of years.
But there isn't a lot of longitudinal data
right yet.

CO-CHAIR LEDDY: But there are
some examples, not in mental health
specifically, but across - although there are
some evidence based practices that have been
found and replicated. So that is starting to
come out like in Colorado is one, mostly
around discharging care planning.

DR. ROCA: Certainly embedded in
this is the presumption that there has been
some failure leading to the readmission or a
quality problem leading to the readmission.
And certainly anecdotally you can discover
that in individual cases. But I'm wondering
if the data would bear that out.

DR. GOLDBERG: There was an article
in the New England Journal a few months ago
where the Congressional Budget Office reported
on what is likely to work to reduce costs. It
was a little unsettling, because they said
that electronic medical record, the primary
care medical home did not - it was hospital
readmissions they projected would only be of
the five or six items they reviewed, it was
only hospital readmissions that were likely to
reduce costs. It was surprising that some of
the other panaceas that we're holding up,
according to CBO.
CO-CHAIR SUSMAN: So PCMH rates could potentially - who knows. Eric?

DR. GOPLERUD: There is some old data and reports from the VA hospital that used the seven-day readmit, and looked at unforeseen consequences. And one of the things they found when they had that psychiatric-only readmit diagnosis is that you got diagnostic fiddling. And so what you had was they would get readmitted for a non-psych diagnosis, or when they had a seven-day readmit, they wouldn't readmit until after the seventh day. People were being kept in 22-hour holding, whole lot of things, because there were some real consequences for their incentive payments.

But so in support of what Reva was saying about all-cause readmissions, if you don't do it, you set it up for people to be diagnostic fiddling.

CO-CHAIR SUSMAN: Joel, did you have another question?
DR. STREIM: Yes, and again this is for Reva or anyone else who is the expert here. What do we know about the - I don't know - the validity of seven-day - 48-hour, seven-day, 30-day in terms of validity, content validity?

DR. WINKLER: To me, what I would say, and I am no expert on this, I think it would be dependent on the reason for the initial admission. And I would ask you all as the mental health experts what is it about that particular condition and the hospitalization which does or doesn't happen during that hospitalization and care transition that - what is the timeframe that would be the most useful for public reporting and pushing and improving quality. The arguments in favor for using more medical conditions like heart failure, AMI, those are sort of a traditional, everybody is comfortable with looking at what is going on for 30 days, but I'm not sure that is
necessarily applicable in the mental health field. I think some conditions might be different.

DR. STREIM: Yes, I think there is a lot of heterogeneity across conditions in terms of time to relapse, time to recurrence. Even if you look at, take a simple example like bipolar illness where you have recurrences that are part of the chronic illness, an expected part of the chronic illness. And some people cycle rapidly and some people cycle slowly. That is the intrinsic nature of the illness itself. The factors we are trying to get at with these measures had to do with how we provide care and how we can influence outcomes, and I think it's very hard to come up with a time interval that makes both clinical sense, but my question was really about what time interval makes sense in terms of quality measurement, and I don't know whether anyone has really been able to tease that apart. Again I don't
know that literature myself.

   DR. MANTON: I wonder if they have
looked at it, if she is still on the phone.

   CO-CHAIR SUSMAN: Is our measure
developer still on the phone at Western Psych?
There is a question here about the rationale
of 48-hour, seven-day, 30-day, and whether you
actually accumulated data that reflects these
readmission rates and how it might inform us
and sort of where the points of improvement
might be in the process.

   WPI REPRESENTATIVE: I don't have
that data available with me offhand, right
now, but we can get that to you.

   DR. STREIM: So are you saying that
you do have comparative data looking at the
readmission rates for 48 hours, seven days and
30 days?

   WPI REPRESENTATIVE: Yes, we do have
seven-day, 30-day, 48-hour readmission rate
data, but I don't have that number currently
with me.
DR. STREIM: Even if you don't have the numbers, can you tell us whether you think the differences are informative about which time interval is most helpful for measuring quality?

WPI REPRESENTATIVE: We believe that the shorter time interval is usually most indicative of the quality of service delivered as the hospital that is discharging, and as the time interval becomes larger and larger, less of the readmission rate can be attributed directed to the discharging hospital. We currently use this information as part of our report cards we do for physicians as an hospital-wide indicator.

CO-CHAIR SUSMAN: Okay, thank you. If there are no other general questions, why don't we go down the list here. This is on the seven-day readmission. We are looking at importance. How many believe completely on importance?

(Show of hands)
DR. WINKLER: Zero.

CO-CHAIR SUSMAN: How about partially?

(Show of hands)

DR. WINKLER: Eighteen. That looks like everybody.

CO-CHAIR SUSMAN: Okay, let's go down then to scientific acceptability. Any new or differing information from the comments of the past discussion?

DR. STREIM: I would just mention that the submissions for all three time intervals for measurement were identical except for the difference in 48, seven and 30.

CO-CHAIR SUSMAN: Okay, then.

DR. PINCUS: For all of these things we basically all agree that that our votes for all of them apply so we can move on.

CO-CHAIR SUSMAN: Thank you very much, Harold, for that suggestion.

Is it the wisdom of the group that we replicate our findings here, and perhaps we
can move to a vote so we have that formal.

I'm seeing a lot of head-nodding.

How many would vote in favor of recommending this measure for acceptance?

How many would vote against, let's see hands please.

(Show of hands)

CO-CHAIR SUSMAN: Eighteen. So the final count is eighteen against, zero for.

DR. PINCUS: Can I make a suggestion that there be interaction with the measure developers about potentially adapting this measure to respond to some of the concerns that we have.

CO-CHAIR SUSMAN: So I'm hearing that one of our parking lot issues, here, is that this general concept is obviously quite important and that perhaps encouraging the measure developer to do some further work would be very beneficial to the field.

DR. PINCUS: It strikes me as a natural thing. We told the Joint Commission
that we weren't going to approve it unless
they did X, and then they did X and we
approved it.

MS. BOSSLEY: Right, you could say
that you would like certain things completed
to these measures, and if those were met then
you would recommend it, and we can take that
to the developer and ask them. I think the
question is, you would have to go really
detailed and give them really explicit
information on this measure. I guess the
question is, for these three measures will you
be able to do that, and will they be able to
then respond back in the timeframe we have, or
is it too big.

DR. PINCUS: My question is - I'm
not sure. If they said that these measures
were to apply to all the readmissions whatever
reason, would that be acceptable?

DR. WINKLER: Some of those
questions I think we can get clarification on,
but one of the major things I heard from all
of you is the lack of risk adjustment as being
the sort of major downfall for these measures,
for this purpose, and that I don't think - I
think that is pretty big to try and get that
fixed too quickly.

   DR. MANTON: It also sounds like
they are working on it.

   DR. PHILLIPS: It almost sounds
like they just need to get farther along in
their development and come back to us.

   CO-CHAIR SUSMAN: I think again,
since you are, I assume, still on the phone
the general sense of the group is that this is
great work but there are some elements
including looking carefully at the
numerator/denominator specifications and the
risk adjustment process that could make this
a very viable measure.

   DR. STREIM: And the other factor
I would add to that list is the availability
of payer data so that you can look across
systems within a region.
CO-CHAIR SUSMAN: So now we are at the 48-hour again. Is it okay - same thing. I thought we'd have to for the safe, but if not, same vote? Okay. Fine.

Well, then I'm going to declare victory and ask if there is any NQF member or public comments?

(No response)

Hearing none, it looks to me like it is now 10 of 3:00. We are sort of ahead. Would it be the wisdom of the group to launch on to substance abuse or take a break? Short break. How about at three o'clock more or less. Thank you.

(Whereupon, the above-entitled matter went off the record at 2:50 p.m. and resumed at 3:04 p.m.)

CO-CHAIR SUSMAN: Tricia and I had this great plan that we were going to alternate facilitation but then we had the workgroup order changed, so you will have to put up with me through this next set of
measures. We will work until about quarter to five and do as many as we can with the first one up being substance abuse, patients, clinical status, recovery and substance abuse treatment.

SUBSTANCE ABUSE, PATIENTS CLINICAL STATUS, RECOVERY AND SUBSTANCE ABUSE TREATMENT

CO-CHAIR SUSMAN: And that group, if you were a member of that, myself, Eric, who else was a member of the workgroup?

DR. WINKLER: It was workgroup four.

CO-CHAIR SUSMAN: Okay, good, so we are on, and the first one we're going to be considering is the milestones of recovery scale.

MEASURE OT3-001: MILESTONES OF RECOVERY SCALE

CO-CHAIR SUSMAN: And I will ask Ian to provide a brief review of that.

MR. CORBRIDGE: So we are working right now on Measure #10: Milestone of
Recovery Scale. And, Heidi, I think is going
down to this at this point, so we'll be there
in a second.

Just a brief description of this
measure. The Milestone Recovery Scale is a
one-item self administered scale that
indicates when an individual is in the process
of recovery from a severe - and I'm sorry my -
does that cover it? I guess my page got
lost.

CO-CHAIR SUSMAN: Severe and
persistent mental illness, the scale is
designed for use with adults who have severe
or persistent mental illness, 18 years and
above, scale measures. We underlined
constructs, level of risk, level of
engagement, level of skills and supports,
combined to create the following eight
categories of extreme risk, high risk not
engaged, high risk engaged, poorly coping not
engaged, poorly coping engaged, coping,
rehabilitating, early recovery, advanced
recovery.

So that was the tag team there.

MR. CORBRIDGE: This is measure #10.

CO-CHAIR LEDDY: It's in a different order if you are looking at this packet. If you are looking at this packet, the decision table, it's in the second group because we decided on the phone it wasn't an outcome measure but we wanted to look at it anyway. So it's like on the fourth or fifth page.

CO-CHAIR SUSMAN: This is workgroup four, so you will find that a little further along if you are looking at these number of ratings.

MR. CORBRIDGE: I believe on the Word document that was sent out for what's being projected up there, I believe he said it was page 36, page 36 for those who are following.

CO-CHAIR SUSMAN: Thirty-four, 36,
35. I mean this is an inexact process.

    CO-CHAIR LEDDY: I have matched up the pages.

    CO-CHAIR SUSMAN: All right, so for those of us who have had an opportunity to look at this thoughts about whether, first of all, this was an outcomes measure or a process measure.

    DR. GOPLERUD: I was one of the publicly disappointed reviewers in that I did not think that this was an outcomes measure. It also really didn't show any change scores. It - most of the measure was not filled in, so it was very difficult to know what to make of this measure because they didn't essentially complete the form. But my sense was it was an interesting area, but we have no idea of reliability, validity, so it's an important issue. Is it an outcome measure? I don't think so.

    CO-CHAIR SUSMAN: Luc and Sheila.

    MR. PELLETIER: I would agree that
knowing where someone is in recovery is an important thing, but I would agree that there were not studies or evidence that the measure is effective for reporting outcomes.

DR. BOTTS: Same here.

DR. GOPLERUD: And also this is a staff reported measure without good anchors, and that has incredible demand characteristics.

CO-CHAIR SUSMAN: So the first step, and then I'll get to Harold's comment or question, is to decide whether this meets the scope or not. And I think we should clarify whether we believe we want to go through the process if we think it's in-scope, so why don't we take Harold and get back to that issue?

DR. PINCUS: So I come back to looking at the importance of scope, we are evaluating the measure or the concept, and so to try a potential understatement, what the concept is behind this. The concept of
measuring recovery seems to be an important concept, but I don't have a good idea of what the intent of this, what - how they kind of operationalize that concept in a meaningful way.

MS. WILKINS: I can respond only because I am somewhat familiar with the use of the tool in California. It's been pretty widely used in some really innovative and strong programs that are addressing many of the outcomes that, in our meeting last fall, we said we really wanted to be looking at. So even though I'm not in that group and didn't actually see what they submitted to us. I am somewhat familiar with the instrument and so I brought a copy of it. The way they look at poorly coping not engaged is, these are folks who - so they are towards the middle of this. It addresses their symptoms; they may have moderate to high symptom distress. They may use drugs or alcohol, which may be causing moderate but intermittent disruption. It
talks about their thinking, they may not think they have a mental illness, they are not participating voluntarily in ongoing mental health treatment. Some of the other measures then get into details like how often are they going to jail, are they in stable houses, so to the extent to which in our discussion of outcome measures last fall, we came up with this really big list of things like are people homeless, are they going to jail, are they managing their symptoms, are they functioning well bundled inside what looks like a really simple list here is a lot of detail about - detail meaning it won't fit on one page. But it's more than just what you see there.

CO-CHAIR SUSMAN: So apart from the issues of the usability, the psychometric properties and so on, I'm hearing that this is a multidimensional composite score which embodies many of the dimensions of outcomes that we talked about at our last meeting. And I wonder you guys in the group have had some
time to look at this, recognizing that many of
us aren't familiar with the instrument itself,
does that meet the scope criteria?

To me, it seems to.

DR. GOLDBERG: I wasn't in the
group. But I was one of the people - I saw
this as an outcomes measure from the
beginning. I can't speak to the science. I
know we'll have discussion of that. But there
are people with severe persistent mental
illness who it distorts or cuts across many
categories of where they live and level of
function and co-morbidities and psychiatric
symptoms. It kind of bundles all those in a
way that allows you to say, what's their
outcome at this point. I mean is their
outcome at this point any better. So I thought
it was on track in some way as a category, and
it seems to me that it is within scope, and
that we ought to discuss the other dimensions
of it.

CO-CHAIR SUSMAN: Okay.
DR. GOPLERUD: I think there are two parts of challenge to this. One is that we didn’t have the detail either; all we had were the eight descriptors. Second is that nobody submitted, say, the global functioning. Global functioning is used a lot. You get a gap score, but it’s a measure, it’s not an outcome, or you could use the basis, or you could use a whole lot of different measures. The measure itself is not an outcome; it’s the use of the measure in a context, either change score or — and so that’s where I had the difficulty with an outcome is it told us about a measure which seemed to have some difficulties, rather than its use in gauging outcome.

CO-CHAIR SUSMAN: So if we look at the underlying embodied behaviors that are in each of these categories, would going to jail a lot or being an abuser be patient-oriented outcomes that would matter? And I would submit they really are. Now it’s hard to know
that from the summary staging, but knowing the
underlying constructs I think it sort of right
within the scope of what we should be doing.
But again that's just one person's opinion.

MS. JAFFE: To me I think the
confusion was part of it maybe was the
author's interpretation of what NQF wanted
was that if the measure shows improvement over
one year using the milestone recovery scale
then that's an outcome. And I think implicit
in the use of this recovery scale - my guess -
is the author's assumption that the outcome
is that they are improving. But they are not
writing it that way. And so it's a little
confusing to me.

MR. PELLETIER: The other confusing
part for me was even in the introduction they
say, it only takes 15 seconds to do this. And
I'm like, not having seen the tool at all,
really, wow.

CO-CHAIR SUSMAN: Maybe they meant
15 hours.
MR. PELLETIER: Because there is a rich amount of information behind it, supposedly.

DR. STREIM: It's like if you are doing a clinical global impression of severity, it only takes 15 seconds to score it, but you know the patient's baseline, you know a lot of information.

DR. PINCUS: I mean it seems to me there is no question that is an outcome thing, and I think the gap is an outcomes measure. I mean it's not a good one. Anything, obviously, but the intent is, I mean clearly the intent is to do this.

DR. STREIM: Was there any attempt to define baseline?

DR. PINCUS: At least what they report here they have actually a fair amount - they don't give any citations but they do report a fair amount of research on this in terms of inter-reliability coefficient of .85, with test, retest reliability of .85, so they
have in - it was also strongly correlated with
the direction with the Multnomah Community
Ability Scale.

CO-CHAIR SUSMAN: I am going to get
Reva and then Bill.

DR. WINKLER: I just want to tell
you that Carol just pointed me in the
direction of where to find this document that
has all this information, and I'll be more
than happy to, when we're done here to go get
it and I'll send it out so everybody has it.
So that if you feel that you need that to go
get a good handle on this measure, we can go
get it for you.

DR. PINCUS: Although it doesn't
get the actual information about looking at
citations for it and actually how they
conducted those assessments.

DR. GOLDEN: One criteria for
assessing this measure which is not in your
master list is validation. I mean you have
basically a provider-generated measure, so the
person being evaluated is the person filling out the assessment. So if you start to go to an accountability measure, then it can be gamed, and the question is, how does an outside entity validate that the reporting is actually reflective of the care. I think that would be very tricky business, and could be an issue for this particular measure.

CO-CHAIR SUSMAN: So let me entertain a vote, if there are no other discussions of whether this is in scope or out of scope, because if it is out of scope then we needn't go further. If it is in scope then we need to do the rigorous work.

So would you please vote first if you believe that it is out of scope. Out of scope, a process measure not sufficiently linked to outcomes.

(Show of hands)

CO-CHAIR SUSMAN: I'm just going by the order up there.

How about in scope, raise your hand?
(Show of hands)

We are just trying to see if you are aware. Abstentions?

(Show of hands)

One, okay. So let's then go on and just go through our process and I think these other issues will probably come up.

First of all, importance to measure the report impact gap in relation to outcomes.

MR. PELLETIER: The same thing that I said before, that the concept is in alignment with the recovery model applied to mental health but we found no studies or evidence that the measure was effective.

It's an important concept.

CO-CHAIR SUSMAN: So remember this is more the importance of the concept of the dimensions being measured as opposed to the measure itself. So I would - when I looked at or now with the benefit of going through these, it seems to me like this is an important concept, that the recovery process,
recovery model as an outcome is pretty important and the patients value that and patient advocates value that highly.

CO-CHAIR LEDDY: So even though the title of the measure says, milestone of recovery scale, we are not voting on the scale itself?

DR. WINKLER: For the importance criteria, the question of measuring this using a tool, perhaps, this one or others that they happen to exist, is the concept of the measure, then you look at the specific characteristics of how the specs are for this particular measure.

DR. STREIM: So we are voting on milestones of recovery not with capital letters but with lower case?

DR. WINKLER: Absolutely.

CO-CHAIR SUSMAN: Joel, you have a wonderful way of distilling things down.

Sheila, any thoughts or comments from any of you?
DR. BOTTS: I thought it met it.

I will talk louder. I felt like it met this measure in terms of an impact and relationship to outcomes. I think some of the other discussion that comes up really comes up in terms of scientific acceptability.

CO-CHAIR SUSMAN: Luc, any further comments? Eric?

DR. GOPLERUD: I agree with Sheila completely.

CO-CHAIR SUSMAN: Are we ready to vote on importance then? Completely?

(Show of hands)

MR. CORBRIDGE: Thirteen.

CO-CHAIR SUSMAN: Partially.

(Show of hands)

MR. CORBRIDGE: Five.

CO-CHAIR SUSMAN: Okay, so we are done with that part. Now let's go on to scientific acceptability. I think this is an area where there probably is some more concerns, at least from my point of view.
DR. GOLDBERG: Based on the section and what they submitted, not this addendum but this one.

DR. WINKLER: We are on capital letters, right.

DR. GOLDBERG: So what is your guidance on that? Do we have to do more of this.

DR. BOTTS: Part of what was in the document were links to the PDF I think of the criteria that were passed around, but they weren't linkable in the PDF that we had, so the PDF is incorporated there, so I'm guessing that they were submitted, but when we reviewed them we didn't have access to them.

MS. BOSSLEY: What we can do is provide it to you, and then Ian, we are going to have them come out on another call again, most likely? You can discuss it then after you have time to review it. That's fine to table it now, if you like.

DR. BOTTS: I just wouldn't want
them to be penalized for us not reviewing what
they probably did submit.

DR. GOPLERUD: I think in this one
it would be useful for us to read the
numerator and denominator because it doesn't
come clearly in the description. The
numerator details is the sum of all clients
who have a higher MORS score at the end of a
specified time frame than they had at the
beginning of a time frame. And the
denominator is the number of all clients who
were given an admission MORS score at any time
during the specified time frame.

CO-CHAIR SUSMAN: So it is sort of
- imagine what you are going to measure at any
time and place, and we'll call it an outcome.

CO-CHAIR LEDDY: It was at
admission or at any time. Too bad it's not
at admission and another specified time.

DR. STREIM: So there is no
attention to speed of recovery, recovery
trajectory here. So if they come back two
weeks later and they get a MORS score and they are no better, that would be actually excluded from the numerator, right, because they are not improved.

DR. GOPLERUD: This comes from the Village, that's where it was developed, and these are the most severely mentally ill, severely mentally ill who are in prisons and jails. So they are really looking at probably a longer time frame of a year or a couple of years and it probably wouldn't say work for acute psychiatric.

DR. GOLDEN: Since we are on the scientific piece right now, it would strike me, people who looked at this, was there any statement about inter-observer reliability. I could see depending on who filled out the tool, there could be great variation.

(Simultaneous speaking)

DR. GOLDEN: And a 15-second assessment, that's interesting.

MR. CORBRIDGE: I'm sorry, just to
interject quickly, I know the measure
developer is on the line. He just sent me an
email. He's having a hard time hearing the
discussion. So if you are speaking just try
to make sure you use the mikes or something.

MR. PELLETIER: It was limited to a
regional sample. It's pretty much California
and they talked about working with someone in
Boston.

CO-CHAIR SUSMAN: So there is
discussion of the reliability testing, is it
primary and secondary rater blind to the other
raters, a total of 105 clients rated by two
individuals, test/retest reliability, two
points in time during a single month in
California, and 381 clients with the interval
ranging from 10 to 20 days. So there is
actually at least some inter-rater and
test/retest reliability, and the correlations
actually are pretty good. Inter-rater
reliability achieved using clients and staff
was .85; inter-rater reliability using clients
and staff, at another place, was $R = 0.86$.

Test/retest reliability, $R = 0.85$. So I think, pretty robust, albeit it in a relatively finite sample.

DR. PINCUS: We really don't have the specific methodology that was used for doing this, and has it been published?

DR. WINKLER: Since the developer is on the phone, they could provide a little background if that could help us.

CO-CHAIR SUSMAN: Is the developer here on the phone, can you hear us?

MHA REPRESENTATIVE: I can hear some of you, though I can't hear others.

CO-CHAIR SUSMAN: What we are talking about right now is the reliability testing and we wonder if you might be able to describe a little bit further what sort of reliability testing has been done, and where and if that has been published.

MHA REPRESENTATIVE: Sure. First of all there was somebody who described or
mentioned the fact that one of the sites that this had been tested on mostly is at our Village program here in Long Beach. We did our major reliability study on that, and that was the study where we did get about a .85 inter-rater reliability coefficient. Basically as it mentions in the article, we had all of our clients rated by up to five different staff, and all staff were blind to each other's ratings, so that was a fairly large number of clients.

We also did another inter-rater reliability where I went to Massachusetts and trained the staff of a large mental health provider in Massachusetts, and that was the study with 105 clients who were rated by various members of staff who were also blind to each other's rating, and they got just slightly higher; that was the .86 coefficient that was mentioned.

So those are the two inter-rater reliability studies that we did.
CO-CHAIR SUSMAN: Are there any further questions about reliability testing?
Yes.

MHA REPRESENTATIVE: I'm sorry?

DR. GOLDBERG: We've had some reaction to the fact that the test can be given in 15 seconds. And what these raters - is that true?

MHA REPRESENTATIVE: It's based on the staff knowing their client that they are rating. So fundamentally I think somebody mentioned the fact that these are for folks who are considered to have a severe and persistent mental illness and who have been in the system for quite some time. Here in California these folks are primarily serve in what are called full service partnerships, so we have very low caseloads, above about one to 15, one to 17. So every month all of our staff rate their consumers, clients on their caseload, and literally takes about 15 seconds, because if you know the client you
know sort of what their risk factors are, what
t heir level of engagement is and what their
level of skills and supports are, so it
doesn't take long at all.

We work as a team on a team basis,
so it's not unusual for everybody in the team
to know everybody on everybody else's
caseload, and that's how we can do inter-rater
reliability that are so high, because the
staff do know members who aren't necessarily
on their caseload, but we are very familiar
with all of them.

DR. STREIM: So another way to put
that is that it only takes 15 seconds to
decide on a Likert scale rating and circle it.

MHA REPRESENTATIVE: That's
correct.

DR. STREIM: But can you estimate
how much time at any cross-sectional
assessment the caseworker or whoever is
following this client, how much assessment
time they actually take to find out how they
are doing, how they are behaving, how they are functioning, how is their last two weeks been going. Because they are interacting with that person, making a clinical assessment, and that clinical data then translates into the 15 seconds scoring.

MHA REPRESENTATIVE: Right.

DR. PINCUS: If you brought an independent -

COURT REPORTER: Microphone please.

DR. PINCUS: If you brought in an independent assessor to obtain the score, how long would it take them to achieve a valid ability to put down a valid score? Although what I'm really asking is, in the real world with the assigned caregiver how long does it take that person who also knows enough of the history that they don't have to reiterate it at every subsequent measurement period. They --

DR. PINCUS: What is the marginal level of effort needed -
DR. STREIM: Exactly, because that is really - in terms of the burden of the instrument and what it takes to actually accomplish this, that is the real measure.

MHA REPRESENTATIVE: Right. I think I understand your question. As we explain in our manual we actually encourage people to use the MORS in one of two ways. You could use it as an individual measure where basically the case manager thinks about how the person is doing, tries to assess them on the three constructs of risk, engagement and skills and support, and then butts up with that. And because they are meeting with their clients regularly, you know, you don't see huge shifts in those underlying constructs from day to day. So we've also done a lot of looking at sort of the stability of ratings over time. And so what I heard somebody questioning well, what is the numerator and denominator in terms of what is the time frame, is that we are looking at periods of a
year to two years in terms of people who may enter the program when they come off the street. They may be high risk unengaged, so they would be rated as a two. But over time we would expect - and that is really the question, we want to look at the trajectory of recovery and see how can different programs do in terms of moving people from a two to a seven or eight, how long does it take on average, those are the kinds of questions we want to use the scale, and that's why we think that it really should be considered at outcome measure.

But the other thing about the way that we have rated folks is that we often encourage our own team to do the ratings as a team, so our teams meet once a week, to discuss how their members are doing, how their clients are doing. So during that meeting, during the discussion, people - different people, different staff on the team, may have different information about how the client is
doing. That is all kind of put together into
- and the client is given a rating based on
that discussion. So much of that team meeting
can be used in that way.

CO-CHAIR SUSMAN: A couple of

further questions.

We have an unusual placement of
microphones, and we have to wait until they
are shuffled around.

MHA REPRESENTATIVE: Sure, no

problem.

DR. HENNESSEY: Hi. Have you done
any reliability studies looking to see what
kind of inter-rater reliability there is when
you compare an individual rating versus a
group rating?

MHA REPRESENTATIVE: No, we have

not done that.

CO-CHAIR SUSMAN: Since these
measures are ultimately being proposed for
accountability purposes, do you have any
standardization timeframe or other
specification here that will make this a more suitable measure for those purposes? In other words if I measure it at one year and Eric measures it at three years, and his population is a little bit less sick because they are not getting any patients who may have fallen into the criminal justice system and yada yada, it sounds like we might do well to say apples and oranges, and that for accountability purposes this measure wouldn't be appropriate. Am I misunderstanding what you are proposing?

MHA REPRESENTATIVE: I think that the common wisdom is that recovery takes a long time, and we are talking in terms of half a decade for a lot of people who come in as high risk unengaged. But I have seen members - we tend to use the term, members, as opposed to clients or consumers - I have seen members come in as high risk unengaged, and be able to reach early recovery within a six-month period, so I think the individual path of recovery is going to be very different.
depending on the individual. But I think that we really want to use this to find and give an idea of what are the typical trajectories of recovery. I don't think that we really know or can really speak to that question, because we don't have a tool that actually has a way of quantifying people's paths to recovery on the aggregate. I mean there are a lot of anecdotal stories out there about how people recover, but we don't know how programs are at actually helping people move through that process. So this is our attempt to quantify this to some extent and say, given the fact that if we had a large group of people who come in at these earlier stages of recovery how long does it actually take us to boot them to the higher stages of recovery? How long does that process typically take? So we are really trying to provide some information to the field about that.

DR. STREIM: Are you collecting data on the mean times that are spent at any
given level of recovery to know --

MHA REPRESENTATIVE: We in our programs we collect this - the milestones every month, and we strongly suggest that in other programs that are started using it do the same. So we really tell people that they should do it less than quarterly so that they can start getting the data points over time and actually have a feel for what progress or lack thereof that they are making. We have also got some papers in press or under review to sort of look at what are those average times in our own program as well as others.

DR. STREIM: And the converse, time to relapse or regression to a lower level, are there data at this point that you have collected on that as part of a recovery trajectory where they may have bumps in the road and setbacks and then advances, two steps back, one step forward?

MHA REPRESENTATIVE: We are looking at that as part of this paper, but I can tell
you that the one study that we have had going
on this, for example, the early data, for
example, the kinds of information that we are
hoping to get out of this is that for all the
people who come into our Village program for
example is that based on our Milestones to
Recovery data, what I can tell you is that
anybody who comes into the program at a
relatively high risk, that is they are a one,
two or three when they come in, is that within
one year if you look down the road one year at
their recovery, there is still about a 6
percent chance that they would be still at
that high risk category. So 94 percent of our
folks after one year are now above the high
risk category if they came in as a high risk
person. So that is the kind of information.
Now is that particularly good for a program or
particularly bad for a program? I don't have
any benchmarking data so I can't tell you
that. But those are the kinds of information
that we are trying to use the Milestones of
Recovery scale to help us to understand.

CO-CHAIR SUSMAN: Eric.

DR. GOPLERUD: I think we have here a really good example of a field developed, program developed measure which is maybe jumping too quickly but is not ready for nationwide implementation and prime time, but not only needs to be encouraged at the local level to develop it, but really to bring in some of the technology of the folks to do the - some of the critical issues around risk adjustment and the questions that we have asked about inter-rater reliability, if you have an outside objective observer, some of the validity testing using different populations et cetera.

It's on a topic that is incredibly important, and it is probably - it may be a measure that could be ready for prime time at some time in the future if developed. On the other hand there are so many challenges right there on the scientific acceptability that it
is very difficult at this point to go forward
I think at a national level and say, yes let's
support this.

CO-CHAIR SUSMAN: Eric and then
Harold.

DR. GOLDEN: Similar comments. I
think that it has great promise as a quality
improvement measure, but because of the
problem of validation I'm not sure it could
ever become an accountability measure. So I,
depending on how you propose the vote, I could
not endorse this or support this
scientifically as an accountability measure.

CO-CHAIR SUSMAN: Harold.

DR. PINCUS: I agree with both of
the previous comments, but also I think the
issues of usability in terms of understanding
sensitivity to change, and what are the
elements that actually influence that change.
So that if organizations are seeking to apply
this as a - it kind of goes to what you are
saying - seeking to use this as a quality
improvement strategy so how do they improve. What are the mechanisms to do that? Would be important to begin to elucidate.

CO-CHAIR SUSMAN: So I'm hearing from the group a lot of excitement that this type of measure is being developed, but concerns about some of the basic scientific acceptability currently, things like risk adjustment, looking at disparities of care, population differences, validity, reliability when you have naive observers or objective observers.

Are we ready to vote on scientific acceptability? Let's go ahead then and completely on scientific acceptability?

(Show of hands)

CO-CHAIR SUSMAN: Partially.
DR. WINKLER: Five.

CO-CHAIR SUSMAN: Minimally.
DR. WINKLER: Thirteen.

CO-CHAIR SUSMAN: Okay, let's move on to usability. We have already had some
comments in this direction. Further
discussion of usability. Do you have
something, Sheila?

DR. BOTTS: I think that Harold's
comments addressed those, and part of this is
just an interpretation and meaningful. You
know you are going in a direction of
improvement, but what that improvement
actually means in terms of outcomes and being
able to apply that as an accountability
measure I think there is a huge gap still.

CO-CHAIR SUSMAN: Any further
thoughts from the group on usability before we
vote?

Okay, completely?

(Show of hands)

CO-CHAIR SUSMAN: So partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimally.

DR. WINKLER: Sixteen.

CO-CHAIR SUSMAN: And then not at
all.
DR. WINKLER: Two.

CO-CHAIR SUSMAN: Okay, let's go ahead to feasibility. Remember that this is a byproduct of care, the issue of burden, ability to electronically incorporate such measurement, exclusions, looking at inaccuracy in the implementation issues here.

Thoughts from the group, please.

MR. PELLETIER: It sounds like the measure is embedded in a practice based on a model, based on the recovery model. Certainly it sounds like this is being talked about all the time. And this is a framework that the inter-disciplinary team uses to talk about patients recovery. So I think those are strengths.

DR. GOPLERUD: I think one of the big limitations is in the material that we were given it shows that this is something that you said it was embedded in a program; in fact it's one of the leading most reputable recovery programs in the country. And the
replicability of it I think is fairly low until we see some evidence that it is replicated. That they don't mention at all things like exclusions I think is really a problem if a measure like this is - are cognitively impaired individuals going to be excluded? Patients with organic brain syndrome, patients who are substance abusers. I mean there are a whole lot of different criteria. And then data collection strategy I think reflects that this is part of the program and hasn't been taken out to more programs to test it. So I think those are real limitations not that they couldn't be overcome, but I don't think at this point that it's ready for that.

CO-CHAIR SUSMAN: So I think the sense that I had is that this is a great start but we are not at the accountability stage yet.

So any further comments on feasibility?
Let's go ahead then and vote.

Completely.

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimally.

DR. WINKLER: Seventeen.

CO-CHAIR SUSMAN: Not at all.

(Show of hands)

DR. WINKLER: Bill left.

CO-CHAIR SUSMAN: Okay, I think we had a robust discussion, have been impressed by the work being done, but - pardon me? I'm getting up to recommendation.

How many would vote in favor of adopting this measure? Yes.

(Show of hands)

CO-CHAIR SUSMAN: And the nos?

DR. WINKLER: Seventeen, Bill left.

CO-CHAIR SUSMAN: So seventeen,

Bill do you vote yes or no? Okay thank you.

DR. WINKLER: Were there any
CO-CHAIR SUSMAN: Okay, so again for the sake of our developer, I think the committee is enthusiastic about the potential of this concept and measure, but there are many issues which the feedback from the group and staff can be passed on, and we sure hope that this will lead to a measure in the future. So thank you very much for taking the time today.

MHA REPRESENTATIVE: Sure, I look forward to getting all of your feedback, and to your guidance in terms of the meeting the qualifications that you are looking for. Appreciate it.

CO-CHAIR SUSMAN: Is there any public comment?

Okay, yes, thank you very much for taking time today. Let's go ahead then and move on to our next which is time for first face-to-face treatment.
FACE-TO-FACE TREATMENT ENCOUNTER

BUPRENORPHINE DOSING

CO-CHAIR SUSMAN: Medication developers? Well, was it really representative here.

MR. CORBRIDGE: Donald, have we heard if Baltimore Substance Abuse is on the line?

DR. OLSEN: We are right here.

MR. CORBRIDGE: They are here.

For those measure developers from Baltimore Substance Abuse, can you just state who is on the phone?

DR. OLSEN: Yes, I'm Yngvild Olsen, vice president for clinical affairs, and the medical director for bSAS

MS. KUHN: And I'm Vanessa Kuhn also with bSAS.

CO-CHAIR SUSMAN: There are a couple of questions around the table of just briefly your organization, who you are, two minutes or less?
DR. OLSEN: Sure. So Baltimore Substance Abuse Systems is a quasi-governmental agency that has the monitoring and oversight and some funding responsibilities for a wide range of treatment services, prevention, intervention and treatment services for substance abuse in Baltimore City, and one of the innovative areas that we have focused on is the adoption of buprenorphine into what previously were kind of drug-free outpatient substance abuse treatment programs to help increase access to effective substance abuse treatment for opiate dependence which is a huge problem, I think as probably most people know, in Baltimore. And the model that we have adopted is to start buprenorphine in outpatient substance abuse treatment programs, and link that to ongoing primary care outpatient medical care, both as a way to continue the buprenorphine, but also to integrate our medical care for individuals with opiate dependence. So that is where
these measures originated, and we really appreciate the opportunity to talk with you today about the two measures we have submitted and our happy to answer any questions.

CO-CHAIR SUSMAN: Thank you very much. We appreciate your taking time. There may be questions along the way. We have a fairly structured approach here, but there may be some issues which we wish to clarify.

Ian, did you just want to go over the specifications overall?

MR. CORBRIDGE: Can do sir. Right now we are currently looking at measure #13, so it's time from first face-to-face treatment encounter to buprenorphine dosing. Number of hours of opiate dependent non-pregnant adults. So the description is number of hours opiate dependent non-pregnant adults aged 18 or older have to wait between the first face-to-face treatment encounter and receiving their first dose of buprenorphine medication.

Numerator statement reads as
follows: opiate dependent patients receiving a first dose of buprenorphine medication.

Denominator statement reads: the event of an adult aged 18 or older, opiate dependent, buprenorphine appropriate, and treatment counseling patients received the first dose of buprenorphine.

CO-CHAIR SUSMAN: Okay, so those are the group. Would you care to address is this an outcome measure or a process measure? I was frankly pretty skeptical that this was an outcome, an outcome that is relevant to patients, and there may well be symptoms or issues that result from a delay that I didn't quite see this as a patient-oriented outcome myself. At least I had some concerns about that. So Richard.

DR. GOLDBERG: Can I make a comment on the extent to which there is data, that this time interval relates to an outcome. Is this an intermediate outcome? Is there good data that - you understand the question
I hope. I'll rephrase it if I need to.

CO-CHAIR SUSMAN: And maybe that's a good thing to put to our measure developer, but is this a causal pathway or intermediate outcome to patient-oriented outcomes that would matter?

DR. OLSEN: Yes, so thanks for that question. This is actually a process measure. It's intermediate outcomes to the ultimate outcome of retention and treatment. So there is some evidence that the sooner patients are - receive medications and the sooner that they are engaged in care, the better the retention of the treatment will be. You are correct, this is an intermediary outcome measure.

DR. GOLDBERG: What is the nature of that data? You say there is some evidence, or you have evidence that the time to starting buprenorphine is tied to retention and treatment? What is the nature of the evidence that exists for that?

CO-CHAIR SUSMAN: Are you still
DR. OLSEN: Can you hear us?

CO-CHAIR SUSMAN: No, did you hear the question?

DR. OLSEN: No, can you repeat the question?

DR. GOLDBERG: Just so you can refresh us about the nature of the evidence that ties the time to dose to your outcome which is, you are saying retention of treatment. What is the nature of that evidence?

DR. OLSEN: There are a couple of studies that we have cited that suggest that the sooner a patient gets engaged in treatment and if you wait three to five - longer than three to five days to get people into treatment that likelihood of dropping out of treatment increases.

DR. GOLDBERG: And where is that - is that published? Is that an accepted scientific finding? That has been reported in
quite a few research studies looking at rates
of show dependent on length of time to first
appointment. It is not specific as far as I
know to buprenorphine dosing. It has more to
do with the length of time between initial
contact requesting service and the first
service, and that is extensively reported on
the NIATx website. Again, there is no reason
not to believe that the sooner you get
buprenorphine dosing that the greater is the
likelihood of retention. But I doubt that
there is any buprenorphine-specific data that
says some interval, at least better than
another, or that it is anywhere different for
buprenorphine than for something else.

One the other hand we have a measure
that is before us which is specific to
buprenorphine dosing, even though perhaps the
committee might be interested in length of
time to first appointment more generally for
either substance use or for behavioral health.

DR. PINCUS: I guess my concern is
that this mere distance from outcomes than a
number of the processy things that we looked
at.

CO-CHAIR SUSMAN: Certainly my
sense in initially reading this is that this
was somewhat removed, and I think it's a
judgment call because clearly there is some
relationship. And how important you judge
that causal pathway to retention and treatment
in the Baltimore patient area outcomes is in
the eye of the beholder. Sheila, what did you
think about that?

DR. BOTTS: I had trouble
deciding. I mean I look at this as an
intermediary outcome that was important, and
it's important to look at where you draw the
line between what we want to include versus
exclude. The fact that there are no other
measures makes me inclined to say, perhaps we
should stretch on this issue. But again I'm
not --

CO-CHAIR SUSMAN: I am going to let
Luc, and then we will get --

MR. PELLETIER: I was stretching too, especially with the developers' discussion of TIP 40 as being evidence, and I wondered whether this particular organization is using that and then trying to get more data about whether something was effective or not, so they were developing a measure to prove what may not have been really strong.

DR. GOPLERUD: It is fairly clear FDA approved buprenorphine because it shows reduction of craving and opiate use goes down if a patient is taking buprenorphine compared to placebo or to other medications. Therefore it's not a stretch to say if you get a patient started on a medication which is known well to be effective in reducing opiate use but it might be linked as a process towards an outcome which is well known.

CO-CHAIR SUSMAN: Rich.

Okay, are there any members of the committee who say this should be taken out

Neal R. Gross & Co., Inc.
202-234-4433
because of out of scope? Maybe we should go ahead and take a vote then.

How many of you believe this is in scope? That it is sufficient as an outcome measure, or as we've stretched things a bit, an intermediate outcome measure, how many of you would vote yes.

(Show of hands)

CO-CHAIR SUSMAN: This is -- right now we are looking sort of -- well, we're going to get the conversation going, and we are going to stop it right here. And against - it doesn't really matter. I think we've got enough. So we are going to go ahead. Too bad. I want to be inclusive; come on.

Okay let's talk about the importance. We've already had some conversation toward that. You know, my concern is for the accountability measure, this was a very narrow focus. And that was my comment up here. And there wasn't a lot of supporting data, there was some. And I think
we have heard the nature of that data already.
So let me turn to Sheila and Luc and then open
it up.

MR. PELLETIER: I thought it was
important, I thought this was an important
topic and the framing of it using the evidence
from the TIP was substantial I thought.

CO-CHAIR SUSMAN: Other comments?

DR. PINCUS: As an accountability
measure I think it's very narrow. If this
were framed as something broader, Eric is
gone, but more like what Eric described as
something looking at a larger set of time,
engagement and treatment in some ways, for a
broader population, it would have more
utility. And so I just don't see this being
picked up a lot except as an internal quality
improvement measure. But not as a large scale
accountability measure.

MS. JAFFE: I have a question for
the staff given that this is a much more
narrow measure than anything that we have seen
before, are there other measures that are this narrow?

DR. WINKLER: Definitely, I mean there are over 600 measures in the portfolio and some of them are very narrow. Your question is, and this is more philosophical than policy, is that appropriate? Is that useful in the grand scheme of things? And we put that to you and ask you to advise us.

DR. GOLDBERG: I find myself thinking of like the term of antibiotics to certain outcomes. But the data that ties that intermediate outcome measure to be acceptable is pretty robust data in terms of the outcomes that they are talking about. And here it's by implication. But it's not here. So there is no reason not to believe that this wouldn't have an impact on retention and treatment which should have an impact on outcomes, but it's not really at the same point of antibiotics in the ER for pneumonia treatment.

DR. STREIM: I would argue though
that for substance abusers it's not a fair comparison to infectious disease; that engagement and retention and treatment may be more challenging with that population and that particular set of health problems. So I think the argument made by the measure developer that it could make a difference, and indeed is an intermediate outcome measure I think is persuasive enough.

CO-CHAIR SUSMAN: And I think the relief of pain and suffering symptoms in and of itself is pretty substantial patient oriented outcome, and if one's suffering longer --

DR. GOLDBERG: Right, but this is a slippery slope. If you let this in the door and you pick up thousands of measures like this that could be submitted and presented for --

CO-CHAIR SUSMAN: I don't disagree, but I --

DR. PINCUS: -- retention and
treatment would be a more - have more proximal
benefit.

DR. STREIM: Well, under depression
you could argue time not to first dose but to
first appointment could be important. I mean
you can imagine similar things --

DR. PINCUS: Right, we almost
knocked out measuring base care as not being
processed - being too process-y.

CO-CHAIR SUSMAN: I think there is
a certain amount of behavior here. How about
gap in relationship to outcomes I think we
have already covered. Anything further?

MS. JAFFE: I guess I wonder if we
would have had more submissions of other sorts
of these where it was time from treatment to
prescription of anti-depressants, would we
have a different conversation? We just happen
to have only one of them, so I think that is
something to consider as well.

DR. MANTON: I also think that the
topic is important to consider. I mean we are
talking about importance to measure and report, and I don't know that the rest of the category will show that it's worth the docking, but I do think that in terms of importance, the measure and report, it's a substantial problem, and I think that whatever we can do to measure the differences that occur because of prompt treatment would be worth looking at. So in terms of importance I think it should be considered.

DR. STREIM: For NQF staff, what do we have in the library for measures of substance abuse outcome? Just curious, I mean this is a process measure, so looking at process.

DR. WINKLER: There are like two or three. Most of the work we've done on substance abuse has been around practices. I'd have to go back and look. But there are very few, and they are process measures. The Washington Circle measures, and I don't think there is much beyond that.
DR. PINCUS: Maybe when we get to the harmonization issues, it seems to me that this is encompassed to some extent by the Washington Circle measures.

DR. GOLDBERG: I'm a little obsessed with the outcomes part. This is an intermediate outcome towards some outcome. Why don't we tell them, present the outcome? What's the outcome that this is intermediate towards, and I'd like to consider that measure. You know the problem, micromanagement, like thousands -

(Simultaneous speaking)

CO-CHAIR SUSMAN: So I hope the measure developers hearing this conversation about where the outcome is. Well, let's go ahead and vote on importance here. Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimal.
(Show of hands)

CO-CHAIR SUSMAN: So we are set on that.

Let's move on then to scientific acceptability. So I mean just to telegraph my thoughts here I thought that the analysis the analysis at least was presented around things like reliability, validity, was very thin, if at all. And I saw that as an important weakness.

Sheila, what were your thoughts?

DR. BOTTS: I would probably echo some of your comments in terms of testing. Again it's looking at it as an intermediate outcome, even the relationship to improve their tension. I mean there's a large suggestion, comes from a lot of clinical trials, whether - I think it's whether we have another process or outcome measure, but a comfort level in terms of scientific validity.

MR. PELLETIER: The developer actually stated that there was no formal
reliability --

CO-CHAIR SUSMAN: Likewise, risk adjustment was not considered or suggested.

No risk adjustment necessary, which I guess probably you could say there should be a standard that is applicable across types of patients. At least that would be maybe more sellable. But if you are going to do different populations across different programs, that might have an impact.

Facts, comments, from the committee as a whole on this?

Let's vote then. Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimally.

(Show of hands)

CO-CHAIR SUSMAN: And then not at all.

(Show of hands)

CO-CHAIR SUSMAN: Okay. Let's move
on to usability.

Again I thought there was just a relative dearth of data.

DR. PINCUS: I think there needs to be some effort at harmonization with the existing NQF measures, because I think they may in fact encompass and be better than.

DR. BOTTS: The notes here say that there are no similar or related endorsed or submitted measures. Is that accurate?

DR. WINKLER: I would have to look at the details of the Washington Circle measures. Those have been endorsed. I'd have to look at the details on them actually.

DR. PINCUS: Initially for those -- the initiation is essentially looking at going from identification to risk assessment.

CO-CHAIR SUSMAN: Other thoughts in this one looking at the Washington Circle?

MS. BOSSLEY: Let me read it out loud. Because I don't think you can read it.
CO-CHAIR SUSMAN: No.

MS. BOSSLEY: So it's the percentage of adults aged 18 and over diagnosed with AOD abuse or dependence and receiving a related service to initiate treatment, assessment of the degree to which members engaged in treatment with two additional AOD treatments within 30 days after initiating treatment. So it's two pieces: initiation and then within 30 days.

DR. BOTTS: So the second piece - so the first piece is the number - anyone who is diagnosed and received the related service and initiates treatment, so just that, the number. And then the second piece is how many days to additional treatment within 30 days.

CO-CHAIR SUSMAN: Okay, so there is at least some overlap at a broad level, whether you think it's important for this particular focused measure or not, I think, is again up to the group. Any other comments on usability?
Then let's move on to vote.

Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimally.

(Show of hands)

CO-CHAIR SUSMAN: Not at all.

(Show of hands)

Okay, let's move on then to feasibility. Do you want to start off, Luc, and tell us your thoughts about feasibility?

MR. PELLETIER: I think what I said here is, performance is limited to a group in a city. Current system features aren't well described; didn't really get a good sense of how burdensome this is.

CO-CHAIR SUSMAN: Sheila.

DR. BOTTS: I thought it seemed to be fairly straightforward in terms of getting the time to treatment within a system, so it seemed that the data would be readily
accessible, the data in the lab, the methodology.

CO-CHAIR SUSMAN: Yes, I guess from a sort of face validity standpoint it sort of made sense that this would be relatively feasible to do, but there were no real data. This is basically one system's ability to do this, and whether it transfers to other settings I think is unknown.

DR. PINCUS: I would think for the most part it's a large system, it would be very difficult, because you have to combine - it's based on hours, and I don't know the time for figuring out the hour of dosing from the time - you know, you couldn't use claims --

CO-CHAIR SUSMAN: So issues of confidentiality. Other concerns, questions, comments.

DR. MANTON: I guess I would suggest that they look at doing a research study first, because it doesn't make sense to me to look at the time to actual treatment
without knowing that it makes a difference.
So I think what I'd recommend is that they do
a research study, come back with what that
shows them, and then look at outcome measures.

CO-CHAIR SUSMAN: The measure
developer does note that data is easy to take
as long as data entry occurs in a timely
manner; data needs to be entered into the
database to do accurate tracking and efficient
workflow, which sounds to me like a separate
process; it does not occur as a routine part
of care if you will.

Okay, if there aren't any other
comments then let's vote.

Completely?

(Show of hands)

CO-CHAIR SUSMAN: Partially.

(Show of hands)

CO-CHAIR SUSMAN: Minimally.

(Show of hands)

CO-CHAIR SUSMAN: And then not at
all.
(Show of hands)

CO-CHAIR SUSMAN: Okay, then let's go on and vote, how many of the group would recommend yes, adoption of this.

(Show of hands)

CO-CHAIR SUSMAN: How many would recommend no?

(Show of hands)

CO-CHAIR SUSMAN: Any abstentions?

Any public comments?

I want to thank the measure developer. I think everybody is very supportive of the concept here, I think there are some suggestions about how to go from where you are. It really would be possible, I think, for us to move on to more of an accountability measure by looking at ultimate outcomes for tension and treatment.

Let's see the next one, same developer, yes, well, let's go. Percent of eligible patients who transfer.

MEASURE OT3-017:PERCENT OF ELIGIBLE PATIENTS
WHO TRANSFER FROM A SUBSTANCE ABUSE PROGRAM
TO A CONTINUING CARE PHYSICIAN FOR ONGOING
BUPRENORPHINE MAINTENANCE THERAPY

MR. CORBRIDGE: So we are moving on
down to #17, Percentage of Eligible Patients
Who Transfer From a Substance Abuse Treatment
Program to a Continuing Care Physician for
Ongoing Buprenorphine Maintenance Therapy.

The description reads as follows:

percent of adult patients aged 18 years or
older who meet eligibility criteria to
transfer from a substance abuse treatment
program where they have been induced,
stabilized on buprenorphine, and received
counseling services, to a continuing care
physician in the community who will continue
the patient's buprenorphine treatments and
will provide other mental health and
social/medical services.

Numerator statement reads: the
percent of adult patients who began
buprenorphine treatment at a substance abuse
treatment program who upon stabilization, on
buprenorphine, and upon meeting transfer
eligibility, ensured stable negative urine
drug screen, responsible with prescription
handling, transferred buprenorphine to health
care services to a continuing care physician
in the community.

The denominator statement reads: all
patients who were inducted and stabilized on
buprenorphine in a substance abuse program,
and to meet the transfer criteria. The
transfer criteria are stated as: ensured,
stabilize, negative urine drug screens,
responsible prescription handling. Regardless
of whether they ultimately transferred their
care to a continuing care physician in the
community or not.

CO-CHAIR SUSMAN: So again I guess
you could ask is this a patient related
outcome. Their tension and treatment, we
probably will have the same set of issues.

DR. PINCUS: transferred. Why is
somebody needing a transfer?

CO-CHAIR SUSMAN: Should we ask the measure developer if they are on?

DR. HENNESSEY: Is what we are talking about then is an outpatient substance abuse treatment program where say someone who is a nonphysician has assessed someone as potentially benefiting from this medication, and so now the person is being referred to a physician who has this expertise; is that what we are talking about?

CO-CHAIR SUSMAN: And other appropriate services is what I understand this measure.

MS. JAFFE: I understand it that they are in a specialty substance abuse program, probably being treated by a physician, and they met some criteria so that they no longer need that level of care and can return to primary care.

DR. HENNESSEY: Okay, thank you.

DR. PINCUS: -- necessarily a path
to outcomes for everyone.

MS. JAFFE: I would think that it might be more a reflection on the comfort level of the primary care physician and not so much on the patient.

CO-CHAIR SUSMAN: Well, I mean the description is patients able to continue and receive maintenance therapy, convenient office setting, other somatic and mental health services, mitigating relapse, continuing care physicians are able to take care of already inducted and stabilized uninsured patients. Their practice office settings do not need to be altered to accommodate time consuming and sometimes difficult and/or uncompensated induction protocols, waiting room disruptions, yada yada. And three, the stable patient condition out of the publicly funded treatment slot and substance abuse program, a new patient in need of service is able to enter the program.

DR. MANTON: It sounds like a
system as opposed to a provider outcome.

DR. HENNESSEY: It sounds like a utilization outcome to me.

DR. ROCA: I could certainly see that it could be a quality outcome if the treatment program made the determination that this is somebody who is appropriate for maintenance treatment, then I think it would be a responsibility of that program to do whatever they could do to ensure that they got into the next stage of treatment which would include maintenance. Presumably not everybody is a candidate for this, and I'd be interested in what the eligibility criteria were. But presumably the eligibility criteria would include being appropriate for more of a long term maintenance buprenorphine treatment that might involve other treatments as well.

DR. GOLDBERG: But this has something to do with getting out of a specialized treatment system to a primary care patient system –
(Simultaneous speaking)

DR. ROCA: But with an appropriate provider.

DR. GOLDBERG: Even with the appropriate provider, I mean conceivable to me they may make their transition and then drop out after a week. So I don't know what the outcome is, just to say that we got rid of some people, we transferred some people to the primary care system, is an ambiguous outcome to me.

DR. PINCUS: That's basically the equivalent of saying that someone who is used to being seen at special a mental health center got transferred to a primary care provider. It may be appropriate for some people, but I don't see how it's relevant --

MR. PELLETIER: The way I read it was that she was describing a community standard that someone is inducted, they go to maintenance to a person who is familiar with this medication and has gone through the
training to medicate this person; that's how I read it.

DR. PINCUS: Right, but what is the counterfactual this person remains in the substance abuse treatment program.

DR. ROCA: Or is lost to treatment.

DR. PINCUS: Right but that's --

MS. JAFFE: I thought I read something in there that you move them out of the specialty so you can make room for a new person.

CO-CHAIR SUSMAN: I mean this is from a perspective of a community health service agency and what their goals are to get patients induced and then get them into ongoing care and a whole range of services. Now whether that's an appropriate outcome measure or not, I think, is the first point here. Is this in scope or not.

DR. STREIM: I am not a substance abuse subspecialist, but however, I would wonder how many primary care physicians have
done the training, paperwork, have the special DEA number which you need for this. I happen to know this, because I actually got this training. I have never actually prescribed buprenorphine, because I do geriatrics, and we don't have too many of those patients. But the question is, how many primary care physicians in the entire United States do you think are actually eligible to prescribe, and is that a common enough phenomenon in any sector of our health system that this would be an efficiency in health care utilization that we would want to measure in a nationally reported measure? I don't know the answer, but I think that is an important question.

DR. MANTON: Actually I think a lot of primary care physicians can prescribe buprenorphine.

DR. ROCA: I don't know how widespread the utilization of this would be, but if you were a substance abuse treatment program that might not be an unreasonable
thing to expect.

DR. BOTTS: I would agree, and I think you kind of get at the heart of the issue is that you have a drug treatment system that is highly regulated both from the patient standpoint and the provider standpoint, and things can potentially get bottlenecked in terms of the turnover. So what you are looking at is efficiency for care, and the numbers involved, the same as large as in my population, no, but for that group it's incredibly important that we do it well.

DR. GOLDBERG: I wonder what data there is once they get transferred, how effective the primary care providers who are licensed and eligible, how effective are they at maintaining these people in treatment. Do we know that?

CO-CHAIR SUSMAN: I don't know that this necessarily implied primary care. It implied ongoing care, and requires ongoing care.
DR. GOLDBERG: Some continuing care providers that are not specialized --

CO-CHAIR SUSMAN: A continuing care physician in the community. I think the reality is that a very very small percentage of PCPs are doing this type of treatment.

(Simultaneous speaking)

CO-CHAIR SUSMAN: In response to my question Ann was saying not so.

DR. MANTON: I think that there is a fairly large percentage, and I think probably for just these reasons, that the drug treatment centers are saying, it certainly isn't 80 percent or anything like that. But I bet just as a ballpark I bet there is maybe 30 to 40 percent. Maybe it's a regional kind of thing.

DR. PINCUS: What evidence if any is that this is proximal to outcomes?

CO-CHAIR SUSMAN: So we have about 15 minutes. Let's first of all vote is this within scope. Is it in scope? Raise your
hand if you believe it's in scope, an outcomes measure. Raise your hands high. Five.

Okay.

Out of scope.

(Show of hands)

DR. WINKLER: Eleven.

CO-CHAIR SUSMAN: Okay, thank you.

That helped catch us up. It is 4:30. We have 15 minutes. I don't know if we want to address the next one which is substance abuse or begin that.

I don't know if you want to go on to tomorrow morning's or do you want to stop here?

CO-CHAIR LEDDY: This is workgroup four, and Ian has evidently split it into three and two because he thought this is about where we would end, right? So the two that you rated are first thing tomorrow morning we continue with this workgroup, then we go on to workgroup three.

CO-CHAIR SUSMAN: So what I'm
asking, and I think we are going to argue up
our time here, we've got about 15 minutes. Do
you want to spend that on the next measure, or
do you want to get out and enjoy the beautiful
Washington weather and see the cherry blossoms
or whatever else is on your agenda.

DR. STREIM: I think it is more
efficient to do it all at once, because we are
just going to have to reiterate tomorrow
morning what we discuss in the next 15
minutes.

MR. CORBRIDGE: Do we want to do it
now.

DR. STREIM: You mean extend and do
the whole thing? That's different if you want
to extend and do the whole thing.

CO-CHAIR SUSMAN: I think probably
starting tomorrow would be the most efficient
use of our time. I know if we can have
agreement on that we'll just wrap up today.
Some key things tomorrow, there is a good
overview of discussions today, where we stand
in terms of measures that we ended up with moving forward to potential endorsement. Most measures that we discussed recommended might not go forward.

Wanted to make a brief note that we will not be in the Brown Rudnick offices tomorrow. We are actually going to be in our offices, which is our meeting floor - I have to send email to everyone, so if you do have access to email. So it is on the 6th floor, however you went to the south side today. Our offices are on the north side. So what you are going to do is, you are going to walk in the building and go to your left, and then you are going to go on the north side of the building, go to the sixth floor, and as soon as you open up the doors you will be right at the NQF offices. We have a similar set up. We are not a lawyer group and so we don't quite have all the plushness of this room, but it should be sufficient tomorrow. And I think one of the main reasons we are moving is that
we do have access to a working phone which will be much more helpful in facilitating the process.

Just to clarify again, you will go in the same entrance right on 13th Street, and you will go to the north side, which will be turning to your left. You can ask the security guard or the concierge down there if you need any help with that.

I want to thank everybody for their hard work and forbearance, and look forward to seeing everybody tomorrow morning.

(Whereupon at 4:34 p.m. the proceeding in the above-entitled matter was adjourned.)
| Page 356 | Neal R. Gross & Co., Inc. | 202-234-4433 |
Neal R. Gross & Co., Inc.  
202-234-4433
Neal R. Gross & Co., Inc.
202-234-4433
two-step 128:3
tying 69:6
type 91:18 97:2
143:22 245:1
311:6 350:6
types 40:7 44:12
334:6
typical 138:12
141:1 218:14
306:3
typically 47:16
111:1 121:17
306:18
understandability 239:3
un-endorsed 94:15
update 93:17
updates 95:20
UPMC 218:22
urge 152:21
urinary 133:8
urine 342:3,13
usability 18:22
306:18
useful 25:21 26:14
27:1 46:14 75:4
171:22 232:15
239:6,12 265:2
268:16 294:4
328:8
usefulness 51:3
user 98:6 147:18
148:6
uses 72:2 121:22
223:9 313:14
usual 103:15,22
94:18
usually 112:17
usual
48:1
useless 45:17
un-engaged 303:3
305:16,19
unfamiliar 28:10
unforeseen 267:6
unfortunately 164:17
uniform 132:17
uninsured 344:12
unintended 168:11
175:5
unique 221:11
unit 243:13,16
217:12 245:10
246:7,16,17
261:10
United 348:8
units 24:18,19
universe 147:1
204:3
University 1:14,15
1:18,21 2:4,14
5:10 6:5,13,19
7:12 10:7 35:20
62:9
unknown 338:9
unquote 66:18
unreasonable 348:22
unsetting 266:14
unusable 212:13
unusual 121:18
300:6 304:7
Unutzer 99:8
306:2 308:22
310:22 323:21
325:12,17 338:15
352:19
valuable 155:20
value 106:8 156:10
171:16 184:11
185:4 238:18
239:22 240:1,4,5
240:16,19 241:7
291:2,3
values 44:16,16
45:20,22 46:1,2,4
58:20
Vanessa 2:15
317:17
variability 195:8
225:18 231:2
variables 41:22
50:1,3 59:8
variation 52:6
295:18
variety 26:14 40:6
133:18 264:21
various 59:8 107:3
261:8 298:17
vascular 91:12
vast 183:13
VA/DoD 113:21
Venn 171:15
verbalizing 203:3
version 37:13
163:22
versus 17:4 85:11
115:7 117:1
126:22 150:4
171:5 196:12
215:5 225:1
304:15 324:17
vetted 78:18
vetting 261:8
viable 275:18
vice 7:15 10:6
317:15
victory 276:6
view 6:2 21:15 71:1
86:10 103:1 138:3
141:4 171:9,13
191:16 219:14

Neal R. Gross & Co., Inc.
202-234-4433