The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Patricia Brennan and Joyce Dubow, Co-Chairs, presiding.

DR.S PRESENT:

PATRICIA BRENNAN, PhD, University of Wisconsin-Madison
JOYCE DUBOW, AARP, MUP, Co-Chair
RICHARD BANKOWITZ, MD, MBA, FACP, Premier Healthcare Alliance
ETHAN BASCH, MD, MSc, Memorial Sloan-Kettering Cancer Center
JIM BELLOWS, PhD, Kaiser Permanente

DAVID CELLA, PhD, Northwestern University Feinberg School of Medicine
ANNE DEUTSCH, PhD, RN, CRRN, Brookings Institution
STEPHAN FIHN, MD, MPH, Veterans Health Administration
LORI FRANK, PhD, Patient-Centered Outcomes Research Institute
BARBARA GAGE, PhD, MPA, Brookings Institution
TED GANIATS, MD, University of San Diego Health System
KATE GOODRICH, MD, MHS, Centers for Medicare & Medicaid Services
JENNIFER EAMES HUFF, MPH, Pacific Business Group on Health
DENNIS KALDENBERG, PhD, Press Ganey Associates
IRENE KATZAN, MD, MS, Cleveland Clinic
LEWIS KAZIS, ScD, Boston University School of Health
UMA KOTAGAL, MD, Cincinnati Children’s Hospital Medical Center
KEVIN LARSEN, MD, Office of the National Coordinator for HIT
KATHY LOHR, PhD, RTI
ELIZABETH MORT, MD, Massachusetts General Hospital
CHARLES MOSELEY, Ed.D, National Association of State Directors of Developmental Disability Services
GENE NELSON, DSc, MPH, The Dartmouth Institute
KENNETH OTTENBACHER, PhD, OTR, The University of Texas Medical Branch at Galveston
GREG PAWLSON, MD, MPH, FACP, BlueCross BlueShield Association
ELEANOR PERFETTO, PhD, Pfizer
COLLETTE PITZEN, RN, BSN, Minnesota Community Measurement
CHERYL POWELL, Centers for Medicare & Medicaid Services (via telephone)
DAVID RADLEY, PhD, MPH, Institute for Healthcare Improvement
TED ROONEY, RN, MPH, Maine Quality Counts
DEBRA SALIBA, MD, MPH, UCLA Borun Center/VA/RAND

MARCEL SALIVE, MD, MPH, National Institutes of Health
BARBARA SUMMERS, PhD, RN, University of Texas-MD Anderson Cancer Center
KALAHN TALYOR-CLARK, PhD, MPH, National Partnership for Women & Families
MARY TINETTI, MD, Yale New Haven Health System

PHYLLIS TORDA, MA, National Committee for Quality Assurance
JOHN WASSON, MD, Dartmouth Medical School
ROB WEECH-MALDONADO, PhD, MBA, University of Alabama-Birmingham
LINDA WILKINSON, MBA, Dartmouth Hitchcock Medical Center

ALBERT WU, MD, MPH, Johns Hopkins Health System
NQF STAFF:

KAREN ADAMS, PhD, MT
HEIDI BOSSLEY, MSN, MBA
HELEN BURSTIN, MD, MPH

SHANA CAMPBELL
SHEILA CRAWFORD
EUGENE CUNNINGHAM, MS
LAURALEI DORIAN
LAURA MILLER
ASHLEY MORSELL
ELISA MUNTHALI

KAREN PACE
LINDSEY TIGHE
TOM VALUCK, MD, JD
JESSICA WEBER
C-O-N-T-E-N-T-S

WELCOME & SETTING THE STAGE . . . . . . . . . .7
Patricia Brennan, University of Wisconsin-Madison
Joyce Dubow, AARP, Co-Chairs
Helen Burstin, Senior Vice President,
Performance Measures, National Quality Forum

ACKNOWLEDGING THE PATIENT AS AN
AUTHORITATIVE DATA SOURCE . . . . . . . . . . 38
Moderator: Joyce Dubow
David Cella, Northwestern University
Feinberg School of Medicine
Reactor Panel:
Stephan Fihn, Veterans Health Administration
Jennifer-Eames Huff, Pacific Business Group on Health
Charles Moseley, National Association of State Directors of Developmental Disability Services

BREAK

PROMISE OF PROs IN IMPROVING PATIENT OUTCOMES;
LESSONS FROM THE FIELD. . . . . . . . . . . . . . . . .110
Moderator: Greg Pawlson, BlueCross BlueShield Association
Partners Healthcare, Elizabeth Mort,
Massachusetts General Hospital
Dartmouth Spine Center, Eugene Nelson,
Dartmouth-Hitchcock Medical Center

METHODOLOGICAL ISSUES: METHOD OF ADMINISTRATION/COLLECTION & RESPONSE. . . .157
Moderator: Ethan Basch, Memorial Sloan-Kettering Cancer Center
David Cella
Reactor Panel:
Lewis Kazis, Boston University School of Public Health
Richard Bankowitz, Premier Healthcare Alliance
Lori Frank, Patient Centered Outcomes Research Institute

LUNCH BREAK
METHODOLOGICAL ISSUES: SELECTION
PATIENT-LEVEL PROs. . . . . . . . . . . . . . . . . . .241
Moderator: Albert Wu, Johns Hopkins Health System
David Cella
Reactor Panel:
Jim Bellows, Kaiser Permanente
Eugene Nelson, Dartmouth-Hitchcock Medical Center
Kalahn Taylor-Clark, National Partnership for Women & Families
Kenneth Ottenbacher, The University of Texas Medical Branch at Galveston

BREAK
KEY CONSIDERATIONS FOR INCORPORATING. . . . . 316
PROs INTO ELECTRONIC HEALTH RECORDS
Moderator: Patricia Brennan
David Cella
Reactor Panel:
Kevin Larsen, Office of the National Coordinator for Health Information Technology
Ted Rooney, Maine Quality Counts
Uma Kotagal, Cincinnati Children's Hospital Medical Center

CLOSING COMMENTS. . . . . . . . . . . . . . . . . . . . . . . . .401
Patricia Brennan & Joyce Dubow, Co-Chairs
DR. BURSTIN: All right, everybody, we're going to go ahead and get started. Good morning. I'm Helen Burstin, the Senior Vice President for Performance Measures at NQF. Thank you so much for joining us today and tomorrow. We're very excited about this meeting that has been long in planning, and I think will really offer us a great opportunity to think about this next stage of measurement that we all want to enter.

I'll do the introduction to the group, and then I'll turn it over to our incredibly capable Co-Chairs, Patti Brennan and Joyce Dubow, who will serve as your hostesses and keep the train moving through the next couple of days.

I just want to personally thank the two Karens who many of you have interacted with, Karen Adams and Karen Pace, and all of our staff, and Gene Cunningham in particular.
who have just done a phenomenal job of pulling this together.

So, I get to set the stage, and part of the reason for setting the stage is that we really want to try to, as best as possible, try to make a distinction between tools and measures. So, I think that's an important issue for us. So, next slide, Jessica.

So, briefly, a little bit about the project scope and the activities before I turn it over to Patti.

OPERATOR: Excuse me. This is the conference Operator. You're not in the main conference.

DR. BURSTIN: Oh, could you please put us in the main conference, Farah. That would be lovely.

OPERATOR: Okay, one moment.

DR. BURSTIN: All right. Great.

Good morning, everybody joining us on the web.

We prematurely started without you for a
couple of moments but welcome.

This is Helen Burstin. I'm just going to give a little bit of the project background before I turn it over to the Chairs for this workshop. So, briefly, a little bit about the project scope.

The first is that this meeting, in particular -- many of you who have spent time at the NQF tables know we tend to talk about endorsement of measures. Well, today we're talking about the methodologic issues, almost really a prequel, we think, to what will ultimately, we hope, be some further work on endorsement in the coming years.

So, essentially, our project scope is under the guidance of the expert panel, essentially, all of you in the room, we're going to have these two workshops to help bring together the stakeholders we think are necessary to really facilitate the critical path, the groundwork we need to get to the development, testing, endorsement, and
implementation of PRO-based performance measures, a pretty heavy lift from I think where we are now. But I think we felt that since there were so many methodologic issues that needed to be resolved, it was difficult to take a big leap and just call for these measures without really putting everybody on the same place, and really beginning to understand what the next steps would be.

You'll have the benefit of two Commission papers, the first of which you have today, thank you to David Cella and his team, on the first one to think through next steps about the selection of patient-level PROs for use in performance measures. And the second one which we'll get to in the fall will be the path to developing reliable and valid PRO-based performance measures that would, in fact, be eligible for NQF endorsement, and could be used for both accountability and quality improvement.

Thanks to HHS for funding this
work. We think it is really important, and we're really glad to be here. Jessica, next slide.

So, just briefly, the time line to give you a sense of where we're going, I've given you a little bit about that today. So, the workshop is today in July. We're going to be working, the paper writers are here with us. They're going to actually work through some revisions of the paper based on the input that you'll be providing over the next couple of days.

We'll have the chance to prepare that draft report, get a chance to hold that second workshop. Here are the dates, September 11th and 12th, at which point we'll have a second paper from Barb Gage and Anne Deutsch of RTI. That one will be much more grounded in how to move into performance measures. We'll have a public comment period from October to November. We'll get the expert panel, all of you again to review those comments, finalize
the papers and then bring them through our approval process before the end of the year.

Next slide.

So, a little bit about putting this in context. We often talk about the Quality Measurement Enterprise. And one important thing I want to mention is the top gold bar there is measure development. It's outside the realm of the others, and that's because very important distinction for NQF is we are not measure developers. We do not develop measures. It's a really important firewall for us, because we are the neutral evaluators of measures. But this does give you a broad sense of how this work fits into the broader landscape of the measurement enterprise.

The National Quality strategy has clearly indicated a goal for having more patient-reported outcomes. There is a standardized measurement process that we'll talk a great deal about in the coming months
that leads to NQF endorsement. But we also recognize it would be very difficult to do much of this without the emerging electronic data platform, also part of the work that we've been doing, trying to then think about the alignment of the various environmental drivers, how do people select measures, how do people find the right measures they want to use. And then, ultimately, evaluating and seeing if these are driving both improvement, as well as improved accountability. Next, please.

So, just a high-level view, I think, of where we're sitting in terms of what we've been viewing in terms of performance measurement.

The first is, there's definitely been a drive towards higher performance. I think as our criteria have gotten more and more rigorous, we are beginning to see a good number of the very basic process measures falling to the wayside in favor of some of the
more difficult intermediate outcomes and
greater measures which we think is very
positive.

I think there is a hope that with
more of a data platform we'll be able to
increasing measure disparities in all we do,
as opposed to the afterthought it often tends
to be now. If we have the data to always
stratify, then we insure that we do that in a
rigorous way.

We're seeing more of a shift
towards composite measures rather than single
process measures or single outcome measures,
trying to get a more complete picture.
Consumers and purchasers, in particular, find
these measures very, very valuable, and
clinicians and others find them useful when
they can be packaged, but still get the
broader view, and lots of different models of
this.

A major part of our work over the
last couple of years has really been about
harmonizing measures. There is a lot of
cacophony currently in the measurement space
of measures that are just slightly different
across different settings of care, slightly
different for different payers, public and
private. So, a great deal of our effort
currently has been around trying to make sense
of some of this, bringing them together,
having a more parsimonious set.

And then, finally, there's been a
great deal of interest in moving towards a
more longitudinal view of what we can do in
measurement rather than everything being very
siloded of what I do in clinic on Mondays
versus what others do in the hospital, versus
the patient's experience at home care, really
making it a more longitudinal view. And if you
do that more patient-focused episode it does
naturally lead you towards outcome measures.
That's what patients care most deeply about,
and especially I think patient-reported
outcomes, in particular. Am I actually going
to feel better as a result of this procedure, or measures along those lines.

If they are process measures, increasingly there is a move to make sure they're process measures that have a direct impact on the desired outcomes. It also moves us towards more measures of appropriateness, and in this day and age hard to imagine that we could look at quality in isolation without increasingly now bringing in measures of cost of resource use to couple them with quality measures including a view of overuse.

So, we had some work done a couple of years back now that Karen Pace led for us with David Shahian as chair of an evidence task force. And this was one of the, I think, really important pieces of work that emerged, was a very clear hierarchical preference for NQF for outcomes, as much as possible, linked to evidence-based processes, outcomes of substantial importance with a plausible link to processes of ways to improvement, although
not necessarily. And then if they are going to still continue to have process measures or structural measures, they need to be closely linked to outcomes. They can't be things so distal from the outcome that just continually measuring and improving those won't ultimately improve what we care most about. Next, please.

So, this is just a visual view. We have two visual views of it for those of you who think differently, see things differently. So, this is what we did last year of analysis of our portfolio, breaking it down by process and outcome measures.

And I realize this might be a bit difficult to read, but just -- as you could see from the lefthand side there are more of the classic areas that have been in measurement for years; prevention, cardiovascular disease, safety, surgery, musculoskeletal are the ones at the top of that pinnacle there.

And as you start going down, you
get to the very small ones towards the bottom there which, unfortunately, include patient experience and engagement and functional status. That's two of the lowest at the end there.

We also broke it down into process and outcomes, so we're actually pleased to see more of that movement towards outcomes. Actually, about a third of our portfolio now is outcomes as opposed to process measures and structural measures.

And the blue here are outcomes. And you can see that in some of the areas there are certainly more process measures than outcome measures. I think work we're going to continue to work on, but in some areas, for example, like surgery, way ahead in terms of thinking about outcomes as opposed to more process fields linked medicine which is my next.

And this is a different view of it, very similar, but this is the way we break
down our portfolio into what's crosscutting
and what's condition-based here. The white are
process measures, the blue are outcome
measures.

And just very briefly in this
lower box here, functional status, patient
engagement. So, two of the very, very
smallest boxes on that overall chart of our
over 700 measures now relate to those two
topic areas that we think in particular are so
important here. And those are, of course,
outcomes, and we want to try to get more of
those across the portfolio. Next.

So, before I hand it over to Patti
just one quick, I think, distinction that is
always difficult to transmit. We're going to
probably need to do this a couple of times
during the course of the meeting. So, the
first is that people often talk about
individual-level PROs, and then they talk
about performance measures. And they are, in
fact, different, and we need to make that
So, the first is, NQF doesn't endorse individual-level instruments or scales. We wouldn't endorse, for example, the SF-12, or a tool like that in isolation. It would need to be as part of a performance measure. So, although they may be very reliable, highly valid in clinical practice or research, those individual patient scores alone are not sufficient to really determine performance of a given entity, or make conclusions about the quality that's provided. But we recognize those individual-level scores are the data that are going to drive the performance measures. What we do endorse, though, are those performance measurements that result in a score for the accountable entity, a hospital, a practice, an ACO, whatever the case may be and use data from all those eligible patients.

So, at the same time, an endorsed performance measure needs to be standardized,
precisely specified so that the specific instruments and scales and scoring must be identified. So, these are highly linked, but at the same time, identifying just the PROs won't get us to where I think we need to go in terms of performance measures. Next.

And just to put this in perhaps a bit more concrete terms, here's two examples of two measures we've endorsed in the last year. The first was for Minnesota Community Measurement. I know Collette is here with us today, which is about use of the PHQ-9, which is a tool to gauge depression. And we've endorsed three measures, the first of which is actually utilization of the question there in the first place in clinical practice paired with one of these two measures, depression remission at six months, or depression remission at 12 months. So, we have not endorsed the PHQ-9. We've endorsed the performance measure that uses the PHQ-9.

Similarly, a very recently
endorsed measure from American Academy of Ophthalmology looks at visual function, and it was the improvement in patient's visual function within 90 days following cataract surgery defined using the VF-14.

Again, we have not endorsed the VF-14. That is a very well validated NIH tool, but we have endorsed the measure that looks at the degree of improvement. So, just to kind of give you that sense of distinction.

It is something that we often find ourselves flipping back and forth. NQF has endorsed CAHPS. Well, NQF has endorsed the tool -- has endorsed the measure that uses CAHPS but we don't endorse the actual tool itself.

So, I think with that, hopefully, I haven't confused you completely. I'm going to turn it over to Patti. So, next slide, Jessica.

CO-CHAIR BRENNAN: Thank you very much, Helen. I want to thank Helen, Karen Pace
and Karen Adams for the work they did to get
us to this point, and Eugene Cunningham
wherever you are for the work you've been
doing.

I am delighted -- yes, can you
hear? Can't hear. Okay. That's never been said
of me before, thank you. I'll try to speak
louder. I do have a Philadelphia slur. I
apologize.

I want to thank Helen, Karen Adams
and Karen Pace, and Eugene Cunningham for
their work that they've done to bring us here.
And I want to thank all of you here and on the
web for your work on the NQF, and particular
in patient-reported outcomes. It's a critical
part of the outcome assessment, and now we're
bringing the patient into the loop.

The purpose of our workshop is to
remind and to expand the idea that patients
are a valid and valuable source of outcomes.

Now, today we're going to be
focusing on identifying some of the
methodological issues related to patient-reported outcomes.

Remember, as Helen first introduced, that there are individual level patient-reported outcomes. How are you feeling today? How much are you able to walk? Can you carry your groceries? Can you play with your grandchildren? These are outcomes the patients may know about and care about, but to translate them into performance measures that tell us about the ability of an institution to provide care requires that we attend to the methodological issues.

Individual-level patient-reported outcomes are valuable to the patient and to the clinician, the individual clinician. They inform the care process, they provide patient feedback and a guide for self-monitoring. And, importantly, they can be contributory towards shared decision making.

We're focusing on the second half of this slide today, the aggregate level, the
performance measure where we're identifying
performance measures that can be used for two
different purposes in health care; first of
all, for quality improvement, to help an
organization, a practice, a group of
clinicians know how to improve. And, secondly,
for accountability, public reporting,
transparency, payment determinations.

Now, we're going to be having a
lot of opportunity for your participation and
feedback today so get ready throughout the
agenda. You see opportunities for audience
engagement, and one will be coming your way in
just a few minutes.

I'm going to turn the podium over,
though, to Joyce Dubow who is our Co-Chair.
I've been delighted to work with Joyce. I
haven't worked with her before and she brings
a perspective from the AARP, and a citizen's
perspective on patient-reported outcomes. It's
extremely helpful to us. Thank you, Joyce.

CO-CHAIR DUBOW: Thanks, Patti. And
I want to add my welcome to everybody. And it's a pleasure to have such a knowledgeable Co-Chair. But, again, thanks to the staff, they've been really terrific.

So, what you see here is the famous bubble diagram. Can't hear? Okay, sorry. Is that better? Okay. I've never had that problem either. Gee, I have to tell my kids.

So, this is the famous NQF bubble diagram that describes the person-centered episode of care. And it is a person-focused model that demonstrates the episode from looking at the population at risk through follow-up care, and the various trajectories that would depend on each individual.

Helen talked about the need to look at patient-reported outcomes with a longitudinal perspective, and this diagram essentially helps us conceptualize and see that framework graphically. And it takes in health behaviors, health quality of life,
functional status behaviors, et cetera.

The next slide, please, helps us identify very clearly what our objectives for today's and tomorrow's -- this is today's objectives, and tomorrow's. This is the -- these are the objectives of the workshop, and they are very clearly presented here.

We want to be able to identify best practices and lessons learned from initiatives that are already underway looking at individual-level PROs in performance measurement.

We want to discuss the major methodological issues related to the selection, administration, and use of the individual-level PROs in performance measures.

We want to discuss key considerations for inclusion of PROs in EHRs, so we need to focus on how this stuff gets integrated into electronic records.

We want to identify the characteristics of individual-level PROs
suitable for potential use in performance measures, and we want to identify the additional set that would be most suitable for further development.

The key here is going to be the interaction and the discussion that this group that we get from everybody -- thank you, there's a good person back there who is -- thank you.

It's very important for us to have an interactive conversation today. Everybody here brings something very important to the conversation, so we want to encourage you to participate and to share your knowledge and your views so that we can move this effort forward.

And with that, we have an opportunity to hear from you. Are there any questions or any observations about anything you've heard so far before we get started with out first panel? Karen?

KAREN: No, I was going to say
let's go ahead and take a few questions.

CO-CHAIR DUBOW: Is there anybody--

(Off microphone comment.)

CO-CHAIR DUBOW: Can you identify

yourself, please? And we can't hear you.

MS. KELLER: San Keller, American

Institutes for Research, and there's an

inherent tension between the longitudinal view

of having the measure standardized over --

patient-centered but standardized over

different applications, and the definition of

the measure at the unit level. So, the units

are going to differ as the patient moves

through the system.

CO-CHAIR BRENNAN: I'm sorry that

engineering degrees doesn't work inside the

Beltway.

If I'm understanding the comment
directly, your remark is that a longitudinal
view of measures is following a patient
through a number of different episodes of care
at different points of care. So,
methodologically you're asking us to consider how different contributors to the care process can be appraised or evaluated by a single point measure -- a single set of measures across a number of points. And I think -- did I get your comment correctly? Thank you.

(Off microphone comment.)

CO-CHAIR BRENNAN: Right. And when you refer to units you mean sites of care as opposed to units of measure. So, I think that's a very important consideration, and that's something we will need to be returning to over the next two days.

CO-CHAIR DUBOW: Anybody else?

MS. LENTZ: Hi, Lisa Lentz, Centers for Medicare and Medicaid Services. And I work mainly on accountable care organizations and physician and group-level outcome measure development. And one thing that we've been thinking a lot about is provider attribution, particularly when we're talking about patient reported outcomes that would be holding
individual physicians and groups accountable
because, of course, the providers are only
seeing patients in the office in a limited
time, and then the patient goes off, and they
adhere or they don't. So, I guess, what I'm
hoping that we can talk about in the next two
days, too, is about attribution, and how we
actually tie those outcomes back to the care
that patients are receiving.

CO-CHAIR BRENNAN: Thank you. If we
could have people go back to this corner area
where there's a microphone, I think that --

DR. GANIATS: This one works.

CO-CHAIR BRENNAN: Good, thank you.

DR. GANIATS: And no one has ever
said that to me before that they could hear my
voice. This is Ted Ganiats, University of
California-San Diego.

I think that one of the issues
that's interesting is that usually we have a
clinical measure that is valid at a clinical
level, and we try to aggregate them into a
performance measure. And sometimes there is a problem because of issues related to are the patients the same from one institution to another, et cetera.

Here we're dealing with measures that may or may not be applicable at an individual level. Many PROs have scoring mechanisms that are designed based on populations of patients and they are not relevant at the individual level, some of them are. So we have that fundamental difference between PROs and most of what makes up performance measures. Plus, we have the other issues that make -- relate to the generalizability, so it will be interesting to see how we can capture that additional element of complexity.

DR. LOHR: Can you hear me with this mic? Okay. I'm Kath Lohr from RTI. And I wanted to go back to a couple of your meeting objectives which are really the last two points, which I think are more for tomorrow.
I was curious to know whether when you say you want to identify characteristics of individual-level PROs, whether you're after sort of criteria for choosing them, whether you're meaning to have a family of attributes for such measures and factors that would sort of be pro or against selecting them. I just want to clarify that that's kind of what you mean when you say "characteristics."

The last bullet is to say pick an initial set, but then I thought I heard Joyce say maybe to pick an additional set, so somewhere along the line today I'm wondering if you could clarify a bit more what you're after, because clearly you have plenty of outcome measures from your earlier slide. And I wasn't sure whether you're meaning to say well, let's find some other ones, or we're starting from scratch.

MS. PACE: I'm Karen Pace. And just -- you're right on your first question that we're looking -- when we say
"characteristics," we're talking about criteria or attributes. We did not use the term "criteria" because we have criteria for our performance measures, and we wanted to make a distinction, but it's -- essentially, you're right. What are the things that would make an individual PRO something we should consider for inclusion in a performance measure?

And it is an initial set. As Helen mentioned a couple of examples, we only have a few patient-reported outcome measures as performance measures. We have many more outcome measures that are more clinical in nature, but we are looking now to identify those from patient-reported outcomes.

DR. LOHR: Thank you.

MR. YANG: Hi, Mr. DerShung Yang with BrightOutcome. And I just want to get clarification on the use of the term "PRO." Are you we referring to specifically only those tools that were assessed by patients
themselves, or are we also including those
that are assessed by providers or care givers
or some other proxies?

CO-CHAIR BRENnan: At the present
moment, the focus is on tool or observations
that the patient individually makes, not an
interpretation by a professional or anyone
else. Thank you.

CO-CHAIR DUBOW: Thank you for that
clarification. It's important. Linda?

MS. WILKINSON: Yes, Linda
Wilkinson from Dartmouth-Hitchcock. I'll be
very interested as the conversation unfolds to
see how we acknowledge the different cultures
and climates in which these measurements are
taken, and such symptoms of cultural behavior
as things like what sort of support is given
the patient to enable them, or to encourage
them to report, et cetera. I mean, I'm sure
these things will come up but that's of great
interest. Thank you.

CO-CHAIR DUBOW: Thank you.
DR. GOODRICH: This is Kate Goodrich from CMS, and this builds off of that last comment, and also a little bit off of what my colleague, Lisa Lentz said.

One thing I'd be interested in hearing about, although I think it may be a little outside the scope of what we're talking about today is the issue of provider buy-in to these types of measures. So, to the extent that people who have -- that we're going to hear about lessons learned. I think it would be very helpful for us at CMS, and also just us within the room to understand how providers, whether it be physicians, or group practices, or facilities, how they -- if they did, and if so, how they developed buy-in into these types of measures. I think that's a major barrier, a major hurdle to the use of these types of measures.

You know, eventually even if CMS requires over time the use of these types of measures we still would like to be thinking
about how we can do that, and how within the
construction of the measure, the
identification of the measure topic, that can
lead to better clinician buy-in.

CO-CHAIR DUBOW: An important
topic. I think we may get to that during the
first panel. We need to wrap-up. I want to
take one more. Who's got the mic? Oh, Albert.

DR. WU: Albert Wu from Johns
Hopkins. Just to clarify sort of the previous
question. So, PROs are obviously from the
individuals themselves, and not the clinician. But are we excluding, for example, parents,
reported measures for children, since -- and
then, therefore, are we excluding all child
measures that are not directly from the child?

CO-CHAIR DUBOW: Proxies count.

DR. WU: Proxies count.

CO-CHAIR DUBOW: It's one of the
challenges, but we need to study that, and the
first paper addresses some of the issues
around proxies, but that's within scope.
CO-CHAIR BRENNAN: Yes. I'd like to make sure that we recognize the origin is an individual who received care. Yes? Speak up more. The focus is the individual who receives the care service, and it may be through a proxy. We'll have to consider issues about family care givers for individuals unable to respond for themselves. And we will be hitting up against the boundaries that might be a little fuzzy. If it's observed by a parent as opposed to observed by a clinician, for example. We'll have plenty of time to talk about that in the next panel, though. Thank you.

MS. PACE: Okay. We're going to move to transition to our first panel, and Joyce Dubow is the moderator.

Just one note from Patti, that we didn't really introduce or give you logistics. Those of you who need restrooms, they're out across the hall through the other doors on the other hallway, and there is coffee in the
back. And we'll also have a break in a little while, but we'll go ahead and get situated.

CO-CHAIR DUBOW: This is our first panel, Acknowledging the Patient as an Authoritative Source. I want to point out that the Planning Group -- is that better? Can everybody hear me? The Planning Group felt that it was very important to start out with this topic for the workshop just to reinforce the importance and the authoritativeness of patient-reported outcomes as a source of important data in health care.

So, not only is the patient a source of information about her own preferences, for example, but also about sometimes a unique source, also about functional status, quality of life, pain, et cetera. Sometimes patients are the only source of information, but we wanted to reinforce that by starting out with this topic.

We think that generally this is a view that's shared by everybody here, but just
for the record, that's just what we had in mind.

We've already addressed the fact that patient-reported outcomes have multiple uses in addition to an expression of preferences, for example. Patient-reported outcomes can be used in quality improvement, public reporting, payment programs, so they have important functions and play an important role in the measurement process.

We've just heard that there are challenges to taking some of these validated clinical instruments and making them into valid performance measures. And that's what we need to tackle with. Today we have a panel that represents important perspectives in addition to the author, who's going to give us an overview, the author of the first paper who will give us an overview of the range of topics that the paper discusses. The three perspectives coming from the disability community, a provider perspective, and a
purchaser perspective are very important.

What we want to do is to tackle these challenges, address them so that we can move forward. If there is agreement that this -- that patient reported outcomes really represent an authoritative and valid source of data, we have to figure out how to move to the next steps. And that's what we hope to achieve.

So, we have, like I say, we have a very, very talented panel. Their bios are in the materials that you've received. We're going to start out with an overview from the author of the paper, David Cella, one of the authors. And we're going to go from there to Charles Moseley of the National Association of State Directors of Developmental Disability Services, Steve Fihn from the VA, and Jennifer-Eames Huff from the Consumer Purchaser Disclosure Group and PBGH.

So, David, will you begin, please.

DR. CELLA: Good morning. Thank
you, Joyce. Do I ask you to advance or can I advance from here? Okay, there's the first one. Thank you.

So, you've all been sent a paper, a draft of a paper, and I think I had a list of contributors to the paper before this slide. Is it not on this set? Okay. There we go. I really want to show this first.

To those of you who said congratulations on a nice draft, these people deserve as much or more credit as I do. I get to be the one standing here, but Beth Hahn, Sally Jensen, and Zeeshan Butt, Cindy Nowinski, and Nan Rothrock all contributed a lot to this paper, so I want to acknowledge them. To those of you who think the paper is terrible, I'll take the full blame for that.

Next slide, please, or two slides.

So, I think the first question -- and I know whenever I work, which I don't often do, but whenever I work with people in the performance measurement field, the first
question I get either explicitly or implicitly is why can't we just ask the clinicians? It would be a lot easier, a lot cheaper. We know our patients, why can't we do that?

Apart from the issue of possible bias and conflict of interest in asking doctors how their patients are doing when they're going to be paid based upon their answers, independent of that in the literature, research literature there's a vast amount of support that demonstrates that clinical providers unfortunately don't accurately capture outcomes that are only logically obtained by direct patient query. Certain symptoms, certain functional areas, I'm afraid to say you have to ask the patient.

Next slide, please.

This is actually some work from Ethan Basch, who's one of the moderators later this morning just showing -- it's looking at various symptoms associated with cancer treatment. In the middle is agreement, that's
the light bars. On the left is where the patient says it's worse than the doctor. On the right is where the doctor says it's worse than the patient. You get the impression here pretty consistently that the error is in favor of the patient acknowledging more problems than the doctor or clinician seems to realize or report on adverse events.

Now, these are people being asked to rate the same thing, fatigue, pain, et cetera. So, there is bias, and the bias tips toward patients identifying more problems and more issues than providers either are aware of or report. Next slide.

And also, there's also work in cancer done by Deb Bruner. When you look at the correlation of adverse events, which is what Ethan's study looked at with quality of life, that is the broader sense of well being and functioning, there's very little correlation between the symptoms that patients have associated with treatment or their
adverse events and their overall general functioning and well being. So, for several reasons it's really important to get this information from the patient. And I realize this may be the choir, but it's an important choir. And if I can help with this refrain please let me know. Next slide.

So, what's the potential for PRO use in clinical care? Well, there are many which I'll run through. You can assist providers in care management, you can enhance efficiency as opposed to the myth that you actually interfere with efficiency, you can improve communication. This is usually at the top of the list, identify patient needs in a more timely manner, sometimes being able to intervene more quickly and prevent problems, and facilitate an atmosphere of patient-centered care.

Despite all these possible opportunities or advantages, routine care assessment is still not common in clinical
practice. Next slide.

One area that's gotten pretty common is the patient experience of care, largely I think because of the endorsement and paying processes around getting information on the patient experience of care through CAHPS and other measures. So, broadening from patient satisfaction where there are questions like did your doctor seem to understand what was important to you? Were you satisfied with your visit? There is an extension in CAHPS to things like reports of actual experience. For example, in the last 12 months when you phoned the provider's office during regular hours how often do you get an answer to your questions that same day? So, a very and much more focus on the experience of care in a very drilled down way, which I notice the all blue -- I think it was the only all blue bar in that figure that Dr. Burstin showed. Next slide.

Clearly, this is an area where we're getting the information from patients
now. So, this concept of patient-centered care, originally Epstein's concept, I think has caught on in the minds of many, and as implemented some places better than others, involves a partnership between the informed and activated participatory patient and family member with an accessible and organized responsive health care system that produces better, more patient-centered and oriented communication, that then logically would improve health outcomes. So, this is the model for why we would think we would want to do this. Next slide.

Now, there are barriers. They've been alluded to, and we're going to start to talk about them throughout the day. Some of the current practices or best practices to minimize self-report barriers including selecting an appropriate method and mode of administration for your context, doesn't mean that there's one-size-fits-all. In fact, that's not the case that one-size-fits-all.
It's important to consider things like the age of the patient, the functional status of the patient going into an episode of care, especially when you consider longitudinal evaluation over time, and the cognitive capability of the patient, whether because of age or because of disease or disability as those relate to your likely need for use of proxies and assistive devices in helping the patients provide information on their own behalf.

In designing instruments, people that use universal design principles that are published and available tend to produce better instruments. There are accepted and approved methods for translating and culturally adapting questionnaires so that they are more likely to produce valid results across different groups whether by language or by culture, or by reading level. And you can produce equivalent versions across these very important sociodemographic and cultural...
differences that our patients manifest.

There needs to be flexibility in
the location from which you get the
assessment, and sometimes there are
differences based upon the location, white
coat hypertension, for example, versus
influences at home, maybe someone is cheating
and having a family member help them at home.
These are all issues that come up when you do
things -- the assessments in the clinic, at
the home, or at some facility, say a nursing
home or hospital.

It requires access to the
technology. If you're choosing to use an
electronic technology, then people -- if you
choose to do it at home people have to have
internet access, or some ability to get to
that technology. Or if it's telephone
technology, they have to have a phone.

And in every setting, particularly
the at-home settings health information
privacy and security have to be considered and
protected.

It's also important within that to address functional literacy, and health literacy. They're somewhat distinct, and they're critical really to delivering person-centered health care. The next slide has an illustration, a diagram of that.

You can divide literacy and technology skills, and consider patients not just in terms of their literacy by doing a rapid literacy test, but whether they have oral literacy; that is, being able to listen and hear what someone is saying, process oral information and speak back in conversation, written literacy, and reading literacy. And even reading literacy is divided into PROs, that is being able to read some text and understand what it's saying. Document literacy, can you sort out figures and graphs. You know what percentage means? Do you know what probabilities are? And quantitative literacy, as well, which gets into this
1 numeracy issue.

   And then on the technology side,
2 if one moves as I would personally advocate to
3 an electronic environment whenever possible,
4 we need to make sure that people have
5 appropriate computer skills or other media
6 skills that might help give them assistance in
7 completing questionnaires, perhaps by reading
8 questions out loud to them if their oral
9 literacy is better than their reading
10 literacy. Next slide.

   So, continuing on with practices
12 to minimize barriers to self-report, there are
13 some circumstances where it might be
14 difficult, or even impossible to directly
15 obtain the assessment by self-report.

   We suggest in the paper that proxy
18 reporting, though it does have problems, can
19 be useful. It's really important if you want
20 to be able to be inclusive to include people
21 with cognitive or communication deficits, or
22 severe disease burden who can't speak for
themselves or respond as to how they're doing, not to exclude them from the picture. And for people who may be able to respond for themselves but they may be in early stages of dementia or malcognitive impairment who might not recognize their impairment as it's evolving, and yet a proxy, a family member would be a good source of that information.

And, finally, for young children who are not yet sufficiently reliable to report the kinds of health status things that we want to capture in performance measurement.

Next slide.

So, that's my introduction. And I think I'm supposed to sit down now and have the experts react.

CO-CHAIR DUBOW: Are there any specific questions about anything? We don't want to have the conversation now, but if anybody has any specific questions about the presentation? Okay, Charles.

DR. MOSELEY: Thank you very much.

It is, as I mentioned, a pleasure to be here. And I was really interested to see the focus of the discussion on the slides highlighting the real need to address indicators that will improve quality, and provide accountability.

I'd like to add one more to the list, and that is to produce indicators and information that's actionable, that can be used by policy makers, by service providers, by family members and others to really make a difference in the lives of the people who are receiving the support.

As I begin today, it's interesting, I feel a little bit like a duck out of the water. We don't use the term intellectual -- in the field of intellectual and developmental disabilities, patient to refer to the folks that we support. We
typically use people receiving support or just people. And I think that's a very important difference because it does not distance the individual from the services and supports that they are receiving, like the word "patient" does.

What I tried to do is organize my comments today around the three particular areas that you asked me to address; how do we best build a value proposition for clinicians and policy makers that individual input is credible?

I think it's important to recognize that the nature of the services provided -- excuse me, important to recognize the nature of the services provided in the populations who are receiving them. Acute care services, for example, typically are time-limited and measured narrowly focusing on a treatment regimen or course leading to some type of cure or amelioration of a condition.

For people who are aging, long-
term care typically refers to support through
a nursing facility which lasts on average
about two and a quarter years per person
according to the CDC.

Many people with disabilities by
contrast receive supports throughout their
life span. The majority of the over one
million people with intellectual disabilities
and developmental disabilities currently
receiving publicly funded supports, for
example, enter the system following school and
continue to receive support throughout their
life times. Many never utilize a nursing home
at all.

Indeed, a disproportionate share
of the support is provided by family members.
Of all people currently receiving publicly
financed services right now, 57 percent
receive them within the home of a family
member.

Finally, it's important to
recognize that the various demographic and
need profiles of Medicaid beneficiaries with
disabilities are incredibly diverse. As noted
by the National Council on Disability and
Managed Care Principles, the type of services
and supports required by an 85-year-old widow
with advanced Alzheimer's disease are entirely
different than those needed by a teenager with
significant behavioral or communication
challenges caused by autism or other serious
neurological disorder, or an adult with
intellectual disabilities who has co-occurring
mental illnesses.

Each may require specialized
medical and prescription -- medical services
and prescription medications in combination
with ongoing personal assistance, but the
composition and competencies of the team
assembled to deliver those services will be
radically different in each case, as will the
types of medical, psychological,
pharmacological, and social intervention
services that are deemed to be appropriate.
What is it that people need? And how do we know that the services will result in the outcomes that are desired? People with disabilities, just like all of us, need direct support to access the community, work, families, and friends. They need training and assistance to enable them to learn the skills they need to function as independently as possible, and to direct their own services.

They need assistance in accessing appropriate health care, therapies, through service coordination and case management, and they need ancillary services, transportation, interpreter services, and a whole wide range of other supports to enable them to fully participate as members of society.

Long-term supports are as personal as taking a shower, eating meals and getting dressed. They're also a matter of public policy. So, how do we develop value proposition for clinicians and policy makers that takes input from people receiving...
supports?

It's important that we develop, I believe, population-specific indicators that are meaningful in the sense that they add value to people's lives, and to state policy, and individual practice; that they're credible, that they address areas of importance to service delivery and the achievement of individual outcomes, that they're valid, measuring what they're intended to measure, reliable, that they produce consistent results across interviewers, raters, and over time, and representative, they're based on a representative sample. And, finally, they include questions that are risk-adjusted so that you can compare -- identify trends and compare state-to-state data, and as I mentioned, national trends, and by addressing broader outcome and performance variables that are relevant to individuals receiving support, state policy and funding decisions.
And we're talking about indicators that measure access to appropriate health care at a point in time, and over time, identifying the number and percentages of people who are receiving -- of people receiving support who are working and are accessing employment, and stay on the job, documenting the percentages of individuals who choose where and with whom they live, and who they spend time with during the day, identify individual and service-related choices that a person can exert over the course of the day, and by tracking the extent to which measures produce data that are used by policy makers, actually used by policy makers, practitioners, individuals receiving support, and researchers.

Now, the National Association in collaboration with the Human Services Research Institute has developed and implemented the National Core Indicators Program which gathers individual-level data, systems data, across 35 states, roughly pulling in information,
individually reported data on people's perspectives on the places where they live and work, the amount of choice that they're able to use, the activities they engage in during the day, the nature of the experiences that they have with the supports they receive, and their individual characteristics.

Currently, we gather information on about 20,000 individuals each year. We're working with a grant from the Administration on Intellectual and Developmental Disabilities to expand the core indicators to all 51 states over the next four years, 50 states plus the District of Columbia. Here I need to really say 51 states, I think.

How do we insure that PRO data is useful to patients as well as other users? This is a really important issue, and I think it's very important to identify and utilize measures that reflect and assess what is important to the person, and what is important for the person.
There are two separate perspectives there that are both valid, they're both important, but for too long we have identified only what's important for the person as determined by someone else, whether it be a clinician, a family member, a guardian, or others. And it's really critical to gather the information from the person receiving supports.

And we certainly have found over the past 14 years in gathering our data that individuals with intellectual disabilities can very easily, the majority of them, report on information about the nature of the supports that they receive, their choices, their outcomes, their goals, and their life expectations.

Provide regular -- the data set really needs to provide regular user-friendly reports summarizing key data trends and issues. And as I mentioned, the core indicators that are now being used by 35
states are being used in a number of different ways.

I recently surveyed the directors of the state agencies supporting these individuals across the country. They said that they used the data to meet CMS and HCBS waiver quality assurance requirements with respect to the plan of care, family involvement, and health and welfare.

Now, the core indicators are important to mention that these are system measures and cannot be used by themselves to really assess this kind of information, but are best used in combination with other information that is gathered by providers and others.

Formulate key policy positions with respect to the kinds of services that should be delivered, compare performance measures and outcomes by diagnostic groups or across key service areas, residential size, for example, employment and health access,
benchmarking system performance in key areas against that of other states, and providing information on key systems variables such as the impact of facility size on quality of life, loneliness, and community access, the extent to which people are able to control and direct the services that they receive, access to employment, improved choice and access to regular health care.

Finally, what are best practices to minimize barriers to individuals being able to self-report? We found that to a great extent you have to go back to the survey basics. Surveys need to be well-constructed, that can utilize alternative methods of data gathering, not only reading a survey questionnaire but having questions explained to them. We found, for example, that iPads are terrific for people who have communication problems because they can zoom right through them and are pretty much in control of the information. Trained and well-supervised
interviewers, the availability of people receiving support to participate in the process.

We talked earlier about proxy respondents. They're very good; the responses that they give are, however, different, and it's important to note those differences even though you may include them in part of the data set.

The development of sound and practical interview and survey administration protocols, data analysis methodologies, and data entry processes, targeting the right people who have access to the data needed, providing consistent and appropriate methodological approaches for analyzing and reporting the data, and providing processes for releasing the data to researchers, demographers, and others to document usage and trends.

I want to loop back around to just mention that it's really important that the
data be used. We're very good, and I think across the health care field at gathering a whole lot of information, producing a lot of beautiful reports and leaving it on people's desks. We need to figure out ways to really drill into the information and use it to change practice. Thank you very much.

CO-CHAIR DUBOW: Thanks, Charles. I think that your point what we call these things is well taken. We are talking about persons, individuals, but for the sake of our discussion today, forgive us if we slip, but I think everybody recognizes what you're saying, and we appreciate that. Steve.

DR. FIHN: Good morning, and I'd like to thank all of you for inviting me here as a representative of the Department of Veterans Affairs.

As many of you may know, the VA has been undergoing yet another major transformation over the past couple of years under the direction of Secretary Shinseki, and
the sort of three tenets of that transformation have been that we are Veteran-centric, that's our word for "patient," results-driven and forward thinking. So, the notion of patient-centricity I think is central to our system at this point, so very appropriate to have this discussion.

And along these lines, for example, VA is investing about $1 billion, for example, in developing patient-centered medical homes at the thousand sites that we provide primary care.

And we do collect information from patients along the lines of PROs. For many years, for example, we've conducted a survey we call the Survey of Health Experiences of Patients of the Shop, which includes patient experience, the SF-12V, that's collected on about 600,000 people a year in a very highly scientific survey methodology.

We've recently started the U Speak for our rehab patients. We have detailed
patient-driven recovery plans for severely injured veterans from Iraq and Afghanistan. We collect PIMS on all our primary care patients. We collect Audit C for alcohol use and the PSQ for depression.

That said, actually, from my view point that's a relatively limited amount of information to get from patients. And we do have plans which we're working on actively now to expand that repertoire, so for example we're developing mobile health platforms with actually a VA App Store, and some of the early apps will be one for pain measurement, another for PTSD symptoms.

We are going to release a health risk assessment connected with our patient portal, which is called "My Healthy Vet." And we're actually experimenting with some condition-specific measures particularly related to ischemic heart disease to measure the outcomes following elective percutaneous coronary interventions.
So, I think from the VA's perspective, and I'm speaking largely for myself here, I think this is a good direction and applaud these efforts. But I think also from the perspective of a system that has over 6 million patients, we've also been acutely cognizant of some of the limitations and difficulties which I hope will be considered in this process.

The paper, David's paper I think eloquently outlines many of the concerns that are fundamental, technical issues, bias, problems with performance, respondent burden, interpretability of measures, privacy; and in our system that's multiplied by 6 million people, so these are not, I think, problems necessarily to be minimized.

And I'm going to reflect for a second on my own personal experience. I reconnected with one of my close colleagues, John Wasson, and recounting a study we collaborated on actually back in the early and
mid-'90s in which we actually randomized
30,000 patients in the VA to feedback of PROs,
both generic, as well as condition-specific to
primary care providers over a two-year period,
and failed to show, actually, any clinical
benefits in doing that.

More recently, we actually
consducted another multi-center randomized
controlled trial which we linked the use of a
condition-specific PRO to well-defined efforts
to intensify therapy for patients with
ischemic heart disease.

We've also done similar things in
heart failure, and in COPD. And none of those
cases actually have we demonstrated a clinical
benefit. And, in fact, in one of those cases,
actually, there was, as reported in the New
England Journal of Medicine, clinical harm.

And can talk about why that might be.

So, I think I would say that as an
organization, and as an individual, I think
I'm quite committed to this notion of patient-
driven care. And I think Don Berwick has, basically, eloquently and very succinctly put that, is that the patient should drive the care. And the patient can't do that unless we know exactly what the patient wants. So, that's clear.

I think jumping from that precept, though, to the sort of notion of performance measures and mandated instruments is a big jump and one that needs to be taken with care and thoughtfulness. And I'm delighted to be invited to be part of this group. I think this is the right group of people to start addressing and confronting those issues, and will be keen to participate.

CO-CHAIR DUBOW: Thanks, Steve. I think your insights about why some of these efforts didn't work would be very helpful. And I hope we can get to that maybe during the question and answer. Jennifer.

MS. HUFF: Good morning. Can you hear me back there? Great. I'm also, along
with everybody else, really delighted to be here today. It's really I think for me personally very exciting to be having this conversation. Early in my career, I had the benefit of working at the Picker Institute, so this is an area that's near and dear to my heart. And the mission there, which I think is very apropos to our discussion was just to see care through the patient's eyes. And I think that's what we're trying to do with the information we're doing here.

It could also be akin to walking in the shoes of the patient, you know. And it brings for some of us that aren't as close to a clinical encounter a real humanness to the work that we're trying to do in terms of improving care.

David Cella had talked about the benefits of pros in clinical care, and I'm just going to focus on one aspect which we've actually heard underscored by the other panelists, which is really getting to patient-
centered care and the growing recognition that we really want to be truly about the patient. And to do that, patients need to be a part of the process.

Asking them about how they report on outcomes and experiences, as well as engaging with them on the results and the interventions is an important piece to creating this patient-centered environment.

And I think we all recognize that PROs also offer a valuable perspective that can't be obtained for other -- from other data sources. In fact, we could even argue that the patient is an expert on some of those areas, like the level of pain they're feeling, their functional status, how their care is being coordinated. And it's not meant to replace other data sources, but it's a good complement to the other information we're gathering as a part of performance measurement.

And I think we'd also say it's important to say that PROs have been
increasingly shown to be linked to improved clinical outcomes. Patients that have more positive experience tend to get better outcomes we've seen in some studies, which is a compelling argument, I think, for all stakeholders in terms of using this information.

And I think we all agree with these particular statements that I've just said, and it creates value, but it's not, necessarily, enough to move the system in terms of using patient-reported outcomes. And for that, I think we have to get really practical on the information that's useful in terms of the value proposition. And that's really creating an evidence base that shows how the use of PRO measurement programs can save money, improve care, engage patients in the care, and then incorporating that information as we gather the evidence into a variety of uses that were talked about, payment programs, public reporting, say,
maintenance of certification or accreditation. I'd also add clinical registries. There are other places where this information could be really helpful in terms of improving care.

A good place to focus initially would be looking at places, and I think this creates part of the value proposition of where there's a lot of evidence of inappropriate care. For example, back pain is a good example of that, or with patients that have multiple chronic conditions.

I'll give an example of a program that the Pacific Business Group has been involved in, and it will show, I think, not just from a purchaser perspective but a variety of perspectives the value proposition I was talking about in terms of the different elements.

PBGH members are using PROs in a program that is called the Intensive Outpatient Care Program. Some of you may know this, as Boeing Corporation did a pilot on
high-risk, high-cost patients. It's essentially what I'd call an ACO for a really chronic condition high-risk population in the primary care setting. It's a care redesign model that includes a dedicated staff person for the intensive primary care management. They do a case rate per member per month to cover non-traditional services, and as also a shared savings component.

What we found in the Boeing pilot is compared to non-participating -- a matched population, there was a 20 percent spending reduction in the program and, as well as, which is where this is really important, they used the SF-12, and the PHQ-9, and they found both improved physical and mental functioning from baseline, and patients reported access to care improved since the baseline, and they also saw a decrease in absenteeism, again as reported by the patients.

PBGH is spreading this model both in Northern and Southern California, and it's
been much easier to spread this model with
this compelling evidence. I think this
evidence really brings the value proposition
to the purchasers that we're working with.
We've been working closely with Humboldt
County. Some of you may be familiar with it.
It's an Aligning Forces for Quality site. And
they've also added using the PIM as a part of
the -- in addition to the SF-12 and the PHQ-9.
And they incorporate it into the patient's
action plan, so they have this information.
They use it in terms of how the care will be
delivered to the patients, and they are
regularly assessing over time how these things
change.

And I think this model has
something for everyone. It's improving care.
It's being more affordable. Patients are more
satisfied with the care they're getting. It
has patient engagement, and the clinicians are
also really engaged in using this information.

In terms of the other area we're
asked to talk about, the usefulness of
patients, I would agree with Charles that it's
really important to use information that's
important to the person, and for the person.
And I also would say I think it's really
important for us to consider how the PRO data
will be used: will patients be using it in
shared decision making; will they be using it
in their treatment decisions or selecting a
provider; because those actions have different
data needs. So, really tying it to what the
use is is important when we're beginning this
path.

And I'd do another plug for
patient-centered engagement. This is making it
useful to patients. This is a place to engage
patients or their representatives as a part of
that process.

The other thing I'd add in terms
of useful to patients, which we've seen with
other performance information is the timing of
when this information is given of really
having those opportunities when the patients need the information, or are making the decision.

For best practices in collecting PROs, I think integration of the PRO systems into clinical care, making it a part of the clinical process, using patient-reported modules and patient portals that are convenient, and I'd also add if there are ways to use cell phone technology in terms of making this really accessible.

And also, I think, with the technology, I think David talked about some of the challenges with using that, so also recognizing what are some of the barriers that would come of that, but it would be a great place in terms of making the information much more available.

And then I'd like to conclude with an experience I had when I was working with a doctor who was an oncologist and primarily treated breast cancer patients. And he decided
that it was really important for him to be tracking the SF-12 looking at the physical and mental functioning of his patients, so he bought a machine that he put in his clinic, and when patients came in for their visit, they filled out the SF-12 questionnaire. It was much like educational testing where you fill in the dots. It has all these dots, and there is a machine where you could put the card right in there, and it spit out the report in terms of the results. It would calculate it.

And there is also a way that you could download the information on a disc, and do different analyses, but it was real time. And then that was given -- the nurse took it and put it in the chart, and it was something that the doctor could use as a part of his visit right then and there. And I'd just like to add that was 17 years ago.

So, I think one of the things you'll hear purchasers saying, as we're always
saying there's an urgency to move things forward and not let some of the real nitty details keep us from making progress. And I think that's a great example of, "We've been doing this for a while." I'm really excited to be here and start talking about ways to really move this forward.

CO-CHAIR DUBOW: Thanks, Jennifer. So, maybe we can conclude that this is an idea whose time has come. And, you know, Jennifer, I hope that during the conversation, and we have about a half hour. Is that right, Karen? That you'll have a chance to give us some insight on how we can get other purchasers to be thinking about asking for PROs to be incorporated into their measurement strategies, as well. PBGH is a fairly enlightened and sophisticated purchaser on behalf of other companies, but it would be really useful to think about how to spread the word to those who aren't quite.

With that, we have plenty of time
for questions and answers. I think that the
panel has given us lots to think about, and I
urge you to just come forth with your
questions.

MS. PACE: And I think what we'd
like to do during these discussion sessions,
if we could ask the expert panel first to make
their comments and questions, and then we'll
go to the audience that's here in person and
on the phone, and then we'll have to ask the
operator to open up the phone lines when we're
ready for that.

CO-CHAIR DUBOW: Does the panel
want to --

(Off microphone comment.)

CO-CHAIR DUBOW: Right.

MR. CUNNINGHAM: And just a
reminder, at each roundtable is an individual
microphone that you can all use, as well.

(Off microphone comment.)

MR. CUNNINGHAM: Oh. At each
roundtable we have --
CO-CHAIR DUBOW: I just wanted to be able to say that.

MR. CUNNINGHAM: At each roundtable there is a microphone to use, as well. Just push it on until it’s green.

CO-CHAIR DUBOW: Is that Al there? I can’t see. Al Wu.

MS. PITZEN: Okay. This is Collette from Minnesota Community Measurement, and I'm sure we'll get to this later, but I just wanted to make sure that we address the use of the validated tools and potential copyright issues with those developer holders.

For example, we're working on a couple of orthopedic measurement sets with definitive charge to access functional status. We obtained permission from Oxford University and the EQ-5D, but in a very narrow scope. So, we can only use those tools for that measurement. And as hospitals and clinics are starting to implement these functional status tools, they would like to use them a little
bit broader than just spine fusion patients.

So, we are continually running into those complications in terms of copyright. Thanks.

CO-CHAIR DUBOW: That's a good point. Do you want to talk about the proprietary issues around that measurement, Helen?

DR. BURSTIN: Sure, I could try.

Oh, that works. It's a great question, and it's something that's come up. Actually, numerous tools have been proposed to us, and people have sort of backed away because of the issue around cost. So, NQF does have a carder that allows proprietary measures to come forward, where part of what is revealed, the measure has to be fully transparent, the Committees have to be able to review it fully, but there is an opportunity to include the charges as one of the considerations under feasibility.

We have not had very many measures make it through that way. Actually, recently
a couple of ICU measures did because it truly
was the only tool around to look at pediatric
ICU care.

But it is a real issue, and it is something that does limit what's available.
And I think if there are opportunities, and I'd be curious to hear, you know, David Cella's perspective from NIH as some of those sort of those sub-elements of those tools may get looked at if there are ways to take some of those building blocks and build them into measures, perhaps, rather than the whole tool.

CO-CHAIR DUBOW: Al.

DR. WU: To save time hereafter, I'm going to go as Al Wu. I think that we all agree that the patient is the authoritative source. And it struck me as we are thinking about -- as we are discussing this, one question which we probably need to ask, which is a little bit of a measurement question, is which the patient, becomes sometimes patients are too sick to answer for themselves. Someone
else responds for them. They then at a later point in time respond. Those two measures are supposed to represent the same person, but we now to figure out how to combine them. Some patients may report things multiply over time. How are going to use that information? Are we most interested in the state of the patient at one point in time, or are we measured in a changed measure, or are we interested in some area under the curve?

So, I think there are a number of issues that we need to think about which sort of relates to which patient, at least at the individual level.

CO-CHAIR DUBOW: Greg.

DR. PAWLSON: I'll agree. I'm not color blind. I actually tell red from green. I had the privilege last night of, I guess, that you'd call it that of spending about six or seven hours on an airplane, and I had a chance to really re-read the paper in some depth. And when I got home and my 5-year old
granddaughter who lives with us engaged me in doing a puzzle with her. And those two experiences together sort of really helped shape some thinking around acknowledging the patient as an authoritative source.

I think one of the things that would really help in this work is to -- and specifically the paper, is to really start with a little bit broader context.

We normally think about it as helpers of a person who needs help, and I think we need a new phrase, because persons is a little bit too broad. But there are people who need assistance of some kind or another, input, and those people are the ultimate source of information about what is going on with them, what they are experiencing, what they think they need, and their outcomes.

And I think that really reinforcing that spectrum of where the patient or where this person who is needing the help is not only the authoritative source, but is
the critical and only source, and then how
that plays out in the interaction with the
persons trying -- the care givers, the people
that are trying to assist that individual, I
think it would really be helpful.

It gets back to -- because I think
it puts into context for clinicians of how we
gather information, and when we gather it, and
when it's really important. And that patient-
reported outcomes are one piece of the puzzle
that is absolutely critical, and has been
lacking, I think, a good deal, but is really
in this whole context of how the person we're
trying to help has to be the ultimate source
of a lot of the key, and then how that gets
played out in this new electronic age. Because
I was also watching my 1-1/2 year old
granddaughter use my iPad, which was very
impressive.

CO-CHAIR DUBOW: Wow.

DR. TINETTI: Mary Tinetti, Yale. I
just want to make one comment about
differentiating patient important outcomes from patient-reported outcomes. And several people have alluded to it, but I think I would like it to be explicitly on the table that a lot of the measures are what researchers think are important. I just want to clarify that.

The second point I want to make is I think I don't want to dismiss too quickly the clinician reported. I think we need to differentiate clinician ascertainment of patient-reported versus clinician reported, because ultimately we want to do with quality improvement is to improve care. And if we have separated what clinicians do versus what we ascertain from patients we're not going to accomplish our goal.

So, I really want to make sure, because I -- both in sort of alluding in the discussion today and an excellent background paper, I think it almost too much dismisses the clinicians that still need to be at the center of the relationship between the patient
and the outcomes, so I really want to make sure we differentiate those two points.

CO-CHAIR DUBOW: That's very important. Patti Brennan keeps making that point over and over again, and I think we really need to keep it in mind.

DR. BASCH: Hi, there. Ethan Basch from Memorial Sloan Kettering Cancer Center. I just wanted to bring up the issue of context of use of measures, because many of the measures that have been discussed are actually generic measures, but as we start to think about really getting very granular about evaluating performance, some of the approaches will really have to take into consideration the context in which patients live, and what they're experiencing.

And if we look at the regulatory context in which many of these measures have been used for many years, context of use or fitness for purpose is really central to the development of a measure. Demonstrating that
the measure being used is appropriate to the patient population is meaningful to that patient population. And generally speaking, this is based upon up front qualitative research demonstrating that what is being assessed is meaningful in that particular population. So, I hope that as we move forward we will keep sight that as we develop patient-reported performance measures, that it's a whole package, and it's not just presenting a measure, but demonstrating that in the context of use for the population of interest that measure is actually meaningful and can measure something the patients care about.

DR. GANIATS: I love you all. I hope you still love me. I'm Ted Ganiats, again, from San Diego, University of California-San Diego. And I will play a bit of a role of a curmudgeon, I guess, because I'm going to challenge the statement that patients are an authoritative source. I don't doubt that they can be, but are they?
And I say that because as a clinician I'm able to sit there and listen to the patient, and the one who comes in with a positive review of systems, who has a positive serum porcelain level for whatever reason, I can then dismiss or partially dismiss, but when put into a performance measure we lose that ability. And are patients of that ilk equally spread among all practices? Then I don't have to worry as a performance measure, but if not, there's a problem. And we know that they're not equally spread.

Are men and women equally likely to respond to a given problem? We know that there are gender differences. If there are gender differences, we have a problem using it as a performance measure. Does mood affect the report of a patient-reported outcome? I believe it does. I think people around the table do.

I am not -- I mean, I'm a family physician. I actually use this stuff, I
believe in it. I'm a strong proponent, but we have to remember the limitations of the patient-reported outcomes, and not assume that just because it's from the patient it's automatically authoritative. I say that with love and respect.

CO-CHAIR DUBOW: It's always necessary to have a curmudgeon. And we appreciate the push-back because, clearly, there are challenges to implementation, and we need to be able to address those. But I don't think you actually challenged whether the patient is an authoritative source, as you simply identified some barriers that we need to address when we get to the measurement part of it. So, you're only a kind of quasi-curmudgeon, and you have to work harder.

(Laughter.)

(Off microphone comment.)

CO-CHAIR DUBOW: Okay.

DR. MOSELEY: I have a quick comment. I think you're raising an important
issue, and it gets back to the notion of context that you've talked about earlier, because we have certainly found as we gather individual responses on people's perspectives on their quality of life, that that changes. It changes with respect to several variables, not only the level of the person's disability or the particular life situation, but also changes with respect to people who are working versus people who are not working.

People who are working clearly are demonstrating more choice over the services that they receive, over the people who come into their lives, over the structure and functioning of their individual support plan.

So, I think it's important to kind of look underneath, just as you suggest, the data to see which group is being representative, and what are the various other variables that could come into play.

CO-CHAIR DUBOW: Steve.

DR. FIHN: Yes, I was going to
respond also. We've actually looked at the geographic distribution of health-related perceptions in our system, and there are huge geographic differences. A good example would be in the Southeast, health status is much worse than it is in other parts of the country. And it, obviously, closely correlates with socioeconomic status and other health-related conditions.

So, again, if we were to sort of use this as a performance measure without some sort of adjustment, we would arrive at some probably erroneous conclusions.

DR. CELLA: This is Dave Cella again. So, I see this particular discussion as, for me, at least, the most important thing for me to take away from the meeting, and it comes back to Dr. Fihn's initial comment about the jump from patient-driven care to performance measurement mandate. And making that jump, I see this group as the group that can help that jump happen.
I guess I -- what I was saying was that I see this group of experts and panel reactors, reactor panels, I said in a conference call that reactor panels sounded like I was getting in front of a power plant.

(Laughter.)

DR. CELLA: And it is sort of a power plant, I guess. But I think this is the -- to me, this is the rub, this is the core of what our challenge is to do here. And on one level you can very simplistic and say well, I'm not sure why patient-reported outcomes are different than any other outcome that has to be risk-adjusted. Maybe you could argue there are more things to adjust for because of culture, and language, and things that go into patient-reported outcomes, and that might be true. But I don't think, in my mind, at least, I can -- I'm here to be educated, that it necessarily is fundamentally or qualitatively different to consider how we adjust patient-reported outcome scores to do fair
comparisons, just as you adjust any other outcome across providers. But I'd like to hear what you think about that.

CO-CHAIR DUBOW: We want to open this opportunity for questions and answers to the audience, as well, as well as to the people on the phone. So, can we ask the operator to open up the -- are there people b-

OPERATOR: At this time, ladies and gentlemen, if you would like to ask an audio question please press *1 on your telephone key pad. We'll pause for just a moment to compile the Q&A roster.

MS. MASTANDUNO: Good morning. I'm Melanie Mastanduno from the Dartmouth Institute. And I'm going to echo something one of our experts, and that is going back to survey basics. And thinking about the response rate among the patients who are eligible to report these measures, whatever instrument we are using.

And I've had the pleasure of
visiting six different sites that are doing some form of patient-reported measures on the ground, and looking at their work flows, and finding out their challenges, as well as what's working well.

And two things that haven't been mentioned this morning; one is the positive attitude among providers when a patient does provide their perspective, and wanted to be acknowledged, thanked, and somehow integrated into that provider-patient conversation as an essential key ingredient.

And the second is the level of trust some patients have for computers, for example, or using a technology when their whole social and socioeconomic circumstance has not permitted them to be really power surfers. And this is a key way we'll collect data. And before we even get to Smart Phones, this is a real cultural barrier from the perspective of accessibility. Thank you.

DR. ROSS: Hi, I'm Clarke Ross. I'm
a new member of the MAP workgroup on persons
duly eligible for Medicare and Medicaid. And
I represent the Consortium for Citizens With
Disabilities, which is a policy consortium,
voltune consortium of 113 disability
organizations.

I wanted to throw out a couple of
paradigm challenges, things that are happening
with Medicaid dollars, both Medicaid managed
care dollars and Medicaid home and community-
based service dollars, and just keep the
thought in your mind as we work through the
two days. And this is the direct empowerment
of people with disabilities.

The payment of dollars not through
an agency, and not through an organization,
and not through a provider, and not through a
professional, the payment of dollars directly
to people with disabilities to make their own
purchase decisions.

We have in several states what's
called Cash Counseling programs. These are
people who employ their own personal
attendants, for example, and make those
choices. And then in the area of mental
illness, we have four states who have financed
independent third-party consumer and family-
operated community-based organizations that
monitor services and engage individuals
directly, service recipients directly. And
these are all people who are in recovery
themselves from mental illness.

They are people with a history of
mental illness who have been trained to engage
their peers. So when we just talk about the
complexity of what we're talking about, these
are models that are actually operated,
financed, and have a lot of momentum behind
the consumer family movement to increase their
use. And I just wanted to remind you all of
those developments where the dollar empowers
directly the consumer and family member, and
doesn't go through all of the kinds of
organizations that we're talking about. Thank
CO-CHAIR DUBOW: Operator, is there anybody on the line who has a question? Thank you for that observation, by the way. Operator?

OPERATOR: At this time, if you'd like to ask a question please press *1 on your telephone key pads.

CO-CHAIR DUBOW: Okay, then let's continue with the audience.

DR. GIOVANNETTI: Hi, my name is Erin Giovannetti from National Committee for Quality Assurance.

One thing that I have not heard in this discussion --and maybe this going to come up later-- is, when you're using patient-reported outcomes as performance measures, is accountability, and specifically talking about quality of life and function.

If we really think that the evidence is there that we can hold providers, entities, health plans, whatever accountable
for quality of life and function outcomes when
we know that there are a lot of outside
factors, social support, housing, income, et
cetera that are impacting these. And just kind
of thinking through are these -- is the SF-12
actually controllable by an individual
provider? Is that something that they can
change by their annual wellness visit?

So, I just kind of wanted to get
some -- hear maybe from the panel in terms of
what you think about function and quality of
life and its controllability.

CO-CHAIR DUBOW: That's an
interesting question, and I can't help but
observe that there is a functional status
measure that NCQA has which is in the HOS.

DR. FIHN: I'd like to respond. And
I'll just expose my own bias here. You know,
I think the -- and in David's paper, you know,
he actually addresses some of the areas in
which these have been used; one, in
particular, hip arthroplasties. In our case I
mentioned we're looking at PCI as a very
directed one.

I find these attractive not only
because in a sense we think about them as
measures of technically how good people are,
did they do a good job with the hip
replacement, or did they put a stent in well.
But I think in terms of measuring a health
system, I like them because they also would
depend upon how well your patient selection
is, are you selecting great people? In the
case of hips, are you doing the appropriate
post op rehab. In the case of coronary
revascularization, you know, are -- one of the
big issues are you actually revascularizing a
lesion that is causing the symptoms? If you're
doing a lot of revascularization for lesions
that aren't a problem, patient symptoms are
not apt to get better.

So, my plea would actually be
let's start with some of these focused areas
where we do have validated measures, where
they've probably are better, if you will, value proposition than the larger sort of much more generic, and I would agree, difficult to effect. There may be in intensive primary care, some groups of patients in which we can alter sort of the global health, but that's a hard thing to move and control. So, that would be my own personal bias, if we're going to get started, to start is some very focused areas where we've got some good evidence already.

MS. HUFF: I would also show my bias and agree with what Steven has said, as well, in terms of really being careful to select what population we're going to be using in terms of looking at these measures.

But to the point around using some of the general status, like the SF-12, I will say what we have found is really -- it does show sensitivity when you look at a chronic b-- a population that has chronic conditions.

So, if you're really careful in terms of
selecting the population for which it has
sensitivity, then it is useful. And we found
that not only in the project that I talked
about in my introductory remarks, but also in
some other research that we've done.

DR. MOSELEY: I just want to
underscore the differences between performance
measures for acute and long-term supports. As
I listen to the discussion of remediation of
conditions and various surgical interventions,
those are really good. And I think it's very
important to have that as a part of the
person's overall treatment plan. But people go
back to life, and they go back to jobs, and
they go back to family members, and they go
back to living in situations, particularly
those who receive live-long supports living in
situations that are paid for and controlled by
others.

One of the biggest factors,
variables, in individual outcomes that we've
documented through the National Core

Indicators Program is state. The state is the biggest variable. And why is that? It's because state Medicaid programs, structure and functioning of their Medicaid programs, although they're all based on the same statutory framework, they vary sometimes significantly from one area to another. The amount of supports that may be available to a family to enable them to keep their son or daughter with disabilities in the family home for a period of time may very vastly. And, actually, since 2007 when the economy has kind of gone off the deep end, the level of supports that are available to individuals has declined significantly in many areas.

These have a very powerful impact on the quality of life that a person has, their ability to control their own services and supports, and their ability to really continue to interact with their families and their communities.

MR. ROONEY: Hi, this is Ted Rooney
from Maine. We're Force for Quality Community, and I want -- this is one of those both-and comments. I absolutely agree we need to focus on those patient-reported outcomes that are controllable by the health care system. And, at the same time, many of you are familiar with the work Robert Wood Johnson has done with Project Match, which is looking into social determinants come out of the University of Wisconsin and others. And if we read that right, it suggests that 80 percent of the health of the population is not due to medical care. We're hoping that the promise of the accountable care organizations does begin to look beyond those things that are directly controllable by the clinician and look at the community.

And at least in Maine, we're actively talking about it. Granted we're a smaller state, we're not going to have 14 ACOs in one environment, but we have account -- you now, ACO-type organizations that are actively
talking about the social determinants in the community because they recognize the limits of health care.

So, yes, I think we need to focus on those things that are controlled by the health care system, but at the same time I'm hoping we absolutely look at the communities, then, to determine 80 percent of the health population, what can we do to support those things in the community that are way outside the control of the physician, but need to be impacted.

DR. PERFETTO: I think I pressed the button too many times. Eleanor Perfetto, I'm with Pfizer.

I want to go back to something that I think it was Ethan brought up a little bit earlier. In the pharmaceutical industry, our most vast experience using PRO data, and I know David knows this well, is in the clinical trial process, and incorporating these tools in clinical trials. And it's a
very difficult thing to do to be able to get something into a clinical trial and be able to differentiate whether or not you're finding some differences because of a treatment that's been provided. So, we've got some experience in looking at whether or not a PRO can change in a given clinical trial environment because of treatment versus placebo, or several treatments against one another.

And I think it gets back to something Ethan brought up before, was this idea of purpose of fit. And we have -- our arbiter in the pharmaceutical industry about whether or not the tool can do what we would like for it to do in this differentiation process is the FDA. And the FDA looks at the data that we provide on the tool that we want to use and gives us the yea or nay about whether or not they think it's rigorous enough to be able to do what we want it to do.

And I guess one of the challenges that I see in this process is: If a tool is
going to be selected, or if a set of tools are
going to be selected, to be able to be used as
performance measures, who is going to be the
arbiter to say that that tool is good enough
to be able to do what we want it to do, and
that it's capable of doing those things; and
then from there, be able to have that tool be
translated into a quality performance
measurement process?

DR. KAZIS: So, think this has been
a very exceptional discussion. I'm Lewis
Kazis. I'm at Boston University. Our group, in
fact, developed the Veterans Rand 12-item
Health Survey which has now been adopted by
the Veterans Administration, as Steve
indicated, and also by CMS as part -- as the
principal endpoint in the health outcome
survey to evaluate the Medicare Advantage
Program.

My view, I think, is that one
needs to be as flexible as possible in terms
of the implementation of these assessment
tools. And that if one moves to very focused
disease-specific assessments, where one might
see an effect to the exclusion of a generic
measure, I think one might lose out and not
get all of the information, in fact, where a
lot of the information that might be conveyed
in terms of the kind of clinical care that's
being rendered.

So, I think it becomes important
to consider not only the disease-specific
assessments, but also the generic assessments
in terms of evaluating, and maybe to consider
in the larger health care systems whether one
can focus on the low-lying fruit to begin with
in terms of those particular populations where
one might get the biggest bang for the buck.

CO-CHAIR DUBOW: Okay. Thank you
very much. I think Erin really highlighted a
really important area, and that is that we are
talking about quality improvement, as well as
accountability, and we need to take that
challenge into account as we go forward.
I hope you'll join me in thanking the panel for an excellent job for getting us started.

(Applause.)

CO-CHAIR DUBOW: I think we have a break for 15 minutes. Is that right? And we'll be back here at 11:00 for the next panel.

(Whereupon, the above-entitled matter went off the record at 10:47 a.m. and resumed at 11:02 a.m.)

MS. PACE: We are going to reconvene. Greg Pawlson, Elizabeth Mort and Gene Nelson, come to the table.

DR. PAWLSON: Okay, we are going to get started now. I'm very sure that there is important stuff being talked about. There are probably about five new grants that are being discussed. All we need are a few more funders in the room, and then we'd never get back to going.

It's always, I think, wonderful in a gathering like this, the conversations that
are sort of offline are just as exciting and
interesting, and sometimes even more creative,
in some senses, than what actually gets
presented. So these meetings do have, I
think, a very important focus, and especially
this one.

This is an area that I suspect
almost everybody in this room has been
thinking about, kicking around sort of in the
background for a very long time. And in
different pieces of it, whether you're talking
about functional status or patient-reported
experience of care, or other aspects of POM,
it's been around for quite a while. But I
don't think it's been put together in this
coherent fashion.

And I do think this is an
incredibly interesting and opportunistic time,
because I think one of the things we have now,
that we didn't have even five years ago, is --
first of all, we have the developing
electronic capability, which I think is
incredibly transformative.

And I wasn't kidding. My one and a half year-old granddaughter was actually going online and finding stuff that she liked using icons, which is very different. It's a new language, in many ways.

So we have that. And then, on the other side, I think, we have this emerging concept of the patient-centered medical home and accountable care organization, which, if it's done right, can truly be a point of accountability and take into account, and factor in, for example, the use and development of community resources as part of their overall mission. So I think we have both a receptor and an effector on site that we didn't have before.

So having gone through acknowledging, in the first panel, the patient as an authoritative data source, what we're going to focus on here is that it can be done. I don't think anyone would say it's being done
perfectly, or as well as we would like to have it done, but it is being done.

And we're going to have two examples of that, which are sort of domestic, U.S. examples. But I would also point out in the paper that it was a very nice little vignette about what's been going on in Sweden, and I am told in our next meeting we are going to actually have representatives from an even larger, very extensive use of patient-reported outcomes that the U.K. has embarked on, and which they've now had about two or three years of experience with. And this was implemented across the entire National Health System, so equivalent to sort of some of the things that Steve talked about that the VA's trying to do.

I think what we're going to try to convey is, this can be done. There are still lots of issues. There are still methodological barriers that keep popping up. But we are making real progress. And here to share their experiences with us are Elizabeth
Mort from Massachusetts General Hospital and Gene Nelson from Dartmouth Hitchcock. And we're going to start with Elizabeth's presentation, since they're a little bit more in the formative stage. Gene's been at this a while, and can show us his scars a little more.

But I think it's a nice balance, because we're sort of looking at one that's getting up and started, and has overcome a lot of the inertia and initial issues, and another that's been in operation for a while. So, Elizabeth?

DR. MORT: Thank you very much, Greg, for that introduction. And thank you very much, Helen and others at the NQF for inviting me to come down and share this exciting story, Lessons From the Field: Early Experience with PROs at Partners Health Care.

We are just beginning, but we're very enthusiastic, and we're in this to stay, I hope. I think how this session today and
tomorrow, and the next part of this workgroup, goes will determine if we're in it for the long run. But at least, we are very, very excited to start.

We are only in the beginning of our data collection phase. We only started data collection in March, and I'll get to that in a minute. I wanted to spend a little bit more time up front telling you about the time and energy that we intentionally spent in setting this up, hopefully for success in the organization.

Partners Health Care is an integrated delivery network in Boston. It was founded in 1995. The founding hospitals of the Brigham and Women's and Mass General kicked it off at that time. We've had several CEOs, and the most recent CEO set off a new strategic plan that was launched in 2010. When we launched that, we were looking broadly at care redesign in two areas. We were looking at primary care and population health.
And we were also looking at condition-specific care redesign, and we selected CABG, stroke, colectomy for colon cancer, AMI, and diabetes as our focused conditions.

When we set this up, we organized this around a key principle, which is that if we were going to redesign care, we were going to redesign care with the goal of improving value. And we talked about this concept for a long time, because our providers, when we asked them to come to the table to work on these projects in 2010, knew we were about to undergo a large change in the way we were paid.

They asked us "Well, how are we going to be paid? How shall we do care redesign?" And of course, in 2010, all we knew is that it's likely to be different, but we don't exactly know how. It's probably going to be something in the order beyond unfettered fee for service. So what we'd like you to do is think about organizing care with
the goal of improving value.

So these overlapping Venn diagrams, we must have shown thousands of times, pointing out that, of course, our goal here is to improve care while keeping it more affordable. So we want to improve outcomes by reducing costs, and obviously the inner section is the value.

So, when we brought people together, we'd had a decade or so of teams working on quality improvement, working on measurements, working on process indicators, working on outcomes indicators.

What really captured people's imagination, and what really has stimulated this work from the get-go, though, was inviting our care teams to think about outcomes that really matter to patients. We want to start collecting patient-reported outcomes. That's what patients, after all, really care about, all the kinds of conversations that we've had this morning.
So I tell you this because we got everybody really fired up about this quest, sort of two years before we even got into the implementation tasks. So we selected five conditions, four of which are acute, based upon an episode starting in a hospitalization, one of which is chronic, diabetes, looking at chronic care over the course of a 365 day period. But we were organizing this work around episodic care population management, and we decided to start with CABG, and we actually added AVR, aortic valve replacement, as well -- coronary artery bypass graft, I should say, and aortic valve replacement -- for the purposes of getting adequate volume to study these PROs. And then we selected diabetes.

We spent about two years. We had a very engaged physician from Israel, Eyal Zimlichman. He may have interviewed some of you in the room. He really led this project, and did all the sneaker work, sneaker power,
going around talking to people, and really
doing a very, very thorough job of change
management, managing us, by interviewing
people. Interviewing researchers,
interviewing folks who went up to Dartmouth,
learned from Gene and his colleagues, to learn
about how this works.

So it was never one of those
interventions where we said "Okay, we're going
to do this. Here are the measures. It's
going to start next Friday." It was all about
getting people engaged, getting iterative
conversations with high-level people, people
very invested in this work from the get-go.

And the kind of goals that we
thought about from the beginning, I put up
here on the slide. We said "You know, this is
likely -- when NQF gets its arms around this,
this is likely going to become the way we do
business in the future. At least, we all hope
that it does. So we want to organize this so
that we're positioned to be ready to catch the
wave when it comes."

So we wanted to make it electronic. We said "Well, we can start it with paper forms." No, no, no. We want to make it electronic. Let's do it with futuristic goals in mind. And then we didn't want to spend time developing new instruments. We thought possibly taking measures, or pieces of instruments, and putting them together would be okay, but we didn't want to start from scratch, so we decided we would use validated instruments.

We wanted to reduce respondent burden, so we decided we would make the instruments short. We may have gone a little bit too short. And we wanted to align this with our overall care redesign strategy, and also some of our paper performance strategy. So again, we spent a lot of time setting the table for this important work.

So a summary of our tool for CABG and AVR is on the left and diabetes is on the
right. The total number of questions for CABG pre-procedure is 17. Ten questions on functional status using the PROMIS-10, some symptom-level questions from the medical outcomes survey, receive health benefits again -- that's a post-op question, obviously, in retrospect -- and health utility from the EuroQol.

So we had 17 measures pre-op, then 21 measures post-op. We engaged our cardiac surgery clinical team. Hours and hours of meetings, and tweaking, and discussions and vetting. It's kind of the way we like to do business there. It takes a long time to get things done.

But on the diabetes side, we had a very robust diabetes team. We decided to use the same functional status measures. We decided to use PROMIS-10. Actually, David Cella was very instrumental. He came to Partners and gave us some lectures, and again we were very intentional and deliberate in
making sure we had run the bases on this.

We added an anxiety measure from PROMIS, a burden of diabetes measure from one of the American Diabetes quality of life indicators. I'm blanking on the exact name of the tool, I apologize. But this was a really, really important measure.

We held focus groups with patients who weren't familiar with this kind of measure. We explained to them what we were trying to do. And one of the patients said to me "You know what I want to hold you accountable for, Dr. Mort? I want to hold you accountable for keeping me as normal as possible. Just making things normal, so that I don't have to think and worry about managing my diabetes, or the symptoms or complications associated with it."

I thought that was very, very instrumental to me, to think about how we want to organize this work. But we found a measure that measured that pretty well, and then the
health utility measure as well.

So the way we wanted to roll this out is, we have this working in a CABG and two cardiac surgery clinics, doing a pre-op assessment and then a post-op at 3, 6, and 12 months. Our plan for diabetes is a baseline measurement and then every 6 months. Both of these instruments, both of these data collections, start in the office.

Data collection. Again, we wanted to kind of channel the future here, so we didn't want to spend a lot of time with our IT folks developing internal data collection systems. So we actually talked to a lot of vendors, and we're partnering with a vendor to do the data collection.

And on the slide you can see, we start with tablets. And all the things that have been said about one year-olds and iPads, and so on and so forth, are very much applicable to the patients that we have tried this with in these waiting rooms. The tablets
work very, very well.

When the patient is given the tablet by the medical assistant or the secretary -- they are given a list of patients who are coming in. They hand the tablet. The patient goes and sits. Workflow-wise, that works pretty well, because they're sitting, and in most clinics you have at least a few minutes to do something while you're waiting to get checked in, and we only have 17 measures. So it has not been a burden or a workflow issue, once you get the group engaged.

The patient is then asked how they want to have their follow-up done, and they can choose between using our patient portal that we call Patient Gateway, again anticipating that we want to move in an electronically forward-thinking way, or IVR with phone operators.

So we really are early in this.

We've only been in the field since March of
2012. We've only collected data on 264 questionnaires. 56 percent of the patients who we've enrolled have chosen a method of follow-up selecting the patient portal, the internet option, so not the IVR. So I do think that's going to be an increasingly popular way to collect information in an asynchronous way.

Our IT folks are working on developing reports, both for patients and providers. These are still being developed and piloted and iterated, but the idea is that these reports would be pushed out to the patients through our patient portal. We're getting good traction with that tool, and increasingly getting more and more of our patients across the entire system enrolled in our patient portal.

And we have the electronic medical record. We are undergoing a massive change. We are actually installing an entirely new clinical and business system across our entire
network. But in the interim, we do have an
LMR, and the goal is for these indicators to
be tracked right along with -- this is hard to
read, but it's a vital sign. It's a flowsheet
for vital signs. And we have these kinds of
things for clinical indicators, like
hemoglobin A1c, blood pressure and the like,
and we'd like to do the same with the
functional status/quality of life measures.

So the feedback -- again, this is
eyearly. The patients -- we have spent a lot of
time with research assistants at the
practices, working with the front office,
working with the staff, working with the
medical assistants, working with the doctors,
in large part to make sure it happens, but
also to learn from them and to improve things,
and to make some iterations as we go, early
on.

Patients say their doctor should
be asking these questions. They like it. The
tablet's fun. And they say they're willing to
answer these questions at home.

The staff experience, the practice administrators, once we make the case for this and they understand it's important, they've been quite flexible in helping us to get this embedded in the workflow. The medical assistants and nurses are very, very eager to get involved in this kind of information, and are great adjuncts to the nurse practitioners and physicians who are actually seeing the patients.

The physician experience is a little mixed, not surprisingly. You know, we've spent 15 years at Partners educating our clinical colleagues about measurements, and we don't have all the answers to "Well, is it valid?" and "What's the tool?" and "What about the scale?" and "How can you know if there's a difference?"

We don't have the answers to all those things that we've been telling people for 15 are years are so important, and I said
"But wait. People are working on this. The NQF has a workshop, and over the next couple of years these things will evolve. And in the meantime, let's get ahead of the curve." That usually gets you somewhere.

(Laughter.)

DR. MORT: It's honest. It's honest. But everyone loves the face validity of these things. The workflow is really an issue, though, because doctors obviously -- this has been alluded to this morning already -- doctors aren't used to getting this information. How does it fit in?

You know, we have our script. We ask the patients a good, open-ended question. "How are you?" Your annual exam, "What are your concerns?" But we have to figure out how to get that piece of data involved in that conversation, so that we can embed it in the workflow of seeing the patients, as opposed to saying at the end "Oh my God, here is this quality of life sheet, let's talk," and it
doesn't work. So people are worried about those things, but people honestly are working hard at trying to make this work for us.

Concerned about "What do I do with the results?" So if I get a critical result, like a potassium of 5.6, doctors know what to do with that. "But what if I get an indicator from one of these scales that suggests the patient's in trouble? Give me the tools to do something with that information. Tell me what to do in terms of referrals, but also make it easier for me to know that someone else is watching for those critical events and flagging me, just like you do for the potassium."

So, just some lessons learned. Most of these I have already alluded to, but we thought it was very important to spend a couple of years doing the change management, doing the research, educating ourselves and bringing experts to the system, integrating it with our data collection on our strategic
We have incredible support from our senior executives, my colleagues at Partners as well as the hospitals, all the way up to the CEO of the entire system, who really believes in this, understands that we have to be cautious about going forward, we don't have all the answers, but it seems to be tremendously promising.

I think I've covered most of this. So I am hoping that over the next couple of days and the next few months, and subsequent couple of years, we'll have the answers to some of the questions that are being raised. But I do hope and believe that this is work that is here to stay.

So again, thank you very much for asking us to come and share our preliminary findings. Hopefully down the road we'll have some more substantive results.

DR. PAWLSON: Thank you very much.

Just as Gene is coming up to start his
presentation, I was really struck with how careful a process you've gone through in terms of change management.

And also sort of something to think about, perhaps for a later question, and that is one of the real hallmarks of Partners has been, for a number of years, it has been doing incentive-based contracting, so that a substantial, or at least a significant proportion of reimbursement has been wrapped around achieving some level of performance in different areas. And I think that kind of integration of payment with professionalism and wanting to do right for the patient is a very, very powerful sort of mover and shaker in this area. So we'll perhaps take up on that.

Any questions for clarification, something that just you didn't understand? I think it was a very clear presentation. Thank you.

DR. BASCH: Just a quick question
of clarification. In the development of the
selection of the measures, was there patient
input, or was it mostly the expert teams of
clinicians that were consulted?

DR. MORT: Ethan, we had focus
groups up front to inform the domains that
patients were interested in. I believe,
though, in all honesty, once we identified the
specific measures -- no, I stand corrected
here. I'm arguing with myself.

We did go back to our focus
groups. Because we had groups of patients who
were advising this care redesign process, and
they were a group that was interested in
parsimony, and they also felt one of the
concerns was "Don't make the questions have
lots of different ways to answer it." You
know, they wanted the response patterns to be
similar.

DR. PAWLSON: Another important
lesson.

DR. NELSON: I think you're going
to cue up some slides, and thank you for
inviting me. It's great to be with all of
you.

I've been asked to speak about the
Spine Center and its experience, and we're
going to start with a riddle, and it comes
from Amory Lovins. How is a kilowatt-hour of
electricity like a day in the hospital?
Nobody wants either. We want cold beer and
hot showers, better outcomes, better care,
lower costs, and use least costs. Value for
money.

So what Amory is saying is, he's
an energy expert. But when we're thinking
about value, it really does focus on the end
user. So that means our patients, our
clients, the families that the patients reside
in. And we have a sense that this is where
the great one, Gretsky, talks about the secret
of skating to where the puck is going to be,
and we think that focusing on person-centered
value, and incorporating patient-reported
outcomes in that, will be really essential.

This is going to be a brief excerpt from about a 50 page technical paper that's available to you. It's available on the internet or in hard copy, and it actually has three case studies: one from Karolinska and rheumatoid arthritis patients, one from Group Health in the Pacific Northwest and primary care patients. And the third case study is the Dartmouth Spine Center, and that's the one I'm going to focus on now. But there's a lot more that you can glean from some systems that have been using patient-reported outcomes for about a decade.

So this idea of value, that we start with an individual living at home or in the community, and then they interact with the health care system -- processes of entry and assessment, and a care plan, what's going to help me become better, and then follow-up over time to see what the outcomes are, what's the new functional status, the new risk status,
the new disease status, if the person has a disease or a condition, and what cost. And what's my experience on the ride through the health care system? How has that treated me? So an image of value that's very person-centered.

And we, like Mass General, have really been focusing on the redesign of care that becomes person-centered and that tries to deliver on value. So you'll see this use of patient-reported outcomes embedded in an effort to redesign care for spine patients. And more than 10 years ago, Dartmouth had spine patients running all over the place. They could have been seen in internal medicine, or the pain clinic, or orthopedics. It was a mess, like much care.

And so the idea was to redesign the care program so that it's one-stop shopping, and that it's very person-centered: back to work, back to play, one back at a time. This is Jim's initial idea. And to use
the patient-reported outcomes to create a new
information environment, and a better
relationship with the patient to achieve the
outcomes that they would wish to receive.

What you'll see in just a moment,
then, is a new information environment that
Dr. Weinstein would say he can't be a good
physician for his patients absent this kind of
information. "It's essential to understand
where the patient's coming from and how
they're doing to be a good clinician," in
Jim's words.

So in a schematic form, the
information environment was changed so that
when a referral is made or the patient
requests a visit to the spine center, that
they are actually requested to complete
information as they're oriented to the spine
center, what does it have to offer, and then
patient-reported information. And that moves
to the initial work-up and plan of care.

And that information can be
completed at home, over a portal, or when the person shows up with a touchpad. And that touchpad or that portal information is uploaded to the electronic health record. And then that is the grist for trying to create a plan of care that meets that person's actual needs, in a way that you'll see amplified in just a bit. And then, depending on their need, they'll go into an acute care program or a chronic care program, or a functional restoration program, or, some people, end-of-life.

And then that data on the patients' outcomes is being fed forward with that patient over time as follow-up occurs, and it's fed back to create a registry. It's fed back for clinical program improvement. It's fed back to become part of a national trial.

So, feed forward/feedback of patient information. This is very dense, but this is what Jim was talking about, Dr.
Weinstein, that this is all based on patient-reported data, and it's meant to be same-page care, if you will, to put the clinician and the patient on the same page about "So, what are my risk factors? What's my history and my symptoms? What are any red flags that I might have? What's my functional status right now, and how is that changing over time?"

On the right hand side, you can see a trend line for physical function and mental health, based on the SF-36 in this case. You can also see pain portrayed in terms of the body, and the patient's experience of my own outcomes. So, "Can I sleep better? Am I able to get back to work?," et cetera. So this is used to create the next step care plan, and it's all based on patient-generated data.

So the patient-generated data can also be used to actually go from the concept of value -- easy to say, perhaps hard to measure -- to measuring value. And this is a
bit complicated, but the same feed forward/feedback patient-reported outcomes system that was started at the Spine Center became the data collection device for a randomized controlled trial that NIH funded, also an observational one as well, and in 12 other centers, including Dartmouth.

And then what it became was a comparative effectiveness research study, to see how people in blue, who had surgery, versus people in yellow, who were treated non-surgically, did at 6 months, 12 months, 2 years, 4 years. And the patients are still being followed.

And this is two-year results, and it's one of three patients populations. It's people with herniated disc. So the average person with herniated disc is portrayed here, on the east side. The west side is a disease-specific measure called the Oswestry Index, and higher scores here mean greater improvement in disability. So blue, -38 on
the Oswestry, versus -24 in yellow, non-
surgically treated patients at two years. So,
favoring surgery.

North is the SF-36 physical component score, and both groups had huge gains, 44 points and 34 points respectively on a 0 to 100 scale. So these are giant gains, surgery a little bit more gain after two years.

My perceived health benefit. "How much was I helped by the treatment that I got?" is on the right hand side, so perceived health benefit, both strong but once again favoring surgery.

But you see at the bottom these better average results did cost more, so this is an estimate of total direct and indirect costs incurred by the patient or on behalf of the patient, so about 25,000 dollars versus 10,000 dollars direct and indirect costs after two years. And in the very middle is the incremental cost per quality-adjusted life
year, and that means about 74,000 dollars more per quality-adjusted life year for surgical care over non-surgical care, which many would consider in the United States a reasonable expenditure.

So these results are then providing good information for research under these conditions. They're also used, on the lower right hand side -- if you go to our website, you'll see different kinds of outcomes and experiences publicly reported for over five years. And so this is transparent, these kinds of results, for people with herniated disc, for degenerative spine, and for stenosis, publicly reported.

This is a prototype, now, and after 10 years and a collection of a lot of data I can show up at the Spine Center and I could see not just results for the average patient, like I just showed -- after two years, what might my results look like if I got surgery or not -- but this is a risk
calculator.

So I would enter in my age and my gender and answer, in this case, four questions about pain on this screen, and I would then get a personalized display on the right hand side of what my estimated results would be for people like me with respect to pain relief. And that's, after two years, the moderate versus the mild levels of pain. Yellow is non-surgery, blue is surgery. So, likelihood of better results.

And then the lower right hand boxes have the face plots, and it shows the proportion of people like me that would be likely to benefit, or not, for personalized risk assessment or benefit assessment, leading to the possibility of very good shared decision making, very good informed decision making, very good patient engagement with better data about what my choices might look like and what they might get.

So wrapping up, we've gone into a
new electronic health record, and so a lot of the functionality that took nine years to build was lost in about a nanosecond on April 2nd. And so we've got this incredible group of people that are recovering the lost functionality, and putting it into the new EHR environment.

There are now 18 different clinical programs that are using the patient-reported outcomes data at Dartmouth Hitchcock, the oldest being the Spine Center, but many others that have been used for more than five years. And we think that it has real benefits for patient care and for research, and for where the health system is going.

We're shifting over to value-based contracts. We're a pioneer ACO. Our basic strategy is better value, better outcomes, better experience, lower costs. Redesign of care programs for people over time to accomplish that. And so we think that this information environment is really essential to
make that happen. The redesign of care is essential, and the patient-reported outcomes, as part of the information environment, is critical to that.

So, lessons learned and a few recommendations. A small comparative study was done at the Spine Center and the Rheumatoid Arthritis Registry in Sweden, and the results are published in a paper that was published a couple years back. And in this small series of patients, as we thought, in Mass General Hospital, patients tended to be positive about giving their information. In this case, 84 percent were positive about the use of the patient-reported outcomes. A statement visit became very helpful, thorough, and informative.

Providers' reactions are mixed, and in general when the provider is actively using the information, it allows the patient to become more involved in their care. "Patients get more involved in their care" is
a quote. It changes how health care is
delivered, and there can be a real shift in
the relationship when you're using the same-
page care approach. We're now together
looking in on my health outcomes and what we
might do best next. And so it can change how
care is delivered, and that can usually often
be appreciated by clinicians, and sometimes
not, because it is different.

Sustainable and replicable. We've
been going for a decade at Dartmouth in a lot
of different clinical programs. Some
recommendations are on the right, and this is
to make these kinds of systems work in busy
places.

Here's five suggestions. Fit the
PROs into the workflow, to make it easier for
patients and providers to do the right thing.
Co-design the system with stakeholder input
for best end user utility. It's got to be
useful for the patients and their families.
It has to be useful for the clinicians and the
clinical teams. You can't just throw it over the transom.

Educate the patients and the providers on how to use the PROs. And the providers have to pay attention to the data, because if I've taken the time to report and it's ignored, as a patient, you're disrespected, and what was the purpose of this? And so the clinicians using the information is critical. Capture data from other sources to improve the utility of the information and then make it better over time. So, thank you.

DR. PAWLSON: Thanks very much. It's interesting, again careful planning and dissemination, and also use in terms of payment enhancement, potentially, again, in the clinical care of the patient and in rapid learning feedback research. And I think having all three of those things as power in this is to me, at least, more than sufficient reason to be doing this. So hopefully these
two, I think, very well-honed presentations
have raised a number of issues and questions.
And we'll start with the expert panel.

Yes?

DR. BASCH: Thanks for those great presentations. Ethan Basch again, Sloan Kettering. Something interesting that's alluded to in these presentations is that the collection of patient-reported outcomes in practice itself, that very act, can be considered as a quality measure or a performance measure, right?

So it's a different way to think about it. One way we could look at your examples is to say "Okay, you've demonstrated the feasibility of measuring various outcomes," but another way is to think of the integration of PROs into clinical care as a structural process measure, which is another interesting way to think about it, if we believe that integrating these into practice does enhance the delivery of care and the
ability of practitioners to self-understand
and benchmark themselves against other
practitioners, and thus continuously improve
their performance.

DR. PAWLSON: Sort of just like
ordering a statin, or a lipoprotein, or
something.

Next?

DR. KAZIS: I'll keep the mic a
little further away this time.

I thought the presentations were
great. Having worked with clinicians for many
years who were on the front lines and are
dealing with the complexities of care, the
demands that are on them, the issue is
information overload.

And I've talked to a number of
clinicians given the electronic medical record
that are really frustrated, and find that the
information is often redundant. It's dated.
It doesn't convey what they really need. They
really don't have easy access to different
parts of the electronic record. It becomes a real challenge.

With new information that we're talking about in terms of patient-reported outcomes, how can we compel the doctors to better understand the importance of this information, so that in fact they're going to use it, rather than tossing it in the wastebasket?

DR. MORT: I think that's a really great question, one that I hope will be remedied by the change in the way we deliver care, and will be delivering it more so under ACOs and payment systems that are captitated, global payment, that sort of thing.

Patients come to me. One patient came to me last Wednesday. She said "Dr. Mort, how come we still have 20 minute annual exams? Aren't you a pioneer ACO, and aren't you doing all these --" because she reads, she knows how we're doing at Blue Cross and everything.
I said "Well, not yet. We're working towards that." But she's absolutely right. She's a health care consultant. She's absolutely right that we need to change the way we actually deliver care to make more access to group visits, patient portal, using non-physician providers, non face-to-face visits.

So we can't do it in the current 20 minute, or 15 minute, or even half an hour, an hour, face-to-face visit. That's just not adequate to deliver all the care, and absorb and react to all the data. So we're a little bit out ahead of it, I think, but I think the answers will be forthcoming as care redesign ensues and more and more practices figure out how to do it.

DR. PAWLSON: And I noticed you were using graphs and stuff, and that's another way. The whole way we display data -- you know, I was also on the plane, and I was looking at the difference between an old 757
and a brand new 737-800, and the display panels are just totally different. And I would guess that pilots get a heck of a lot more information the right way in the new cockpit. So that's a nice, I think, thing to think about, is how we deliver information.

I think we want one more question from the audience, and then can you unmute the phones to see if we have any phone questions?

So first, I saw somebody back -- one more expert, and then somebody in the audience. I saw a hand.

DR. GAGE: Barbara Gage, Brookings. I found the presentations very interesting. I did have a question for Gene, and one of the outcomes was a very important outcome. It was a two-year out outcome. How did you collect that from the patient, and in your comments about sustainability, is that something that you have on an ongoing basis? And if so, how are you funding it, or doing it?
Those particular results were part of this larger NIH-sponsored trial, so people were followed up in that case as in a research study. We had a meeting last week with the ortho group, and the issue was "Let's make sure that we attain 90 percent baseline PRO data, and 80 percent follow-up PRO data for at least two years." That we have to do this, it's important for our ability to, again, provide care and measure the results.

And so that becomes the design challenge, to get the work processes and the patients and clinical teams engaged enough, and the design good enough, that we get 90 percent intake and 80 percent follow-up over two years.

DR. PAWLSON: Isn't that close to the British experience?

DR. NELSON: And to get at that, it's a mixed-methods approach whereby portal at home, possibly IVR has been mentioned,
iPads in the office for people that can't report that information, advance scheduling those people 30 minutes early to complete the essential information, so that everything works well.

DR. PAWLSON: And two years may be a bit of a stretch, but the British actually base payment on getting responses from patients in the three to six months post-surgical, so there are some levers out there.

I think we had one question back in the audience? Go ahead.

MS. MASTANDUNO: Melanie Mastanduno from the Dartmouth Institute. Just one point to add to Gene Nelson. Those orthopedic providers were very keen on having some of their clinical team staff participate in the review of the responses, so that screening for positive results that land on the doc's desk are the ones that are part of the workflow, as opposed to noting the results and integrating them into the record.
So screen for positives, and that will reduce burden on the physician.

DR. PAWLSON: Thank you all very much, and thank you to the panel. That was very well done.

MR. CUNNINGHAM: Real quick, just want to check with the operator. Do we have anyone else on the queue for questions or comments?

(No response.)

MR. CUNNINGHAM: Operator?

OPERATOR: Once again, if you would like to ask a question, please press star-one on your telephone keypad.

There are no audio questions at this time.

MR. CUNNINGHAM: We do have one quick question from the back.

MS. LENTZ: Thank you. Lisa Lentz, CMS. I did have two questions, one for Elizabeth and one for Gene. For Elizabeth, I wondered if you could just elaborate a bit
more on the process for involving patients in the selection of the domains and the measures. And for Gene, I wondered if you could elaborate more on how you've translated economic data, such as QUAL-Es, into something easily understandable to consumers on the website.

DR. MORT: Involving the consumers, patients, was really interesting, because the first couple of focus groups we held, we even had difficulty as focus group facilitators -- it wasn't me, it was people who were trained focus group facilitators -- describing what we were trying to get at, in terms of quality of life and outcomes. And what was informative was that patients weren't even thinking about it that way, because no one had ever broached it with them and asked for their thoughts.

But once they got it, we just had -- you know, we did it through focus groups and trained facilitators, to get people to
vocalize on what they thought was important.
And that's pretty much the methodology: focus
groups a couple of times for each one of these
projects.

Your question raises a question
for me, as to how are we going to do that on
an ongoing basis. So thanks for asking the
question.

MS. PACE: Okay. Could we have
our next panel come on up? We'll have to,
maybe during the lunch break, get a chance to
ask more questions.

DR. NELSON: Well, I'd like to
just answer this other question.

MS. PACE: Okay. Go ahead.

DR. NELSON: The question about
QUAL-Es on the website. We did not put the
QUAL-E, just as the cost data, on the website,
at this point. It's pretty complicated. But
what is on the website is an estimator of, if
you have a procedure or not, what your
expected out-of-pocket and insurance expenses
are going to be. So there is that kind of proactive cost information. Satisfaction is there, yes.

DR. BASCH: Hi again, I'm Ethan Basch, Memorial Sloan Kettering Cancer Center. I'm an oncologist and an outcomes researcher. I run a research program focused on informatics and patient-reported outcomes. I am delighted to be here. Thank you very much to the organizers for this invitation, and many of us here are quite excited to see this topic being discussed with such methodological rigor.

I also stand between you all and lunch. I often find myself in this position at meetings, probably because I'm a New Yorker and I talk fast. So you know, I can just speed us along. We have a little over an hour, and so I'm just going to set up our session here very briefly, and then hand it over to our panelists.

So, understanding the patient
perspective, or their experience with care, involves more than just developing a questionnaire. It's really a whole package, right? It's the questionnaire, but also the way that it's administered, the way that it's interpreted, and then how it's acted upon.

Our panel now is going to focus on the second piece of this: how the questionnaire is actually administered once it's been developed. And this is vitally important for a couple of reasons. The first is that how instruments or questionnaires are administered can actually affect the information that you get back, right? How you ask the question affects the answer that you get, and we really need to be very careful that we don't alter the meaning of what we're getting back.

But the second reason, which I think is actually the most important, is around missing data, missingness. There is a real risk in real-world populations of having
systematic missing data that's not at random
from particular populations: populations at
risk, populations who are traditionally
underrepresented or hard to reach, those who
are the sickest. Oftentimes, the patients
whose perspectives we may care about
particularly. And if we're not careful in the
way that we administer our questionnaires, we
can exclude those patients and lose their
perspective.

There's another issue, which is
that practices that are particularly good at
eliciting responses from their sickest
patients may actually look worse than those
practices that actually don't get as many
responses back from their sick patients. So
there are all sorts of biases that can be
introduced by the way the questionnaires are
administered.

So our panel is going to focus,
really, on three broad conceptual areas around
administration of questionnaires. The first
are methodological issues. Methodological issues are things like the mode of administration. We've heard a little bit about IVRS automated telephone administration, there's web administration, good old fashioned pencil and paper.

There are true scientific or methodological issues that are related to mixing these up within a population, to developing a questionnaire in one mode and then converting it, or looking at equivalence in another mode, and so on.

There's an area of increasing interest called CAT, or Computerized Adaptive Testing. David Cella's about to give us a demonstration of this in action using an electronic questionnaire, so I'll leave it to that, and it will also be discussed by one of our panelists.

And then the second issue is feasibility. So particularly with a very large implementation, in a large population,
substantial infrastructure has to be developed. Personnel need to be trained and put in place. This can be cumbersome. It can be complicated. And it needs to be sustainable. And there are real barriers and lessons to be learned from other contexts, so we'll be highlighting those.

And then the third issue is around community or population engagement, patient or person engagement. And this really has to do with enlisting populations as our partners. As we've heard already, there are focus group approaches, but there are also community outreach approaches. With some of these very large engagements, there's a real need to engage people or patients as our active partners. So, methodology, feasibility, and patient engagement.

We have three learned panelists today. I won't belabor their introductions. Their full bios are in the distributed materials. We have Lewis Kazis from Boston
University. He's a professor of health policy and management who directs the Center for Assessment of Pharmaceutical Practices. We have Richard Bankowitz from Premier Health Care, who is the chief medical officer, an internist and an informaticist. And then from the Patient-Centered Outcomes Research Institute, Lori Frank, who's the director of patient engagement research.

I'll stop there, and once again introduce Dave Cella, my old friend, who's going to give us a demonstration of CAT in action for about 20 minutes.

DR. CELLA: It's me again. Thank you, Ethan. This is going to be the densest, most technical part of the day, so bear with me.

Before I get to CAT, I thought I would come back to the paper and review some of the key points made in the paper for those of you who might be seeing or hearing some of these things for the first time. Reminding
you this is a session for methodological
issues, so forgive me in advance for getting
into what might seem like some technical
details.

So I'm going to talk about method
and mode of administration, and also the
source of data. And it's important to get
into this to some degree, because decisions
have to be made about the method by which you
get these data, and there are costs and errors
associated with surveys, however you go about
doing them. It's important to select the most
appropriate method for a particular question,
and to try to stick within that method when
possible. We can come back to that.

And most of all, know the impact
of a particular methodology that you're using
on errors and costs. Methods and modes differ
along various dimensions, which the paper
covers; that is, the degree of interviewer
involvement from none to complete, sometimes
something in between, the level of interaction
with the respondent, or the person providing
the information, the channel of communication
that gets used, and also the degree of
technology.

One way to look at that is to
start from the source, so the source is either
going to be the person himself or herself, or
a proxy or observer on behalf of that person,
usually selected as a second choice, but
sometimes an essential one. And then mode, in
this context -- in the paper, at least, we
referred to mode as the recorder of the
information.

So if the person is providing her
information directly on a piece of paper or on
a computer screen, that's self-administration.
If the information is collected by talking to
another person, that's interview-administered.
We'll use that distinction for mode for this
context, at least. And the method, then,
would relate to whether you get that on paper,
whether you get it on a computer, or over the
telephone.

So, proxy reporting. What are the pros and cons? It's useful, particularly when it's difficult, or not even possible, to obtain PROs directly from the patient. It allows, therefore, broader inclusion and a more representative range of patients. It minimizes the missing data problem that Ethan alluded to, increases the feasibility of longitudinal assessment, because you may be able to start with patient self-report, but then need to move to proxy, which is better than moving to nothing.

And so in that regard, proxies can substitute, or they can complement, patient assessment. There are situations where you might want both to be done concurrently. That may be a luxury in some settings, but it's something that a strong case could be made for that. You can involve proxies to assess patients as they think the patient would respond; that is, the proxy responding for
what they would believe, if the patient were asked the questions, what the answer would be, versus the proxy giving his or her own perspective on the patient, which can be different than what they think the patient would say. "I think my husband would say he's just fine cognitively, but I've noticed slippage" would be an example.

Evaluating agreement between patients and proxies is something that can and should be done. Usually in the literature there is better agreement -- it's never great, but there is better agreement when the rating is of something observable, like physical function or activities of daily living, being able to do things or function in the world, and less agreement when it's about something mental or in the social realm, such as pain or cognitive status and emotional status.

The magnitude of the disagreement can be minimized with careful attention, but keeping in mind that disagreement between them
sometimes is actually useful, as I alluded to earlier with the example of someone with early cognitive impairment.

So, how about mode, that is, the recorder? So, let's just consider the objectives of the assessment, and then the resources that you have available. So, there are advantages to self-administration. One is cost. You don't have to pay an interviewer. You often get better disclosure, or more disclosure of issues and problems on a non-interview self-report, and people can proceed at their own pace.

Disadvantages are that there's more potential for missing data, and it really does require that you have up-front careful attention to survey design, using best practices in survey design, because it's very easy to do bad surveys, and not so easy to do good ones.

Interview administration has advantages of allowing you to not worry so
much about survey design, because the interviewer can make up for the problems in the design, and it's useful for patients that have reading problems or writing problems or vision problems. The disadvantage is, of course, the cost and the potential for bias, because the interview is a social exchange, and in that social exchange that can influence the way people report how they're doing. And they tend to under-report in an interview, compared to self-administration.

Concerns about the effects of mode on data quality. So, the reliability is actually high for both. That's good news. Response effects tend to favor self-administration, but they're inconsistent. We weren't asked to make a recommendation, but if I were I would recommend, when possible, self-administration over interview.

Just to kind of run through this fairly quickly, the paper and pencil versus electronic, if you consider electronic to be
either by computer or by telephone -- and of course, there's a range of electronic administration options.

Paper and pencil has low start-up costs but more downstream cost issues related to data entry errors, scoring challenges, and getting it incorporated into electronic health records. So you may save money up front but lose it later.

Electronic has advantages of being interactive, very practical, more integrated, easy to incorporate into the electronic health record, but there's the up-front cost that's incurred by setting up electronic data capture.

There is a potential for differences between paper and pencil versus electronic based upon things like the impersonality of the method, the cognitive burden on the patient, who may find it easier to use the tablet -- I find it interesting that we use the word "tablet" in the medical
setting for small computers -- Control over
the questionnaire, which can be more easily
managed in some settings when it's paper, and
communication style.

The increasing evidence of
evidence equivalence is encouraging. That is,
there aren't a lot of differences between
different methods of administration. As new
methods are developed, it's critical to
compare them to existing methods.

Probably the most vulnerable is
telephone interview administration. First of
all, for example, getting an in-clinic
assessment and then having people call up on
the phone and get information later is the
most vulnerable to having a systematic bias,
where you look like you're improving because
you're having that social exchange. But apart
from that, if you're careful about the
longitudinal picture, you can minimize if not
reduce this bias. And across these methods,
patient privacy is always a concern.
So we looked at this, and PROMIS -
now, PROMIS, to those of you who don't know,
is the Patient-Reported Outcomes Measurement
Information System. It's a large, nearly 100
million dollar investment over an almost ten
year period from the NIH to develop generic
but responsive patient-reported outcome tools
in various domains that cut across multiple
chronic conditions.

And within that project -- I'll say more about it later with the CAT demo -- we looked at a mode of administration, or method of administration by this paper's terminology, study, comparing paper and pencil to computer to IVR, Interactive Voice Response, and to PDA, to handheld device or smartphone.

And we found, happily, that there were no meaningful differences found between modes of administration. Now, interestingly, remember what I said earlier about interview tending to boost up scores. The IVR was done
by a computer. And this is my belief; it's not something we can prove. But I believe that the reason the IVR was equivalent to paper and pencil and computer, which is great news for PROMIS, is because the respondent knows it's not a person they're talking to. They know that they're interacting with a machine, and therefore they're treating it, I believe, more like an impersonal exchange, and they're providing information. So it was not a live interviewer.

This less than one and a half points on a hundred point scale just illustrates that -- the vertical dotted lines represent what would be, even in the most conservative sense, an important difference of two points. That's two tenths of a standard deviation either side of the average. The estimates, whether you're comparing PC to paper and pencil or to IVR, or to PDA, for fatigue, physical function and depression were always virtually identical to the PC
administration.

And I think this came up earlier, as well, and somebody alluded to it. We did ask what they preferred, and people preferred the computer administration. They like it, and I agree that our experience with tablets is quite positive. They're really easy to use for people that you might otherwise have literacy concerns about.

So how about the setting? Get it in the clinic, get it at home. One of the bigger problems is mixing the two. Although it can be done, it should be done with caution. The strengths of getting it in the clinic are you're getting the real-time assessment, it's easy to implement electronic administration because you can feed it right into the electronic records.

The limitations were, as we heard earlier, impact on clinic flow, interruptions in the assessment based upon clinic flow. We once had a study that we couldn't get done in
the waiting room because they were so
efficient hitting one of their other
performance measures, and not having long
waiting times, that the waiting time we
anticipated to fill out the questionnaires
wasn't there, so we got missing data because
of the group hitting on the waiting time
performance. And patient distraction, anxiety
can be a problem in clinic, as well as staff
burden.

In the home setting, the strengths
are that it minimizes impact on clinic flow,
minimizes staff burden -- so, sort of the
opposite strengths on the home side. But
there are limitations to accessibility and
privacy and security and patient safety of an
anxious patient, or if an alert comes up and
the patient's at home, you have to engineer a
system to take care of that.

Last couple of things. Ethan
mentioned missing data and how to manage that.
There was a fair amount in the paper about
that. I'll just say that there can be bias introduced by missing data. There often is bias introduced by missing data. I won't go through the methods to do that, to help enable some time for discussion.

And there can also be this influence of, over longitudinal assessment, patient adaptation or even response shift, where the patient's own sense of what a number means on a scale changes over time.

And you may then think that you've measured change, when in fact you're actually measuring the patient's internal barometer for what an 8 means on pain, because they never knew what a 10 could be until they had a 10, and so now their former 10 is now an 8, as an example.

So this leads into the CAT demo. Most things that you've used in the past, and we've all used in the past, have been built on classical test theory, which estimates a person's level based upon a summing up of all
the questions they answer, like the way the
SF-36 is scored. And that produces a test-
dependent measure. You have to ask all the
questions, and you really are dependent on
administering that entire test.

Item response theory is test-free. You can create different tests from pools of
questions and estimate the underlying thing
that you're measuring, whether that's pain, or
depression, or fatigue, physical function, et
cetera. It enables you to do a customized
assessment that includes Computerized Adaptive
Testing, in which you can tailor the questions
to the individual that you're measuring.

So you can have shorter
questionnaires that maintain good precision or
accuracy, even at the individual level, and
you don't have to have those long tests to do it. And patients don't have to complete the
same sets of questions along the way. They
can have different questions administered at
different times, if you prefer to have that,
or if the CAT selects that.

So the demo, this is just to give you the framework for PROMIS, this divide into physical, mental and social health, and then within each of those areas there are as many as 40 different banks across pediatrics and adult, on the physical side measuring symptoms and physical function, on the mental health side measuring various affects, principally negative affect but evolving positive affects, behaviors and cognitions, and on the social health side measuring social relationships and social function.

So I mentioned this pool of questions. They're calibrated. If you have a calibrated set of questions, meaning that every item is a measure of that underlying thing, that underlying trait, that's called an item bank. And when you have an item bank of calibrated questions, any subset, including one item from that bank, can be used to provide a score for that domain. I think that
has a lot of possibility in this kind of
setting, where you might want to have
provider-based measures of something like
depression or fatigue or physical function
that don't require long assessments.

The metric for PROMIS is a T-score
metric with a mean of 50, standard deviation
of 10. It's referenced to the U.S. general
population, 2000 census demographics.

So the tools from PROMIS are
derived from item banks. They involve
Computerized Adaptive Testing, which is a
dynamic testing -- I'll show you an example of
that -- using fixed-length forms, or you can
do health profiles of 29, 43, and 50 item
length. And then there's this global health
index that Dr. Mort mentioned earlier, which
is 10 items measuring physical and mental
summary scores.

Okay. So, here's how the CAT
works. We assume from the beginning --
remember, it's referenced to the general
population, mean of 50, standard deviation of 10, so that's what's represented in the lower panels, here. That's a distribution of around 50, a normal distribution of depression we assume in the general population. So we're just going to assume that any given person at the start of a test has a score of 50, and there's a large confidence interval around that score.

The best item in the depression bank to start the CAT is, lo and behold, the question over in the top right, "I felt depressed." And then the answers are never, rarely, sometimes, often, always in the past week.

The curves you see in the top right show you the probability of responding never, that's the black, rarely, that's the red, sometimes, that's the green, often, that's the blue, and then always, that's the light blue. So the probability of responding each of those answers increases as you get
more depressed, so the more depressed you are, the more likely you are to say you're frequently depressed.

And the lower plot shows you how much information that item has on that same metric. So it asks that question, the person gives an answer. And now we'll just run through the first question, the first answer. The person says "rarely." That statement has a T-Score of 52, the standard error's a 4, and finds the next question that's going to be most informative, in this case "I felt like a failure." You see the response characteristic curves and the information curves that the computer knows. That's why it picked that question.

The person says rarely, it estimates the score of 53, a little more depressed. Standard error goes down to 3. Then it picks the next question, "I felt worthless," says rarely, and then estimates 55, standard of 2.
Now, we can go on and on, and if you keep going -- that's three questions asked so far. If you keep going on, it goes on to eight questions, but you'll see the estimate didn't change really much at all, vastly, between 54 and 55. So we asked eight questions, but really only needed to ask three in this case, and got the same estimate. So with three questions, got a very precise estimate of this person's depression. That takes about 15 to 20 seconds.

Now, the last couple of things that I want to show you because I think it's pretty germane to this discussion. We have a fatigue item bank as well in PROMIS, and you can imagine these different programs in cancer, arthritis, heart failure, joint replacement, pain management, using different methods, different items, CAT's different short forms. You're going to get the same metric and the same meaning.

The PROMIS investigators did this
in a research setting in COPD, heart failure, low back pain, depression and cancer, and in each case measured fatigue, so I'll show you the fatigue example. And this is an animation, so you've got to kind of go quickly through it.

So it starts, remember, with mean of 50, standard deviation of 10, to remind you of that. COPD patients starting over at the lower case, stable are around 56, exacerbators are around 63, and patients that go from exacerbation to stable actually, lo and behold, go from 63 down to 56, which you'll see down there under the dotted line. Heart failure transplant patients start at around 58, and drop down to around 47.

Depression patients start up over 60, more than a standard deviation, and get better with regard to their fatigue when their depression is treated, after one month and then after three months. And now back pain, the same thing with back pain. So in all
these cases, you have different clinical
areas, but the same metric on fatigue, being
able to compare changes.

You see, they're all starting more
fatigued than the general population, and all
moving in the right direction after treating
the clinical problem. Same thing with cancer,
but we'll move ahead.

This is the last concept I want to
put forward, because I think it's also pretty
relevant to what we're looking for in this
context, and that's a PRO Rosetta Stone, or
PROsetta Stone, which is a project that we
have to link many of these different measures
-- PH29, CESD -- with the PROMIS depression
measure, as an example. We've done that. We
haven't published it yet.

And therefore, you can express --
you can administer the PH29, you don't have to
give that up. But you can express it as a
metric. We heard earlier from Dr. Burstin
that NQF does not endorse the instruments;
they endorse performance measures. This is a way one could think of it, that this metric of the mean of 50, standard deviation of 10, referenced to the general population, is a potentially endorsable metric that you can get by asking PROMIS questions, or PH29 questions, or CESD questions, but you're putting them on that common metric I showed you with fatigue.

This is a sample from work that we've done, comparing a fatigue questionnaire that we developed earlier in the cancer setting to the PROMIS fatigue T-Score. You can look them up and use these in your reporting.

I think that's the last slide.

Thank you.

DR. BASCH: Great. Thank you. In the interests of time, we're going to move on to our first speaker. Lewis Kazis, why don't you come up to the podium to speak, so folks can see you?

And while you're coming up, I'm
reminded to mention that, in addition to
PROMIS, which is an NIH initiative, there is
also a second government initiative under the
NCI called PRO-CTCAE that some of us in the
room have been involved with, that uses some
advanced methods to develop questions to allow
patients to report on issues related to
adverse events, or safety and risk, for anyone
who's interested.

DR. KAZIS: Thanks, Ethan. I just
want to mention that it's a delight to be here
today, in a conference that's really dear to
my heart. I've been involved in the use of
patient-reported outcomes -- in their
development, in the methods, and in their
implementation in the context of performance --
for more than 25 years. And in fact, BU
just gave me a clock for being at BU for 25
years, so that sort of reminded me how long
I've been at it.

What I wanted to mention before
going into what I was charged with talking
about today was that the VR-36 and the VR-12 were developed in the VA under the support of the Health Service Research and Development Service. And the VA, I think, has been at the forefront of patient-reported outcomes and performance measures for many years, going back at least 20 to 25 years, when in fact performance measures in the VA took on real import in terms of VISNs and their organization and resource allocation, and so forth. And I think the whole move in that direction began with Ken Kaiser, who was in fact the founder of this organization.

So the VR-36 and the VR-12 have been adopted by the VA in terms of some of the assessments that they're currently doing and have done historically in the area of performance, and have been adopted by CMS and the Medicare Advantage program specifically in terms of principal endpoints in their evaluation. The VR-12 is now used in highlighting those particular plans in a
particular two-year cycle that are either negative or positive outliers.

We've also recently developed and published a utility metric that's generated from the VR-12 that allows one to begin to look at cost-effectiveness in the context of plans, so that information is out there. And also, it's in the public domain, and free of charge, in terms of algorithms, imputation, and contextual fixes, as it has been supported by the federal government over many years.

So my charge today was to talk about the issue of bridging measurement tools, and the very first slide is what I consider to be a binding framework for the use of legacy measures. Those are the historic measures that David Cella talked about, and relate specifically to a set of items in a particular questionnaire, like the VR-12.

Just to mention that the IRT approaches have informed legacy measures, and have been, I think, an important methodologic
advance in terms of allowing us to perfect and
come up with even more precise legacy tools.
And clearly, I think the legacy measures have
informed the development of item banks, as
they have been used in the context of CAT, the
Computer Adaptive Test. And clearly, I think
this becomes a very important aspect.

So there is clearly synergies
between legacy measures, IRT, and the item
banks that they have helped to inform. I
think as we move forward, the more bridges
that are established between legacy measures
and item banks, the better.

As I see it, there's no real
silver bullet in terms of any single
assessments, or even bank of assessments, that
will work totally in terms of the complexity
of our health care system, and I think one
needs to move in the direction of bridges
that, in fact, tend to combine the use of
assessments that, in fact, have been used for
many years, the legacy measures, and the newer
CAT models that are out there.

So I'll just mention a couple of points in terms of what I consider to be the advantages and disadvantages of the legacy measures, and then talk very briefly about the IRT CAT measures.

The legacy measures have been extensively tested for reliability and validity across many settings, over many years and populations. A good example is the VR-12, which has now been administered to well over 5 million individuals in the VA, outside the VA, in terms of CMS and the Medicare Advantage program, and in other systems of care. In an average week, I receive about six requests for the VR-12 with their scoring algorithms.

Fewer resources now are needed to implement the legacy measures compared to CAT, and that clearly is an advantage, because the CAT measures do require resources in terms of their development implementation in clinical settings. The expertise to implement them has
matured, and they can be integrated with new technologies, and have been in the context of the internet.

Disadvantages, of course, include the time to complete the instrument that David mentioned. It's usually longer than the CAT. Instruments are less flexible to update and calibrate compared to the CAT, and they require larger samples to avoid spurious results.

If one looks at IRT and CAT, the advantages clearly are, they estimate personal level traits within subsets of items, they usually require smaller sample sizes, and they're less vulnerable to floor and ceiling effects.

The disadvantages, I think, clearly have been controversial, and some of them involve the differential item functioning. That calculation may be problematic for multidimensionality assessments, and that's where the probability

Neal R. Gross & Co., Inc.
202-234-4433
of responding in a different category varies across different subgroups, given equivalent levels of the underlying attribute. And an example of that would be age, for example, or gender, where in fact the DIF calculation may not be what we would want it to be.

So clearly, there's no silver bullet here. The IRT CAT is an important methodology to be implemented, but there are limitations in its use. They do require front-end technology to implement the instruments, and additional assistance is usually necessary to facilitate successful patient-technology interaction. So there is a resource-intensive aspect to its use.

Clearly, I think, real advantages are, it's useful in assessing change, there's greater precision of measurement when compared to the historic measures. There may be high start-up costs, and the software and hardware is commonly proprietary and expensive. So clearly, there are advantages and some
disadvantages.

The other issue is the computer literacy of the population that one is dealing with, and whether the CAT method would require a population that's computer literate. So one has to consider populations where, in fact, it may not be as advantageous.

So I do have more slides here, which will be in your packet. When you testify before the Congress, you usually say "It's going to go into the Congressional Record."

So, let me summarize here. The issue here is that I think there's no silver bullet in terms of one assessment tool, or even a simple set of assessment tools, that will work in the context of a very complex health care system. I think hybrid approaches become really necessary, that bridge legacy and IRT CAT approaches for purposes of application to systems for measurement performance, so that's really important. If
you take away anything from this talk, I think
the issue of flexibility and hybrid-ness
becomes really important in terms of what's
adopted and what's going to be used in a very
compact clinical setting.

             Mixed mode approaches are
necessary, so that flexibility in the
protocols is possible in real world settings.
And David Cella, I think, did an excellent job
talking about the different modes of
administration. And clearly, I think, for
those non-methodologists, there needs to be
fixes in terms of the empiric data that, in
fact, might be biased by whether, in fact, the
administration is done at home, whether it's
in the clinic, or whether, in fact, it's face
to face, whether it's through computer, and so
forth.

             Those things can impact on your
results, and one needs to adjust for those.
There are contextual fixes. We've spent many
years developing contextual fixes in the
development of the VR-12, so that if there is a user out there, we have an algorithm, and those fixes can be made, so that when one generates a score from the VR-12, that metric can be considered to be reliable.

So, mixed mode approaches are necessary, and there need to be fixes for these things. And then the issue of missing data, which is a fact of life in real world settings, adjustments for missingness is required to adjust for bias in results. We've developed an algorithm called the Modified Regression Estimate which, in fact, controls for missingness.

So if you were to have as few, for example, as three items present in the administration, when you've administered the VR-12, one can generate some scores on the basis of that. So, one needs to consider that. One needs to consider the biases that, in fact, would be generated as part of that missingness.
And that's all I have to say.

DR. BASCH: Great. Thank you very much. So, Richard, you're up next.

DR. BANKOWITZ: Thank you. I'd like to speak about the experience that some of our premier members are obtaining in implementing some of these instruments. We are an alliance of 2,600 hospitals across the U.S., and we try to accomplish a lot in a collaborative methodology. So we now have an ACO implementation collaborative consisting of 30 health care systems, and in that context of that laboratory we are trying to gain experience with some of these instruments.

So we're looking at a variety of things. We have two of our members, Fairview and Geisinger, who are looking at the Dartmouth Institute Primary Care patient-reported measures pilot. We have two of our members who are looking at the Southeast Minnesota Beacon Community and Mayo Clinic patient-reported outcomes. And then we've got
three -- South Coast, Bay State, and St. Francis -- that are trying to implement the Dartmouth Institute "How's Your Health?"
We've also been approached by the Gallup-Healthways, to see if we'd like to take up their well-being index, but so far that's in the very early stages.

So I'd like to report on some of the very early information we're getting in the use of the Dartmouth "How's Your Health," and I think that might help inform some of the methodological questions.

So this is very early, and we've got a very small sample size, but I think we're getting some very interesting information from it. First of all, the data comes in in a variety of methods, so some are filled out in the doctor's office, with paper and pencil. Some are done via the internet, before the patient appears. And none of these three systems has anything like the sophistication that Gene Nelson showed you or
that Liz Mort showed you with the data infrastructure.

So one big problem was, what do you do with the data? If it's filled out over the internet, how do you even get it to the physician? There's a concern about simply using email that's not secure, so you need to have a secure email server. One of the institutions is trying to put it on their portal, but then the question was, how do you get it to the right physician? It's fine that it's in our portal, but how do we make sure it goes to the right physician at the right time, so that it gets incorporated into the record?

None of these systems can integrate with the EHR at the moment, so the best they can do is scan a document and that goes into the EHR as basically a photograph. I mean, you can't search it or do any kind of structured analysis with it. So that's a big problem, just in terms of implementation.

But we're getting some very
interesting feedback from both the physicians and the patients. Some of the feedback we've had from the providers -- and we only have three, but they've said things like "The use of this instrument really establishes rapport with the patient." "It gives me a jumping-off point for a discussion." It helped me identify patients who had inadequate knowledge of their condition." "It helped me identify problems at home that I was unaware of, problems with their feelings and social phobias," so quite a lot of information. It provided risk stratification for future hospitalization," so maybe identifying patients of particular risk. "It enhanced patient empowerment." "The patients feel more included." "I can identify confident patients, and vice versa."

And then one physician said "The most value came from identifying patients who felt unable to manage their help problems. This was really a good use of the tool, but it
took a lot of office time." So it's not necessarily the case that just having this information in front of you is going to make the physical faster. It often makes it go longer, because you've got more issues to deal with.

The patients had a variety of responses. First of all, most of them said they would recommend it to others. 85 percent said they would definitely recommend the How's Your Health survey to other patients. One patient said "I took the survey, and I'm healthier than I thought I was." And that's interesting, and I think it also has implications for response shift. If the instrument itself is making people have a different expectation of their health, that's an interesting finding.

The patients say "I'm glad to know my physician was interested in what I'm thinking." "It's good to have the information available before I see the doctor." "I learned
new things." "It gives me time to think about
the answers to the questions." "I think it
helps patients who don't ask questions while
they're at the office." "I liked the reading
materials." And then some said they wished
they had more ability to put explanations in.

So as I think about these
responses -- and they are limited -- I think
there are a couple of lessons, maybe, to be
learned. One, as we discuss these
methodological issues, I think it's important
that we not let the perfect become the enemy
of the good. And I think it really depends.
The first question you have to ask is "What
are we going to use the information for?"
Because that may dictate how precise we need
to be, and it may dictate the operating
characteristics of the test that are required.

So as I think about this concept
of useability, I think it's going to be the
most important concept as we go through the
measures endorsement process. So it may be
not precise enough to say this is a performance measure. Is it a performance measure for internal performance improvement? Is it a performance measure for transparency? Is it a performance measure to compare two institutions? I think that's an important question.

And so, as I thought about these responses from the physicians and the patients, they tend to fall into two classes. Class one is actually dependent on the answer, so we found some patients that are at high risk, or we found some patients that don't know a lot about their health, and we can act on that. But then there's a second class which is really not dependent on the answer: establishing the rapport, being more included, having the patient feel the physicians are more engaged. That takes place just by the process of the instrument alone.

So, how would we incorporate those two things into performance measures? It came
up earlier as a structure measure. I think of it almost as a surgical checklist. It's something that should be done. I'm not sure we would take each item on the checklist to compare institutions. Institutions might want their own running tallies, so they know if they're getting better. But the use, I think, is really going to be key.

So that's one challenge, and I think the biggest challenge. The second challenge, as I think about these measures -- and this has also come up today. We're moving from measures of sickness and illness to measures of health and well-being, which I think is a very good progression. But I think when we do that, we have a dilemma. Because as we move to those measures of health and wellness, we then begin to have shared accountability, right?

So yes, the ACO, the health home, is responsible for some things. But then we have the school, we have the community, we
have access to fresh food and playgrounds. We have a lot of items that impact health and wellness. So how do we apportion that accountability? It's going to be a very important question. Not one of the traditional methodological questions, but I think a key one going forward.

So, I'll stop there.

DR. BASCH: That's great. Thanks very much. So we're going to finish up with Lori, and then we'll take some questions from the audience.

DR. FRANK: Great. Thank you, Ethan. So now I'm between you and lunch, and also between you and the audience engagement piece, so I'll move quickly through this.

I do want to discuss what patient engagement means, how it is and is not currently being implemented in the course of PRO development and evaluation, with the hope that reframing patient involvement in PRO research can improve measure development and
testing and enhance the value of PROs for use in clinical settings and for performance measurement. And I use the term involvement very mindfully, a point I'll return to in just a moment.

PROs offer one way to capture outcomes meaningful to patients. They are not the only way. They don't always succeed, and they're not the only way to do so, which I think is a point we all need to keep in mind.

When I was in kindergarten, I remember we were promised that by the time I was in sixth grade we would have the ability to make phone calls to people and actually see the faces of the people on the other end of the phone. And I'm glad the technology is finally here, but I still carry with me the disappointment of all that delay, and the lost opportunity, with the length of time it took to bring that to be. And that's how I feel about use of PROs generally, but method of administration for PROs specifically.
In the last century, I was part of a panel on the promise of ePRO, I think with some people here. And again, it's been a bit disappointing, the rate of adoption. I think that patient engagement might be a way forward, and I'm particularly interested in how that can apply to the tremendous contribution that the PROMIS initiative stands to make, if only we can get the uptake there, like we need it.

I also want to say that I think that -- I appreciated David Cella and his co-authors' paper, with its thoughtful consideration of all the methodologic issues that we need to think through before going full bore towards PROs in performance measurement, but before we continue down this track, or in keeping with the innovative technology theme, the high-speed MagLev, I think there are a couple of other trains that need to be connected, and those would be patient-centeredness and patient engagement.
In his 2009 Health Affairs piece, Donald Berwick defined patient-centered care as "the experience of transparency, individualization, recognition, respect, dignity, and choice in all matters related to one's person, circumstances, and relationships in health care." I think that there's a lot in that definition, and a lot actually in that piece, that's valuable as we consider the role of engagement in improving measurement.

We do need to make that distinction between patient-centered clinical care, as he's talking about it in that piece, and patient-centered research. For research, we need to further differentiate between clinical research and methods research, and right now, as you know, PCORI is funding clinical research, but we did just release the funds for the methods research. And out of the 50 funded projects, 11 of those dealt with PROs. Nine of the 50 deal with Computer Adaptive Testing or other technologies. So we
are moving forward, but I think that there's
a role for funding agencies here in this
dissemination and uptake.

I would add to clinical care, and
to research, when we're thinking from a
patient-centered standpoint, performance
measurement, obviously. What is the value to
adding engagement to a performance measurement
view?

In their discussion about the
paper, about the potential for PRO use in
clinical care, David and the co-authors
mentioned patient-provider communication and
identifying patient needs in a timely manner.

I think it helps to view those
sorts of statements from a patient-centered
perspective. Improving communication is a
form of patient engagement, which enhances
patient-centeredness. Identification of
patient needs ensures patient-centeredness.
So together, the engagement and patient-
centered orientation, from the clinician and
from the health system, can improve health outcomes.

And I think that was well catalogued, some of the empirical evidence for that, in the paper. And certainly there's evidence of the value for patient involvement in improving content and construct validity of our measures. But I suspect that, without going further, we will miss out on some opportunity for some more meaningful information that we could get to improve our measurements. So it's an exciting opportunity, and I'll say more about that in a moment.

I just wanted to review, then, principles of engagement. You heard some in the quote from Donald Berwick. Trust and transparency, leading to respect. Partnership and collaboration, including co-learning and communication. There's an inescapable interactive element to patient engagement, a relationship element, which is why shared
decisionmaking can be considered a form of patient engagement.

So I have this virtuous cycle here of engagement, as just a way to show us that there's engagement in research. Right now, a lot of what's happening in PRO methods research is involvement, where patients are subjects, but this is a giant step back and showing an enterprise view of truly engaging patients at all phases of the research process, and not just as subjects. And the idea is that there's some measurement value that we can recognize with this view.

The next slide shows the same schematic, but using performance measurement, then, as the organizing principle. And there, too, I think few have so far contemplated -- we've heard from some this morning, thankfully -- what taking this engagement view can really do to improve the methods behind performance management. It can really anchor us.

So this morning, Lisa Lentz
mentioned patient attribution. I think this framework can handle that. Linda Wilkinson mentioned culture. Who's defining culture? I think this engagement framework can help with that. On the first panel, we talked about walking in the patients' shoes. Certainly, this is a framework that helps us to achieve that.

How much collaboration do patients really want? Nobody knows the answer to that. We don't have good data. But the idea is that there's an ethical argument to including patients, engaging patients this way. We're at the upward ends of our ability to use empiricism here. I heard a lot of good, empirical questions. Is clinician performance improved by use of PROs, for example, one of the questions raised this morning. But there's the idea that we need to accept the principle of engagement before we can move forward for some of the methods improvement.

In Table 3 in the paper, it's
important characteristics and best practices
to evaluate and select PROs as performance
measures. I would suggest that we add patient
engagement in development and testing as one
of those important characteristics. We still
need to develop metrics for patient
engagement, so there's a lot of work to be
done, but I think that that would help us to
reframe in a positive way.

Under content validity, perhaps
that evaluation, whether the outcomes are
patient-centered or not. For performance
measurement, they need not always be, but
asking the question might help to improve our
output.

In the discussion specifically of
method of administration, the authors
reference patient burden. How do we know
what's burdensome to a patient? We're
inferring, and certainly as researchers we're
also patients, so we can figure this out. But
to remember, might actually disadvantage us to understanding the patient view. So incorporating patients into the research team, then, is a way around this conundrum.

So I'd say consider a continuum from low patient input to high, proximity to patient voice to distance to patient voice, and think about what we're doing. A psychometric evaluation study, where is it on that continuum? The patients are certainly providing input, but through channels that have been engineered by the researchers. Cognitive interviewing is bidirectional. We're talking to people. It's qualitative. But here, again, patients go off-topic all the time, and it's actually our job to keep them in the channels that we, as researchers, have engineered.

Focus groups, another great opportunity to collect patient input, and there's a lot of value to be derived from it. But there, too, we're missing an opportunity,
by limiting ourselves to these methods, to
really sit down and get the full benefit of
researcher wisdom plus patient wisdom
together. So on that continuum, then, putting
the patient at the center would be patient-
centered outcomes research.

A lot of barriers noted in the
paper for use of PROs, I think many of those
might be system-centric and not patient-
centric. Just a point to note.

Of interest, the authors say "For
those developing or modifying measures
according to principles of universal design,
they're encouraged to consult with relevant
experts." Well, who are those experts? Will
it include persons with disabilities? Is it
going to be about us without us, as people
say? But it's another excellent opportunity
to improve measurement through engagement.

One idea, then, is to create a
task force on measurement error and invite
patients to participate. Are there some
things that we could learn about improving our
methods from the patients, that we just
haven't thought to ask? I think that method
of administration is a great entree into this
particular type of thinking.

A quick point. There were some
questions specifically about proxies. The
FDA, as you know, has a taxonomy of PROs,
ClinROs, observational measures. I think in
the case of proxies, we could come out with a
better taxonomy. There are true proxies,
people who really can accurately report for
the individual. Some parents can really tell
when their child is fatigued, for example.
There's quasi-proxies, people who can report
but with non-ignorable error. Then there's
just poor proxies, people who don't do it
well.

And to Mary's point earlier, I
think we need to always be mindful, are we
treating the reporter as a proxy or as an
informant? Someone who can have some insight
into the patient, but not complete patient reporting in their stead. And Ethan reminded me, too, about always going to the proxy in the case of kids, or in my area, for those with cognitive impairment and dementia, when actually there's the possibility to obtain accurate reporting.

And the question here is about truth. How do we know that we've gotten to the truth, to the accuracy? To a certain extent, I think we've been a little lazy in the field, and more phenomenological research, which is a form of patient engagement, can help really get us to that truth of what's accurate. When do you need the proxy, and when is the patient truly the accurate reporter?

Missingness was also raised, and my only point here is that missingness might be a form of revealed preference, and there, too, engaging patients as part of the research team could help lead us to a wider understanding of causes of missingness.
Some view patient engagement as a shift in the power relationship. If you think about it, the patient completing the survey in the parking lot holds a lot of power. They're going to ruin our study. They're going to ruin our performance measurement. So let's acknowledge the power that each party has here.

It's a scary notion, to think about giving away this power, but I think it's one that's worthwhile and that can help us with improving our PRO measures, and improving our measurement overall.

DR. BASCH: That was terrific.

Thanks so much, Lori.

You know, I should mention Lori and Mary Tinetti, who's here, and I, have done a fair amount of work within PCORI around patient engagement, including issuing a couple of contracts to do a landscape overview and a systematic literature review around methods for patient engagement. And our conclusion at
the end is that we really do need systematic research on approaches to engaging populations that will be informative to the scientific enterprise. I really do see a lot of synergies between PCORI's interest here and NQF's interest here, PCORI on the research side.

So we have about 15 minutes for Q&A. Just to set this up before we start, we really do want to focus on administration methods. We recognize that many of the themes cross over to other areas, but that really is our focus. So we have Lewis, who talked about scientific challenges, Richard, who talked about implementation issues, and then finally Lori talking about the special challenges in developing PRO measures for performance evaluation, because we need to actually engage with those from whom we're gaining information. We'll start in the front.

DR. FIHN: So, this isn't specifically about administration, but one of the themes here, at least that I've heard, is
heterogeneity of multiple approaches. Is that in conflict, ultimately, with sort of organizational imperatives right now, in terms of accountability measures? What we see a lot, where we try to convey the complexity of measurement, at the end of the day what often trumps is comparables in a very competitive marketplace, and the question of sort of how those trade-offs will work out when we develop a very complex and rich system of measurement. And at the end of the day, for accountability reasons, not for improvement reasons, but we didn't name that as one of the goals here, there's going to have to be some reconciliation for comparability across systems and organizations, or even within systems.

So, how does this all play out?

DR. BASCH: From the panel?

DR. FRANK: I think the PROsetta Stone is a great example of the way to begin to really cross communicate.
DR. BASCH: Could you be a little more specific about what you mean by heterogeneity? Do you mean heterogeneity across patients or across contexts, or do you mean the measures themselves?

DR. FIHN: Yes.

DR. BASCH: You mean the measures themselves. So, go ahead, Lewis.

DR. KAZIS: I think it's an excellent question. We were involved in a study done a few years ago that was published in Health Services Research comparing the Veterans Administration to the Medicare Advantage plan. And we looked at mortality, and then we looked at measures of outcome using the VR-12. In that context, the VA actually did better in the adjusted analyses. In terms of the differences, they were quite dramatic. And for those that would be interested in discussing that further, I'd be glad to talk about what, in fact, we hypothesized as why there were differences, but the
VA did a lot better.

Now, in that context, we had the luxury of similar assessment tools across the two systems. Going forward, I think that it's going to be a real hodge podge in terms of the assessments that are used nationally, depending on the organization and what's adopted, and so forth.

I think what is absolutely necessary is that there are adjustments that are developed to deal with those differences, differences in assessment tools that are used. I think David Cella can speak to the IRT and CAT, which I think will allow for item banks that, in fact, might permit comparisons across different systems of care.

DR. BASCH: I think Albert had a question.

DR. WU: On the topic of missing data and biases that might come with missing data, there's another thing to consider, and that is that there may be biases in present
data. I don't actually have any data from our system on this, but I'll give you an example. I was teaching a course, and we wanted to get student evaluations, essentially. Student satisfaction data. And we wanted to get our response rate up. We always got pretty good response rates. We then made it mandatory. We actually gave people a point on their final grade if they would turn in their evaluation.

And what happened to our evaluations? We got 100 percent response rate, and our evaluation went down, because the bias is that people who are more satisfied are more likely to respond. So I think the idea about looking at response rates is something that we do need to consider, and it's just another topic.

DR. BASCH: That's a great point, actually. I think at the next meeting we'll have another presenter from the NHS PROMS initiative across England, and this was actually a phenomenon that they observed.
across the U.K., which is there were generally
lower scores for provider systems that had
higher response rates, and they've now had to
adjust for response bias.

Do we have a question over here?

Go ahead.

DR. GANIATS: I came here -- this
is Ted Ganiats. I came here today open-minded
but concerned that we would not be able to
come up with accountability measures, and I'm
just really happy to say that I was right.

(Laughter.)

DR. GANIATS: And everything
that's been said today is just absolutely
fascinating and absolutely wonderful, and it
promotes clinical use, it promotes patient
engagement, and it promotes quality
improvement. But I've yet to see anything
that gets to my methodologic concern, and
nothing that gets to my practical concern
regarding accountability, as was mentioned
earlier.
And the reason this is important now is because, if we're able to limit ourselves to quality improvement, that helps address the methods approach that you asked that we discuss during this question and answer period.

The methodologic concern that I'll just throw out is that, in general, performance measures come from guidelines, and guidelines come from evidence. And we are bypassing the evidence, and we're bypassing the guideline, and creating a performance measure without the structure of a guideline. So we're going to be inputting into practice requirements prior to the guideline having been created.

And if we want to do that for quality improvement, that's fine. But I don't think that's good for us -- the NQF won't, but who's going to create the quality measure, sans guideline, which the NQF is then going to evaluate without the evidence and without the
It can happen. I hope it happens, and I think it's years down the way. We have the phone without the vision right now. But for quality improvement, I think we have a lot of good evidence that we should be moving in that direction, and that will help us, then, in selecting the methods for the questions, which is what you wanted to address this time.

DR. BASCH: Yes. That is a long, complicated question. One could argue that guidelines generally describe the phenomena to be measured, as opposed to the measures of those phenomena themselves, which is what we're talking about here. But it is a blurry line. I think we probably would have to leave that for another -- let that linger with us, a very important --

DR. BANKOWITZ: Can I speak to that momentarily?

DR. BASCH: Go ahead.

DR. BANKOWITZ: I think it's a
great question, and it also relates to the question that was asked at the front of the room here. And I think one way to look at it is, we might want to think about leading measures and lagging measure.

So a lagging measure is a big dot measure. It tells you how well you have done. It's too late to change it, but it tells you how well you have done. So mortality, 30 day mortality might be a lagging measure. And there's no guideline that says "Don't kill people," but it is a measurement that is valid. So if we can come up with a big dot measure, then we need to ask these questions about how to we adjust it.

But we might have leading measures where there might be more heterogeneity, which might vary from place to place, which might not be so well documented. And I think that might be one way to address it.

DR. BASCH: Did you want to follow up quickly?
DR. GANIATS: Yes. I mean, just real quick, I've been on guidelines. I've been on performance measures. The 30 day readmission for heart failure, I co-chaired the committee that created that. Why 30 day? Why not 15 day? Why not 45 day? Where's the evidence that we're going to draw the line? And there was no evidence. Personally, even as chair, I didn't like the measure because it is arbitrary, and I think that there are huge problems.

Most of NQF-endorsed measures are process measures or structure measures for a reason. There are a few outcome measures, and I think that we have a problem when we are creating outcome measures without the guideline, and we are pushing practice through a performance measure accountability instead of through a guideline process.

Believe me, I'm in favor of us being better. I'm just worried about the process. But I'll let others talk now.
DR. BASCH: In the middle.

MS. TORDA: Yes, I am Phyllis Torda, from the National Committee for Quality Assurance, and I'd actually like to pick up this thread and the suggestions about structural measures.

And I think, as Helen noted in the very beginning, there is a maturity progression from structure to process to outcome, and that one of the things that it might be really helpful for this group to do is to lay out what should guide thinking about structure, process, and outcome measures for patient-reported outcomes.

And it does sound to me like we might be getting to the point, based on the presentations, where there is a good evidence base for a structural measure. Structural measures are good at signaling the need for implementation, encouraging implementation, without stifling innovation, because you're not being overly prescriptive. And then, as
you learn more, you move from structure to process to outcome.

So I think there's a progression there. I probably agree that there's a lot of methodological reasons why jumping to outcomes might be premature at this time, but I think that thinking about that progression would be very useful.

DR. BASCH: That's a very good point. We're going to go over here, and then we had a comment here.

Go ahead.

MS. HUFF: This is Jennifer Eames with PBGH, and I want to thank the panel for having a discussion on methodology that incorporates the component of useability, because I think we can't really separate the two without knowing how we're going to be using the measure that really influences the methods.

And I think there's a general belief of quality improvement, you don't have
to be as rigorous, public reporting, some more
rigor, payment, the stakes are higher. And
I'd like to challenge that belief in terms of
who's using the information for public
reporting and payment. You look at the users,
and I'll say that purchasers have a much
higher tolerance for error in this
measurement, and would rather have information
now as opposed to waiting for years for it.

Also, for consumers, there was a
study done by Judy Hibbard and Arnie Milstein
who looked at the tolerance for
misclassification when comparing providers,
and it's much higher amongst consumers, I
think, than what public reporting programs
usually do. So again, I think this is a
really important area, of tying the methods to
the useability, and looking at who's really
using the information.

And per the conversation around
using structural measures or checklist
measures, I just want to add a word of caution
to that discussion. Because I think those measures are good. They're good at encouraging implementation, but they don't necessarily get at the quality of the implementation, and I think that's where the outcomes really come in, and why people are interested in outcomes is it gives a sense of how well the implementation is really occurring.

DR. BASCH: We've been given authority to go about five minutes over, so we're going to go to the back here, and then we're going to go to the audience and the phone.

DR. PAWLSON: Greg Pawlson, from Blue Cross Blue Shield Association. I have to weigh in on this, the need for guidelines for outcome measures. NQF, and I don't remember, Helen, the exact group, but I remember it was a task force. Superb discussion, and it was written up and published by NQF, about where guidelines really come in.
And there was, I think, a fair consensus that, for most outcomes, you don't need a guideline. Pain, patient-reported experience of care, we've never had guidelines that said "Patients should be treated with respect and trust and so on." And I would say it's the same for many of the patient-reported outcomes that we're talking about.

Now, as those get into areas that are more process-oriented and need a guideline like, to bring up one, Ted, hemoglobin A1c levels, then clearly there is a need for guidelines. I think it's a different kind of evidence base that doesn't come out of guidelines, but there is some evidence base in most of the measures that we've been talking about today, that some clinical intervention does have some impact, especially at an aggregate level.

You can't hold an individual physician responsible for some of these things. You can for patient experience, but
you can't for some of these others. But clearly an accountable care organization, if they don't have any effect on that, on mortality or on readmissions or on a number of other -- and I should stick to patient-reported outcomes -- then I don't know what we're doing in health care, frankly.

So I wouldn't be so pessimistic about our ability to have evidence-based -- not guideline-connected, but evidence-based outcomes with very high face validity, that can now be measured and can show change, especially over time, at an aggregate level.

DR. BASCH: I think this speaks, really, to the need for research, rationale, consensus. And to Ted's comment, if you look at a measure -- for example, in oncology, we look at post-chemotherapy nausea, there are actually guidelines that recommend improving nausea after chemotherapy, but patient-reported outcomes are generally not used to measure that. So in some cases, there are
guidelines. In other cases, it would be quite easy to develop consensus around those domains that merit measurement. But it does, again, speak to Lori's comment for the need for empiric research in this area.

One could argue that, in fact, the rationale for developing these measures is, in fact, patient engagement, that engaging with populations to determine what is important to patients as outcomes would then logically lead to the development of measures of those outcomes that are important to patients, and then we could do research to demonstrate that the measures being developed actually yield meaningful differences in measurement between practices, and that's probably the continuum.

To the audience?

(No response.)

MR. CUNNINGHAM: The phone? Operator, can you tell us if we have anyone in the queue for questions or comments?

OPERATOR: Yes, sir. As a
reminder, you may press star-one to ask a question.

There are no questions at this time.

MR. CUNNINGHAM: Thank you.

MR. BLUM: Hi, Steve Blum with Forest. I just wanted to make a brief comment relative to proxy. There's been some discussion lately about potential utility for proxy or observer-assisted reporting, where there may be an opportunity to either assist with recall or frame the question in a way that the subject would understand, or provide some context for their response, which may address some of the shortcomings with both proxy or patient-reported, maybe get closer to the truth by putting the question within a context that the subject is able to respond in a way that they wouldn't otherwise, if they did it by themselves.

DR. FRANK: Yes, and I would just comment that, in the United States, we have a
political and a legal system that cherishes
the autonomy of the individual, but that's a
great example of when shared decisionmaking
with the family, being those whom you're
sharing it with, comes into play. And that
crosses over into measurement, then, as well.
Absolutely, it's patient, dyad, or family
grouping as the measurement unit.

MR. CUNNINGHAM: Are there any
other questions from the audience before we go
back to Barbara Gage?

OPERATOR: There are no questions.

MR. CUNNINGHAM: Thank you. Barb?

DR. BASCH: Go ahead.

DR. GAGE: Thank you. The
discussion is changing a little bit. There
were comments about, really, the insurer's
perspective, although we didn't call it that,
when we were talking about accountability and
value. And typically the accountability is a
measure of reduced readmissions which, while
we all talk about it as a measure of quality,
much of the force that's behind it is a reduction in the cost.

But some of this discussion has just broadened to identify patients' preferences as a value metric, even though an insurer may -- I mean, that's not a common metric from an insurer's perspective. They're typically more concerned about costs.

DR. BASCH: Panel?

DR. BANKOWITZ: I think one could also look at the purchaser's point of view. I think large companies who are employing and insuring their populations do want to have access to quality care. They are concerned with the quality of decisionmaking. So maybe if you take one step back in the chain, and look at the purchaser, that might be helpful.

DR. FRANK: And I would just add that preference-concordant care may actually improve efficiency.

DR. BASCH: Richard, to follow up and to draw that out, would you say that, on
the payer side, there's actually a preference around improved symptoms, because of its relationship to utilization?

DR. BANKOWITZ: To answer the question, I do think that a useful metric from the purchaser's point of view is probably healthy days at work, speed with which one returns to work, this kind of metric which would be useful. I don't know how it plays into patient-reported outcomes, but I think clearly employers are concerned, increasingly, with the physical, mental and emotional well-being of the workforce. So it is becoming a broader, I think, discussion.

DR. BASCH: I think we have time for one more, and then we'll finish up.

DR. KOTAGAL: Hi, Uma Kotagal from Cincinnati Children's. I wonder if I might pull this together a little bit in a broader context. So if we take children with asthma and they are readmitted, it is not only a cost issue, but it's a missed school day issue, and
an absenteeism or presenteeism issue for the
parents, and therefore for employers. And the
evidence for that is pretty strong.

Now, if we want to reduce
readmissions for asthma, and we publish this
data, it turns out that in parent self-
confidence and understanding, recognizing
early symptoms of asthma is important. And
when we first began this measurement in about
45 practices in Cincinnati, with about 15,000
children with asthma, half the time the
parents reported that they did not feel
confident about managing their child's
illness.

When this got filled in in the
waiting room, and the form was given -- this
was a paper questionnaire -- to the physician,
they were shocked at this response. And they
said things like "Really? You don't know how
to manage it? We've been talking about this
for so long."

Now, of course, when we recognized
that half the parents of these 15,000 children
didn't understand early signs of respiratory
illness for asthma, or how to prevent a
hospitalization, and began to work on it, we
not only could reduce our hospitalizations
significantly, but we improved functional
outcomes for the children, in days missed and
so on.

So I think it's difficult to think
about this conversation in a unidimensional
way without recognizing that there's both a
developmental sequence to it -- i.e., we don't
quite know how to incorporate all of this --
there's a research significance to it, how are
we going to measure the right thing as what's
important. But there's a larger context of
connecting the individual, the family, the
society and the context. And when we
dissociate these pieces, we end up with lots
of interesting conversations that don't
necessarily result in better health.

So I just want to use asthma as an
example, to bring together the utilization
question, the child question, the parent
engagement question, and the employer
perspective.

DR. BASCH: It's a very nice
comment to finish it off. I want to thank our
terrific panel and everyone in the audience,
and hand it back to Karin.

MS. PACE: Just one quick
announcement. We have lunch for the expert
panel and authors. For the audience, I think
you were given as you came in some quick
places to go that are within a block of here,
to go get some lunch. We'll reconvene in 45
minutes, so 1:55. Anyway, thank you again for
your participation, and enjoy your lunch. And
we'll reconvene in about 45 minutes. Thank
you.

(Whereupon, the above-entitled
matter went off the record at 1:11 p.m. and
resumed at 1:55 p.m.)
A-F-T-E-R-N-O-O-N  S-E-S-S-I-O-N

(1:55 p.m.)

MS. PACE: We are going to go ahead and get started. So I will have the panel take their seats. And Al will start us off.

DR. WU: Thanks. You almost said, "Take your sleep." And I realized that we are post-lunch. So we will try to keep this lively.

So this is the session on selecting patient-level PROs. I have to check my notes. I am Albert Wu. My institution thinks I'm from the Bloomberg School of Health, but I am listed here as being -- I think our corporate member is the Johns Hopkins Health System.

So we have been -- we, like quite a lot of other big academic medical centers, have recently decided to go with a large vendor-provided electronic health record. And so we are now scrapping our legacy system that
has actually been working reasonably well finally after 15 or 20 years.

And we are faced with a number of choices. And one of the choices that we are faced with is -- well, it's an opportunity, really, to incorporate patient-reported outcomes into the electronic health record. And part of the reason is that our product comes with a built-in patient portal that has a couple of PROs built in, the PHQ-9 and the RAND-36, a close cousin of the VF-36, I think. But those are the only things that are built in. And we now have the opportunity to customize our system to include other PROs. So the question is, which ones?

For those of you who are not among the cognoscenti of this field, there are thousands of PROs, dozens of generic measures, quite a few utility-based measures, and then hundreds and hundreds of specific disease or treatment-specific PROs. And so which of these are we going to include in our
electronic health records?

These conceivably will be used for research in the future, which we're certainly very concerned about. They could also be used for quality measurement, internal quality improvement, and ultimately for accountability. So we've sort of got our eye on all of those things.

There are a number of issues we are confronting, one of which is that there are many proprietary measures that are among the best tested, not a surprise, really. Should we be using these? Are we going to have to pay for them? Can we get a site license for our five hospitals and God knows how much in total dollars going through or are we going to pay on a per-use, per patient-use basis?

How do we know whether a tool is good enough for research? And, perhaps a little more problematic, how do we know which are good enough for clinical use? And, not
identical to that, which are most suitable for use for making comparisons across organizations? Are there tools that have known properties so we know how they are related to other variables, particularly things that are complicated, like patient personality? Do we know anything about how personality is related to the scores that people give?

We have suspicions. And Ted Ganiats was sort of voicing a little bit about that. Some of my patients have profiles of patients who have sort of a very similar personality profile, maybe because they are nicer than I am. And so how do we take that into account or can we?

We understand that tools need to be blended for research use and clinical use and quality use. Which are the most acceptable to both patients and providers? Because, if nothing else, very selfishly, that is going to affect whether people give good
answers, whether they give valid answers, whether they give reliable answers, whether they take it all seriously. This is another way of saying that it has got to fit into the work flow, but patients and providers are a part of the work flow. So selection is important.

We would like things to be interpretable, both by patients and by providers and ultimately by others we may be accountable for. We would like them to be actionable, if possible. So I hope we talk about some of these things. I am sure we will.

You have slightly detailed files in the back of your materials, but we have got a terrific panel of reactors, each of whom is going to speak for seven minutes. And I will try to be mindful of time.

Jim Bellows from Kaiser Permanente; Gene Nelson, whom you have already heard from but you are going to get a little
bit more from, from Dartmouth; Kalahn Taylor-Clark from the National Partnership for Women and Families; Ken Ottenbacher from the University of Texas at Galveston. And, even before that, I think David is going to tee things off for the rest of our discussion.

MEMBER CELLA: Sorry. It's me again.

(Laughter.)

MEMBER CELLA: Okay. Well, I think we're getting into some things that many of us have been waiting to talk about: how to select PROs for performance measurement.

I think it is fair to say we probably could all agree that we want instruments that measure person or patient-reported outcomes in a standardized way and that we can know something about the properties of the questionnaires.

There are a lot of guidance documents available to address attributes of patient-reported outcomes used in research,
but there is little guidance regarding
attributes for patient-reported outcomes to be
used as performance measures.

So we drew from one of the more
frequently, perhaps the most frequently, cited
guidance on user PROs in research. And that
is the medical outcomes trust.

So you will see in table 3 of the
paper and text accompanying it an outline or
structure that draws heavily from the medical
outcomes system original criteria.

So next slide. So there are some
differences, but before I get into maybe a
couple of the key differences in selecting
PROs as performance measures versus research,
there are I think more similarities than
differences in terms of wanting something that
is reliable and valid and interpretable, et
cetera. And we'll come to that.

But one thing that's pretty
different about performance measures, unlike
research, is that it is really important to
have short length. And the second thing that is pretty obvious is that the stakes are different and the stakes are higher in use of PROs as performance measures than in most research studies.

Established PROs have a lot more evidence behind them in terms of their usefulness as performance measures than the newer PROs, the ones we have talked about, like using IRT, but the newer ones tend to have better measurement properties. So you have this always-present tension between going with something that has been around a long time many have used, familiar with, lots of literature, like the SF and VR tools. They are limited in being static measures that are scored using classical scoring methods that do require you to administer the entire thing and not have flexibility around that.

Future direction, which I talked about this morning, in IRT-based measures, like PROMIS, is a contrasting way to look at
Next slide. So what we're going to do now just to sort of get the discussion going is to go through the criteria in table 3 and review recommended characteristics for PROs as they could be useful in performance measures. And we're going to pick on the Western Ontario and McMaster Universities' osteoarthritis index, or WOMAC.

Many of you are I'm sure familiar with that. It's one of the instruments that's been around for a while. It's been used with knee and hip osteoarthritis. It's got 24 questions covering the range of the last one to 14 days, Likert-type with 5 discrete choices and a 100-millimeter visual analogue format is available as well.

There are three sub-scales: one on pain, one on disability, and one on joint stiffness.
conceptual and measurement model. In other words, the documentation of an instrument that we ought to be looking for should include a description of the concepts that are being measured and the intended populations for which they would be used and how those concepts organize into a measurement model.

The target of what you are measuring, the PRO you are measuring, should be a high priority for the system or it's probably not going to be sustainable.

And so if you look at WOMAC as an example, actually, it does not do so well there. It is more one of these long-term, has-existed-for-a-long-time instruments where the factorial validity has not held up, has not been deemed to be adequate from a conceptual and measurement model standpoint.

But its grades get better as we move on. So let's go to the next slide.

Reliability. The internal consistency should be at least .7 or above for group-level
comparisons or use and .9 or above for individual purposes. And also that is the internal consistency and also the other aspect of reliability. The more commonly referred to one is stability or reproducibility.

It often depends upon the time window, shorter, of course, being better because people don't change. You want to be testing people at two points in time where they don't change to show that something is reliable as a rule or applied to the same line twice.

So the Cronbach's alpha for the sub-scales are quite good on the WOMAC, from .86 to .98. So that's great. And the stability has been adequate for pain and physical function but less so for the stiffness sub-scale if you look at the literature.

Next slide. Next is validity, three main types: content, construct, and criterion validity. A limited number of
instruments have been validated and with the validity of these sorts in the context performance measurement.

The WOMAC development involved expert clinician input and survey input from patients as well as a review of existing measures. So it's got good content validity. And the scores correlate well with satisfaction with arthroplasty and other clinical outcomes. So the validity looks pretty good on the WOMAC.

Next slide. Responsiveness. So it's an aspect of validity. You want to be sure that what you are using is going to pick up important change or be responsive to it. It is important in performance measurement, obviously, because there is an expectation that you will detect change or hopefully improvement or there may be consequences associated with it.

And if you are going to expect to tie action to scores, then you do want to be
able to be sure that the scores will change as a result of action. So if you're going to have actionable assessment, then you want to be able to make sure the assessment actually is responsive.

And the WOMAC has adequate responsiveness and ability to take change in response to focused clinical interventions.

Okay. Next slide.

Interpretability, which really comes out of experience with the measure and applying it in context. You need to know what a low and a high score represent; what is the average; and what is the standard deviation; what is the referenced population that you are pulling the average and standard deviation from; and what is a meaningful difference; and sometimes, you know, if it's know, what's the lowest likely meaningful difference, or a so-called minimally important difference; and how do you know when a change over time is meaningful in a person; how do you know when a change over
time is meaningful in a group, a group of people.

We talked about this earlier. For performance measures, if different PROs are used for the same concept, it's going to be important to have a link or a crosswalk that allows you to be able to interpret scores on one instrument in relation to what they would be on another or how they might relate to that same underlying trait or thing that is being measured, like depression or physical function and to apply the criteria that you learn and use for determining what is clinically meaningful in this setting.

Now, the WOMAC does have population-based age and gender norms. They do have a published minimally clinically important improvement that I use. And you can use the WOMAC to translate into utilities for economic evaluation. So because it has been around a long time and been used a lot, it does have good interpretability behind it.
Burden is the next item. And that relates to the time and the effort and other demands on the respondent as well as on the administrator, on the clinical staff.

Performance measures, PRO assessments then probably in this context do need to be as brief as possible. And reporting should be done in real time.

There is a short form available for the WOMAC. And the average time to complete it is a little under five minutes. So it seems to do well in that criterion as well.

Next slide. Alternatives modes and methods of administration. We walked earlier. Several people have commented that it is important to be able to be as flexible as possible without being blind to the issues that one needs to consider.

There are validated mobile phone and touch-screen platforms for WOMAC. So it does seem to have a good means of alternative
modes available.

Next slide. And cultural and language adaptations, again, hopefully done in a way that is responsive to getting truly, not just literally equivalent translations but semantically equivalent translations on culturally fair, if not equivalent, translations to be able to use across people of different cultures and languages.

The WOMAC is available in dozens of languages. So that's also a good class, a passing grade.

Next slide. This is the last one, which is in the context of electronic health records, critical features of performance. Measures will be the interoperability of the measure, having automated real-time measurement and reporting capabilities. You can get that instant report and not be in the way of the clinic flow; instead, perhaps even facilitate it and preferably sophisticated or at least stay to our analytic capabilities.
with the information that is brought into the EHR.

Electronic data capture of WOMAC, which is available, should allow for fairly easy integration within the health record. So I'll pass on that one as well.

Go to the next slide. That may be the last one. I thought it would be useful just to kind of run through. That's the table, table 3. Maybe we should have a tenth criterion of actionable from the discussion and maybe have some discussion about that.

And my thanks to the developers and validators of WOMAC for allowing me to use them as a guinea pig for this introduction.

Thanks.

DR. WU: Thanks, David. And thanks for keeping us on time.

So who -- Jim, are you -- it's time. So Jim is going to talk to us about a couple of things, characteristics that identify PROs as suitable for use as
performance measures, relevance for use of

PROs that are developed for controlled

research to be used in clinical practice. And

who knows? Maybe more.

MEMBER BELLOWS: Thank you.

So the space of curmudgeonly

having been taken, I'm going to venture into

slightly provocative and talk a little bit

about what I see as the real importance of

useability, including actionability, which you

just mentioned, and meaningfulness in this

space. And perhaps they're even greater

importance than some of the other technical

aspects we talked about.

So when I look forward from where

we are now, I see two visions of how the whole

PRO enterprise might develop. And, on the one

hand, I see a vision much like what they've

done at Dartmouth, which to me is a thing of

such incredible beauty that I almost wish I

could get back pain and go to New Hampshire

and experience the Dartmouth model with all
the integration of the clinical care that we have seen.

I know that in Kaiser Permanente, we have a couple of uses that are working out terrifically, the use of PHQ-9 to drive a treatment model for depression, which is really based on treatment to remission, instead of counting how many medications we are doing or how many follow-up visits we are having, really has the potential and is transforming our depression care.

We're working with an instrument we use after total joint replacement that is being terrific in helping us identify the people who need the most follow-up. So that's the part of the vision I really love.

I see another part of the PRO transformation off into the future that could look not nearly so nice and that could look more like patients feeling spammed basically by their clinical systems.

Here are some of the evidence
points I have that bring that vision to bear.

One is that I know that in all of the surveys
that we do in our system, our customer, our
patient experience and care experience
surveys, our responsive rates are trending
notably downward over the last few years on
CAHPS and HCAHPS and those satisfaction
surveys.

I know that our responses to
clinical reminders, whether it's they'll get
your colonoscopy or breast mammogram or
whatever, our responses to those are going
down over the years. And we're going more
quickly through each new technology.

When we first knew how to use our
computers to send out letters on paper, it was
like a great thing. And people responded.
And then they stopped. And then we started
using IVR. Oh, that's so cool. People
respond to IVR. And then they stopped. And
then it was email. And then it was text
messaging. And with each one, there's a
shorter and shorter message.

(Laughter.)

MEMBER BELLOWS: And so my fear is that if we're not judicious about how and where we use patient-reported outcomes, the patients will start to feel spammed by us asking all of these questions in much the same way they do in other things.

The business of patient-reported outcomes is so important and so precious to us I think in a clinical context that we need to be careful and judicious about how we proceed so that we don't push patients over their limits and get them to the point where they stop responding.

I know for myself, I am at the point where when online surveys pop up in the middle of an internet experience, I just dismiss, dismiss without even thinking about it and similarly. And I don't want to get there with patient-reported outcomes.

So how do we avoid that? To me
one way to avoid that is that I think the traditional formula in a room of moving toward performance measure would be to consider the reliability and validity of measures first and only secondarily their usability and feasibility.

And I would propose that for the patient-reported outcome space, that we make the useability of measures; in particular, their meaningfulness and their actionability by clinicians and patients, of paramount importance and not even consider a use for performance measure until it has been demonstrated that it can be used productively in a clinical system, that it can be appreciated and acted upon appropriately at scale and in real time and only then move on and that if we get ahead of ourselves and start pushing out a plethora of measures that are coming not from the position of improvement but of accountability and payment, that we will end up in a space that is not
ever what we intended.

So I think that is the basis of my idea. So I guess there were a couple of specific questions. From the research context, most of the measurements we have that have the greatest validity have the most number of items on the scale, duh. They're most valid if they are really big patient populations. But it's those measures that have a lot of items on a scale and go to a really big population that are going to most saturate our patients' willingness to respond.

So to me moving judiciously means going to narrowly targeted patient populations that have a specific episode of care where the care is very salient to their needs. That's like the people who have had a total joint replacement or CABG or whatever and are very interested in what is going on with that care and then, departing from the practice that has been so common in research but going with extremely short instruments that everybody can
I understand.

So it's a little bit different than the typical criteria, but that's what I think will serve us best in the long run as we develop and build.

Thank you.

DR. WU: Thanks. That was very timely. I remember a time when I used to answer my telephone at home.

(Laughter.)

DR. WU: But those days are long gone. And so I think that this patients feeling spammed notion is one that is going to stick with me.

Next we've got Gene Nelson again. Gene, are you ready to go?

MEMBER NELSON: Ready to go, right. I am not as good at being curmudgeonly. So I won't try.

The first question was, when can general health status measures be utilized? And when should condition-specific measures be
utilized?

And in general, I think at this point in time, it's wise to use both general and condition-specific measures when you are focusing on a particular clinical population so that we can start to learn what are the benefits and drawbacks of both. So conditions such as spine problems and heart failure and total joint replacement, depression would be good candidates for both general and condition-specific.

I think it's very important to use general health status measures under some conditions. So, for example, at Dartmouth Hitchcock 76 percent of heart failure patients have two or more co-morbid conditions. The median is four. So to look at the whole patient, it makes sense to use a broader general health status measure that captures physical health and mental health and function.

Also, when screening for problems
that may be important but can easily go undetected, go to use a general measure, so, for example, people with open heart surgery or AMI recovering or postpartum, it's easy to miss mental health issues or in the new annual wellness visit that pays for prevention that Medicare is offering for the first time, they're screening for both functional problems and health risk problems. And so using general and specific measures is helpful, specifically there a depression measure as well as a general functioning measure as well as health risk indicators based on health behaviors.

The second question is, are there any setting-specific issues for selection of PROs? Short answer, yes, many. First, what are the clinical populations that are being cared for there that matters to them? What are their needs? So, really, knowing your patients and then tuning the use of PRO measurement to the common likely issues for
the patients and then tuning the measures to
the work flow, as mentioned earlier, to try to
create a richer and more useful information
environment for both the patients and the
clinical team. So it's easy to say but design
in and design it well so you get the
useability is not like falling off a log.

There are also issues of patients'
acuity and cognitive abilities as well, of
course, that does vary pretty dramatically
from setting to setting or point of time
within settings.

The third question was, what
conditions would be most sensitive to
measuring changes in patient health status and
outcomes? So here chronic conditions, such as
mentioned earlier, heart failure, depression,
ischemic heart disease, Parkinson's disease,
low back pain, osteoarthritis, rheumatoid
arthritis. All of these tend to have an
impact that rolls through the person's
well-being, physical, mental, and roll
impairments, all fairly likely.

So certain chronic conditions would potentially be good starting points and, as mentioned earlier, by any other surgical conditions. Total joint arthroplasty, bariatric surgery might be a really good place, open heart surgery, spine surgery.

The third one is a little bit different in terms of most sensitive to measuring changes in outcomes. And that's people that are not necessarily in serious trouble yet, but they're at high risk of avoidable death or of the onset of disease or the accumulation of new diseases.

So health risk status measurement, including health behaviors, and biometrics for people, such as cardiometabolic syndrome, or minor hypertension but at risk for other conditions or high health risk behaviors.

So screening for avoidable risk of death or onset of disability would be an area where risk measures could be sensitive. We
estimate that about 5 percent of the population aged 30 to 49 account for about 25 percent of the risk of avoidable death. So there's a lot potentially there.

We haven't talked very much about health behavior measures and health risk today, but I think if we wish to go upstream to promote health and to prevent disease and disability, very fruitful area.

Last question was, what is the variation in patient-level scores related to clinical interventions? I put together rather eclectically one slide here that shows for some general health status measures, -- this is from the sport trial -- the differences in people having surgery or not having surgery on a zero to 100 scale of physical function.

And so a 44 versus 30, 23 versus 32, for example, herniated disc and stenosis, depression, the PHQ-9, as many people recognize, that a certain degree of elevation is indicated.
And then getting to remission means less than five, a score of less than five. And some health systems and some providers are much more successful than others doing that. Within our Dartmouth Hitchcock system, we have tremendous variation at the clinical pod level on this measure. Carpal tunnel syndrome using a disease-specific measure, 3.1 down to 1.8, by people that are looking at patients with the problem suggest very substantial improvements, et cetera.

So a lot of the work that has been done on clinical populations using either generic or condition-specific measures and associated with different interventions and different kinds of health care can show very substantial differences that would make a large difference in how a person is doing and feeling physically, mentally, and in their role.

DR. WU: That was great. Thanks.

I like this last slide.
I do wonder, for screening purposes, a lot of the conditions that people have might sort of result in some general depression, some haloed depression of everything, but you would only suspect that something is going on but not know what it is.

And I wonder if sort of something which is generic but very comprehensive might be a better screening test, I mean, something like the old SCL-90, which went through a whole laundry list or review of systems of systems. I wonder if that is something that would be better for that purpose.

But shall we go on to Kalahn?

DR. TAYLOR-CLARK: Sure. Thank you.

I have to say that much of this conversation has brought me back to nearly 15 years ago. I have a Ph.D. in health policy and health economics. And the first day of the program, we walked in. And the professor said, "So, you know, we're going to calculate
these qualities."

    And I said, "Well, what if a quality falls below zero?"

    And the professor looked at me really strangely and said, "What do you mean?"

    I said, "Well, what if somebody is living below a zero and actually living worse than death?" because zero was death. And that was when I realized I needed to be part of the patient community and the consumer community to talk through these issues a little bit better because I felt that the economics weren't exactly measuring what I was hoping to measure through the patient perspective.

    So that brings me here today. And I appreciate the opportunity to speak on the consumer side.

    I was asked to actually address only one specific question, which was what measures are important to patients. You know, I kind of went through this a lot, especially as we were sitting there and having the
discussion at our table about what would really be important to patients but what is also useable and feasible for the system to do so. I'm always thinking about that balanced perspective in providing this perspective to you.

The first thing I will say is that obviously -- and this has been brought up many times today -- that the measures that have been put forth so far through CAHPS and others have been really useful for research purposes but not necessarily as meaningful to patients. That is something that has been brought up, but that needs to be reiterated as we think through what we are going to use for measures of both quality improvement as well as accountability in public reporting.

That is going to be extraordinarily important when we think about how patients actually read what this measure means and then are able to use that for decisions on terms of their preferences or in
terms of their decisions about the providers and the treatments that they receive.

So I want to put that forward because a lot of the measures that we think about might not even necessarily exist in many of these validated instruments. I hate to say that because I hate to take two steps backward. But I do think that there are ways that we can think through this.

So the measures that we have really been thinking through are some of the ones that actually were mentioned. So expensive surgical procedures certainly are of import.

Shared decision-making, particularly in maternity care, these are measures that are both of import, obviously, to purchasers but also very much to consumers. And really thinking through some of those measures will be important measures of depression, as Gene mentioned.

But I wanted to also say that the
measures of patient experience shouldn't completely be excluded. I think we have been mostly talking about patient-reported outcomes today as levels of functional status and some of these other types of measures. But I want to reiterate that patient experience is extraordinarily important to patients.

I will give you an example. In my former life, I was a research director at the Brookings Institution, where we were looking at a pilot study where we worked with a major health system that was now going to become an accountable care organization.

One of the points that they made, they said, you know, we really need better measures of patient experience because what we are finding is that some of the experiences our patients are having are actually having an effect on our retention.

Let me just take you just through a moment there. They said specifically care coordination measures from the patients'
perspective as well as organizational access, which is usually defined by CAHPS as their ability to get after-hours care and things like that, those two particular measures were the most defining in actually predicting the patients' retention, the retention of their patients in the system, that they really wanted to actually define care coordination a little bit better than what CAHPS had been doing previously but to really actually include that in all of their measures, both for quality improvement as well as for accountability in public reporting. Those are just two examples of experience where I recognize that they have tended to be process measures but can lead specifically to outcomes of import to patients.

The other piece that I would say is around the confidence piece. And certainly patients' confidence in their ability to make decisions in terms of self-management and care, either from the patient perspective or
the proxy perspective, are going to be extraordinarily important.

And then the final piece that I would say around the measures specifically is around this idea of access to social and community resources. I want to take a step back on that one. Ted Rooney suggested for Maine -- got up and actually made a real interesting point. He said, "Well, we can't always put the onus on the provider." And I think others have talked a lot about this, that, you know, we're defining these measures and then we're suggesting that the health care system is the only fix-it.

The bottom line is that that is obviously not true. Social determinants of health, your ability to get transportation, your ability to actually access the system, et cetera, et cetera, are going to play an extraordinary role in your ability to get better.

And so one of the points that we
were actually trying to make again when I was working at Brookings was to think about what a measure could look like in the health care system that would allow us to get as well as we could on what the system could do to actually provide access to those other resources.

And we actually came up with a potential measure that said, how well does your provider or did your clinician -- sorry. I think the specific measure was, did the provider link you to helpful and useful social and community resources that helped you to either self-manage or take care of your diabetes or your asthma? And this gets to one of our points about who we are actually looking at.

These were very narrow disease-focused, chronic disease-focused, surveys. And I think I would echo the point that Jim makes that we do want to have sort of general questions around general satisfaction
and experience and outcomes, but we also might want to look specifically at conditions that are of import to these particular groups.

Obviously care coordination is going to be important for certain groups and not necessarily for others, the same with the social and community resources. This is going to be extremely important to patients that have the need to get to these social and community resources.

So as we think through what the measures should look like, I don't want to take patient experience off the table. I also want to try to think a little bit outside of the box about how we can actually link other measures that are going to be of import to the use, to patients' use, of the data.

And I will end there and take questions later. Thank you.

DR. WU: Well, that was terrific. Thanks very much. And in sort of the out-of-the-box spirit, these could possibly
fall into sort of the health behavior bucket, but they really are things like confidence, self-efficacy, outcome expectancy, which, in fact, health systems should be in many ways accountable for improving in patients.

I know the Society for Behavioral Medicine has been thinking about this a little bit. And maybe it is worth revisiting a little bit more.

Finally, Ken, can you clean up?

DR. OTTENBACHER: Thank you, Al. I am going to go ahead and use the podium, I think. It might make it a little bit easier for you all to hear me. I would like to begin by thanking NQF and the conference organizers for the opportunity to participate. Okay. Didn't work, huh?

(Laughter.)

DR. OTTENBACHER: Okay. Let me try it again or I could sit down. Again, I would like to thank the NQF and the conference
organizers for the opportunity to participate
in the workshop.

My task today is to comment on
issues associated with selecting patient-level
PROs. Specifically I have been asked to
respond to the impact of patient
characteristics and baseline values on change
in PROs and on conditions or circumstances
under which stabilization or no change may be
a desired outcome.

Measuring change requires a
context. Targets for assessing clinical
change include the individual or patient, a
group or facility, and the population.

Defining the context is an
important first step in assessing change.
Approaches, methods, and even conceptual
frameworks may differ from one context to
another.

Assessing change has a long and
controversial history. A common approach has
been to calculate the absolute difference
between change from baseline to follow-up. This is referred to as the change, gain, or different score, a variation of this approach is to compute the percent of change.

The limitations of these approaches are widely known. Ward and others demonstrated in the 1960s -- and I quote here -- that differences between scores tend to be much more unreliable than the scores themselves. This quote is from a widely referenced text, "Problems in Measuring Change," published in 1963.

Percent change continues to be a widely used measure, despite its limitations. Percent of change is sensitive to variations in baseline distribution. If the range of baseline values is large, the variance associated with the percent of change increases disproportionately and the sensitivity of this approach is reduced dramatically.

An extension of change or gain
scores is to adjust for baseline using analysis of covariants or regression models. Sophisticated statistical approaches have been proposed to adjust for baseline scores and other covariants using general estimating equations or hierarchical opinion modeling. These procedures examine interactions and relationships among baseline score, time, other covariants, and intended outcome. While they represent powerful methods, they are not without controversy. An article in the American Journal of Epidemiology in 2005 by Glymour and colleagues provides caveats regarding adjusting for baseline values using sophisticated statistical models. A variety of indices have been proposed to address some of the limitations of traditional approaches. These include reliability change index, the residual change score, and various applications of effect-size measures. Each of these approaches has
strengths and limitations that have been described in the literature and are referred to in the draft NQF methodology paper.

The approaches described above can be classified as distribution-based, meaning that change or gain scores are expressed in terms of an underlying distribution and rely on sample variation, standard error, and measurement precision.

An alternative is anchor-based approaches. Anchor-based approaches require an external or independent standard, or anchor, to determine the clinical importance of change. The anchor-based approach provides unique opportunities for stakeholder involvement in determining the focus and magnitude of change. It introduced other measurement challenges.

Another alternative is referred to as adaptive measurement of individual change. Adaptive measurement combines the methods of item response theory and computerized adaptive
testing to provide a different way of thinking about a scale's ability to detect change. The adaptive change approach has a number of advantages, including the ability to demonstrate when a measure includes better precision across the range of a trait.

Combining IRT and computer-adapted testing was described previously and is being examined in several applications of patient-reported outcomes.

There are instances where stability or the ability to maintain a level of performance is the desired outcome; for example, in degenerative conditions, such as Parkinson's disease. This is also true for conditions that involve cycling diseases, disease processes, such as multiple sclerosis.

For some populations, the best outcome may be to maintain a desirable level of function; for example, maintaining a level of mobility or self-care that allows independent function in older adults.
Providing evidence of stability may be analytically less complex than assessing change that requires adequate data points, which can be a logistic or resource challenge.

I will conclude with a couple of observations regarding other challenges relevant to assessing change or no change in PROs. One important challenge is how to risk-adjust. Quantitative methods for risk adjustment in managing various forms of bias have advanced dramatically in recent years.

The challenges of risk adjustment will be particularly important as patient-reported outcomes at the individual level are translated to performance measures and quality indicators at the facility and population level.

A second challenge is how to facilitate the evolution of patient-reported outcomes to include patient-centered outcomes. The Affordable Care Act and the creation of
the Patient-Centered Outcome Research Institute have placed increased emphasis on the role of stakeholders and consumers, not just in the assessment of outcomes but as partners in the decision-making process regarding the content of what should be measured.

A final challenge is how to monitor, identify, and reduce unintended consequences. We are all aware of examples of unintended consequences resulting from changes made in health care that were implemented with the best intentions.

We don't want to create a patient-reported outcome or quality indicator that becomes the kudzu of health care.

(Laughter.)

DR. OTTENBACHER: My major advisor will respond to telling his Ph.D. students that doing outcomes research is a lot like raising children. You always think you're going to do a better job next time.
(Laughter.)

DR. OTTENBACHER: The careful and thoughtful assessment of patient-reported outcomes, as reflected in this workshop, and the high-quality scholarship evident in the draft of the PRO methodology paper indicate that we really will be able to do a better job next time in creating quality indicators that reflect patient-reported outcomes.

Thank you.

DR. WU: Thank you. I will try to unravel myself from the kudzu to say that, interestingly, some of the problems that plague individual-level change, measurement of individual change, are less of a problem for measuring group change.

So many of you with an epidemiologic bent may have more insight into this than I, but for some of the tasks for quality measurement and accountability, we actually may still be able to look at change scores in one form or other.
Anyway, great discussions. And every kept time sufficiently so that we have I think -- now, do we have 20 minutes or do we have a little bit more? Did we start a little late? Great.

So first comments from the panel? Oh. Yes, please?

MEMBER CELLA: I don't know if I could just say something about people who are gathering their thoughts. My experience of Dr. Ottenbacher's brief presentation was that it was really a marvelous overview of a lot of work that has been done and that we have learned about measuring change. So congratulations. Very well done in this short period of time.

It got me thinking about something that I think could work very well in this process that could bring also in this sort of patient engagement side as well in a group like this.

To do a technique that really
draws from educational testing and looking at achieving standards. And that's bringing together experts. In this case it could be experts like people here in this room along with patients that represent the patient perspective in the community, showing them data.

They are blind to scores. They don't see numbers. They just see profiles. They see people. Here's a person who's got this much pain. Here's how they're functioning.

And then you go through -- I won't go into details of the methodology, but there are various methods that allow groups to form a consensus opinion about what kind of difference would matter. And you can also put in extra methodology to get a value for how much would that be worth.

The essential thing would be to bring experts, people like you and patients in the community, together to come to consensus
on units of change that would matter. And you also articulate why they would matter. And then these could become standards that get set for achieving either maintenance or improvement.

DR. WU: Ethan?

MEMBER BASCH: I can comment on some of our research within the oncology world around symptom-specific or demand-specific or population-specific measurement versus more generic. And in general, adherence rates over time are much higher with very content-specific measures, measures that are more common when there are fewer zero values. It is actually a very strong relationship between how specific we are to the world of the patient and the patient interest and willingness to report.

Similarly, interestingly, there is also improved adherence when the provider is involved in the process of collecting the data when, for example, a letter is sent to the
patient saying, you know, "Dr." so and so or
"Nurse" so and so "is very interested in your
participation in this information." In
general getting buy-in from providers to do
that is much greater, again, with
context-specific measures.

So it seems the closer we get to
the context of use or the fitness for purpose,
the more everybody kind of perks up and feels
that this is relevant. And they become more
enthusiastic about participating.

DR. WU: More comments from the
panel? John?

MEMBER WASSON: Yes?

(Laughter.)

MEMBER WASSON: Hello, Mother. I
saw a moment of silence there. I didn't want
to let it lie fallow.

In particular, in regard to Jim
Bellows' points, I just wanted to go on the
record as saying he is right on but not far
enough. By that, I mean we are all operating
in a current paradigm mindset that is good Twentieth Century thought process. But the median is becoming the measure.

And we have to recognize all of us use cell phones, for example, and we're still talking about portals for EMRs. And we're wondering why patients don't respond to new methodologies. And the answer is they're not in control. You don't want to answer surveys that you don't control nowadays. You want to control. We have to deal with that reality.

Just to echo the point that was just behind me, in our experience, we have been doing this a long time in primary care and other practices. When the clinician should say, "This is part of care. This is part of the service to you," the response rates one time around, even on the internet, are well over 80 percent, not this 30 percent cap stuff that goes out. That makes life so much easier for analytics, et cetera.

While you are serving the patient
needs, you can put in a few measures that matter. And each patient needn't have the response burden of the entire instrument because you can aggregate it over many patients to get at the accountable care organization.

So, in other words, all I am saying is just what Jim said, that the useability and the service to the patient, put that first. And then we can fold in measures for accountability quite easily. And we'll have high response rates and get away from a lot of the issues that Ken just articulated. You know, how do you deal with risk adjustment, et cetera, et cetera?

So I just think there are practical approaches. I know there are. And I think Jim's right on the money. And I worry that if we stay too locked into the old style paradigm that is still in the back of our mind of mail-out surveys, you have to complete all or most of the items yourself, we're in
But if we look at it from a population point of view, where individuals in my practice get three items and another individual randomly gets three items, you can still aggregate it up to my practice.

And, then, finally, you can crosswalk the PRO measures across practices very easily. There are just so many technical solutions to a lot of the issues we're talking about.

A Cronbach's alpha stands in our way. All right? A Cronbach's alpha stands in our way because it doesn't allow that thought process.

DR. WU: So the ghost of Alvan Feinstein is smiling, I think.

MEMBER WASSON: Absolutely. He would call this quantiphrenia.

DR. WU: Yes. On the other hand, I think that, even though some young people are thinking nostalgically about remember when
we used to have PCs.

I think that patient portals don't necessarily have to be the enemy. They can be friendly. They can also be ported, incidentally, to your cell phone. But in many ways, if they designed properly, they can provide the patient with a lot of value. They can help them schedule appointments; cancel appointments; look at lab tests; communicate directly with the physician in a confidential way; and, incidentally, fill out PROs. So I am not sure that that is entirely dead.

MEMBER WASSON: Knowing they are all intended.

DR. WU: Not yet. There. How's that?

MEMBER LARSEN: I am Kevin Larsen. I am from ONC. A little perspective from where I came from before, which is Minnesota. And we have been doing essentially patient-reported outcomes for about five years as part of structured programs called
Minnesota Community Measurement.

You know, in my own clinical practice as a primary care doctor, we were doing asthma control tests and a PHQ-9 test for five years as a statewide measurement program.

My practice was in a country hospital, which 20 years ago had thrown away all patient forms because the patients never filled them out. Our health fiduciary rates were low. We didn't get any kind of response to anything on paper because of reading levels and language levels.

So we had a system-wide strategy where we actually administered these verbally to patients across our home care organization. And it turns out that actually happened quite a bit within the State of Minnesota.

In my own practice, I actually was an early adopter because I was a medical director. And so I had to make sure that I was doing things that had to happen.
So I actually administered these myself for three years, asked all my depression patients their own PHQ-9 scores. What I found is I got 100 percent response rate. And mostly it was because we used this as a tool to jointly decide about patient care. So it wasn't just for the patient, and it wasn't just for the doctor. It was used right in the point of care. I would show them a graph on the computer, "Here's how you were last time. Here's how you are now. I think your depression is worse. Let's think about escalating your therapy."

And, all of a sudden, this tool that we think of as an accountability tool became the visibility tool for the patient and the doctor to really think through what should they do about care right now.

And to my mind, that's the most exciting news of patient-reported outcomes because it really moves health forward. It
doesn't just beat people up for not doing something that we think they should do.

DR. WU: That is really what we're talking about, about getting things into the work flow for both patients and providers.

More comments? Yes? I see a hand. It's attached to Ted perhaps.

MEMBER GANIATS: It's Dr. Ted. I am sorry. I was being so good. Then Dr. Wasson had to speak up.

Everything that is being said is very good, though it's contradictory. And Dr. Taylor-Clark brought up some excellent points. And Dr. Wasson brings up some excellent points. And this guy -- I can't read his name tag -- brings up some excellent points. The problem is that they all contradict each other a little bit.

If what we want -- what do we want the patient-reported outcomes to be? Do we want them to be clinically actionable, important, real time? Then that is going to
decrease the usability of an institution to be able to increase retention rates.

Do we want it to be part of a pay for performance to try to improve quality over time? That may or may not be as actionable at the patient bedside? Do we want to help employers choose a health plan that might require something different?

So I think all of the points are excellent. I just think that we have to keep reminding ourselves that sometimes we're talking about different types of PROs. And not all of them are going to be able to be used for all of the purposes.

Gosh, as a clinician, I would love to have everything be useful all the time, but I just think we have to be really clear about what is going to be used for quality improvement, what is going to be used in clinical practice, and what is going to be used for the other purposes.

DR. TAYLOR-CLARK: I actually
don't think that what you are saying is that
we were contradictory. I actually think that
you made something much more clear in what we
said, which is that as we're thinking about
criteria -- I know we're not using the word
"criteria," but as we're thinking about how
we're going to choose these measures and what
we're going to actually end up using them for,
we have to make the distinction of whether
we're going to be able to use them for
purposes of accountability or useability or
whether we're going to be able to use them in
the clinical improvement process or for
self-management.

I loved the example that you just
gave about using the outcomes for shared
decision-making, which is ultimately what you
did. And so there are different ways that
these measures are obviously going to be used,
but I think that that has to be a part and
parcel of our thinking around how we're going
to develop and design the criteria for these
things.

DR. WU: Great. Shall we open
this to the back of the room? And also,
Operator, can you please open the phone lines?
We're moving on to the back of the room.

MR. CUNNINGHAM: We will pause for
the phone lines. Operator, if there is anyone
in the queue?

DR. WU: Operator?

MR. CUNNINGHAM: Are you still
with us? Farrah?

DR. WU: Hello?

THE OPERATOR: Yes. If you would
like to ask a question, please press *1 on
your telephone keypad.

MR. CUNNINGHAM: Thank you.

DR. WU: If anyone in the back of
the room wants to use their phone to ask their
question, that would be fine.

(Laughter.)

MR. CUNNINGHAM: I am not hearing
anything on the phone.
DR. WU: One size does not fit all. Any questions in the back? Comments? Lew? Lew, could you please move to the back to ask your question?

MEMBER KAZIS: Hi. I had two questions. One was for Dr. Taylor-Clark. I really liked your comments, especially the ones related to thinking out of the box.

RWJ has had a real investment in looking at community health and the issues around assessment tools that, in fact, are supposedly gauging the physician when, in fact, there are much broader issues as the community. There are issues of access, something as basic as, does the ambulance, in fact, come within ten minutes of an MI?

The question I have is, where do you see those types of metrics moving, in what direction; and how, in fact, can one begin to think in terms of operationalizing them in the context of our health care system?

DR. TAYLOR-CLARK: I think that is
a really tough question. And certainly those
that have been part of this what we call the
social determinants of the health world have
had this as a problem, you know.

And so the way that I would
actually see the measures as not necessarily
at this point, as Jim described, not
necessarily being used for purposes of
accountability or payment for the provider
but, one, as a monitor for understanding how
especially new delivery systems, integrated
systems, are actually working to improve the
health of their patients, not just the health
care processes of their patients and certainly
outcomes are getting us to that.

So one would be for monitoring,
but the second is, as we think about
accountability in terms of public reporting --
and Al and I had this question. I said,
"Well, what is accountability?" You know, I
have to go back to the basics on everything.

And he said, "Well, I think it's
actually, really, about public reporting."

Well, that's going to be an extraordinarily important place for these types of measures to be for patients to be able to make decisions about, if they can, make decisions about, where they get treated and how they get treated.

But it will also be a check on the system, not necessarily on the individual providers but on the system, to be able to say, "We need to be able to provide resources."

And my example of linking to social and community resources is one of potentially many that we could consider and I think that we should consider. But I think it is going to put the onus on the system to really think through how they engage with other systems.

And I think that is where we want to go. That is where I would like to see us go. I don't want to stick just with health
care. And I don't think that this is a revolutionary concept, but I think we spend a lot of time talking about health care systems and structures and we put the onus on providers. And ultimately this is about improving health broadly.

And so in order to do that, we're going to have to start developing the linkages and the relationships with other types of systems so that we can improve health.

And I think that in order to do that, we have to start the monitoring somewhere.

DR. WU: Great. We've got something back. I see a waving microphone.

MS. BARANOWSKI: Thanks. My name is Rebecca Baranowski. I'm with the American Board of Internal Medicine.

I have to say, first of all, as someone who has a background in test development, it was really exciting to hear the discussion of CAT and DIF and IRT. It
took me back to a previous lifetime.

There was some discussion about the need to keep patient surveys short. And I have often heard that if patients consider the questions relevant, that length is not so much of an issue.

Can anyone comment on that?

DR. WU: Well, I will say from our own experience that that is true. If someone feels like you are -- we are doing a lot of service with oncology patients. And if we are asking them questions about their own health needs, the things that maybe gaps in their goals or achievements, they are happy to talk about those things all day long on paper or elsewhere.

On the other hand, they can be very annoyed by very short surveys that they don't think are worth their time.

Other comments?

MEMBER KALDENBERG: I think that's true. I think the respondent's interest in
completing the questionnaire or an interest in the topic of the questionnaire will lead them to answer more questions.

So if a patient or any respondent looks at a questionnaire and says, "I don't understand its purpose. I am not particularly interested in it," they're not going to respond; whereas, if, as Albert said, it is something that either has been demonstrated as important to you, it's a part of your treatment plan, it is important to you because "I'm interested in my health. You can ask about as many questions as you want." And I think there's a lot of literature that says that's true.

DR. WU: David?

MEMBER CELLA: Yes. You know, I guess I would like to give an answer that's in between what lies under your question and things that Drs. Bellows and Wasson were saying. So yes, patients like to be asked questions about how they're doing almost
universally and providers that aren't used to that and you see them in research studies. You know, you will hear patients say, "I think these questions should have been asked all along." And they're very happy about it for the most part.

But, coming back to what Dr. Bellows and Dr. Wasson were saying, there's a point at which they reach a limit. And sometimes that point comes by being asked the same questions over and over again, you know, every three months or for some period of time.

I've had the same experiences with seeing technology be exciting for a while. And then you lose it unless it has become really ingrained in the clinical setting.

So I completely agree and from experience as well, not as much as Dr. Wasson, but we really do have to make these relevant. But I don't think, unless I misunderstood what you said, Dr. Wasson -- well, you know, if I'm doing IRT, I can agree
that Cronbach's alpha is standing in our way, but I'm not sure if that's what you meant because Cronbach's alpha is in classical theory and you don't need it in IRT. In fact, it's irrelevant. And you can get very short, very relevant assessment.

So I think there is a solution here. And just the other part of my response to that is that it's not just the patient, of course. It's the provider's side. They have to continue to be enthusiastic about it because they're the ones paying for it to be done. So it's really a partnership in that respect.

DR. WU: I will add to that. You know, whenever you ask anyone to do things repeatedly, their interest gradually extinguishes. I think it's a physiologic response that we can't do anything about.

On the other hand, if you, we, respond to someone's complaint about their satisfaction with an element of their care and
tell them what you're doing about it or,

alternatively, if you respond to some bit of
feedback that they give you about the amount
of pain they're having and tell them what
they're doing about it, the next time they're
much more likely to continue to respond.

So I think that, again, what we do and
how we handle all of this is very important.

MEMBER NELSON: One of the cases I
referred to but didn't explain very much was
the rheumatoid arthritis registry in Sweden.
And there the survey that patients are
completing is actually fairly long. It takes
them about 20 minutes.

These people have a serious
disease: RA. And it used to be that this
information was collected for research only.
It was a registry. And now it's actually
collected so that the nurse or the doctor
looks onto the patient's current and past
results against the treatment plan so that the
next steps in the care plan can be responsive
to their disease activity and how they are doing.

And, by report, when this scientific registry was redesigned to actually help the patient, patients like their care much better. It absolutely changes the fundamental clinician-patient relationship. It makes it much better.

They've now added a care designer to help the 60 RA centers in Sweden do this well, embedded in the care design with a person with RA. It helps the new centers take it on.

And the trends over time show that people are starting RA sicker, but they end up at 12 months having much better disease scores. And they're doing better.

So I think it is possible under a focused context to actually have this information be extremely relevant to me -- it's all about me as a patient -- to the relationship and what that relationship is
doing for my care plan and actually seeing if my outcomes are getting better.

And under those conditions, how long it takes me to complete the survey, 10 minutes or 22 minutes, isn't relevant because now we've got an instrument panel that guides me and our care and how we're doing to monitor our care.

And that's becoming a model that the rest of Sweden wishes to adopt for all relevant registries. And Michael Porter is popularizing that in Sweden and now back in the U.S.

So I think there are new ways of doing this that can be rather -- what would the term be? -- disruptive.

DR. WU: I think that rings very true. We're coming to the end of our time. I'm going to ask one more time if there's anyone on the phone: Operator or anyone else.

THE OPERATOR: As a reminder --

DR. WU: Yes?
THE OPERATOR: As a reminder, if you would like to ask a question, please press *1 on your telephone keypad.

DR. WU: Wait. I think I'm getting a call from Kathy Lohr.

MEMBER LOHR: This is just one minor observation or question. In talking about shorter forms versus longer forms, at least for some ways of answering questionnaires, have people been trying to build in ways that you can stop and then go back?

Because I get this kind of questionnaire in various ways and various purposes, not so much for health care. And beyond about 10 or 15 minutes, you're either tired or you get interrupted. And if you can't save what you've done and then go back to it when you're either not so distracted or whatever, it's extremely annoying. And you might not the next time ever start knowing that you are likely to be interrupted and
can't save it and go back.

So I'm really just asking whether for people who are implementing a lot of these things, through portals or almost any of these mechanisms, do you build in ways for people to say where they have gotten and go back?

MEMBER BELLOWS: We don't have it yet, but our her vendor is delivering it in its next release very shortly.

You know, to me, this is partly coming back to the submerging theme that has come out so great, which is that it's all about context.

It's not just about the measure. It's about the measure used, how often, with how many patients, with what follow-up that's really sensitive to people's care and with what degree of control. And it's just built out so beautifully.

DR. WU: Great discussion. I think we are going to wrap up. I would like to thank the panel members and members of the
audience and, of course, people on the phone.

MS. PACE: So we are going to have
a short break. And we'll ask you to try to
reconvene at 3:20. And Kathy is clapping. So
thank you, panel.

(Applause.)

MS. PACE: So we will try to
reconvene at 3:20 if possible.

(Whereupon, the above-entitled
matter went off the record at 3:10 p.m. and
resumed at 3:27 p.m.)

CO-CHAIR BRENNAN: Good afternoon,
ladies and gentlemen. I am Patty Brennan. I
am co-chair of the panel today. I am not sure
everyone is as excited about this session as
I am, but I am very excited about this
session. And I think we will get some more
people excited about it shortly.

When we speak about
patient-reported outcomes, we often look to
efficient ways to both understand the process
of care that a person has been engaged with as
well as a place to store the assessment,

perhaps that assessment of their outcome

having dual use for their individual care

processes as well as the accountability and

quality improvement goals that the NQF has.

The electronic health record as we

know it today is probably not the best place

for it, but it's a starting point to think

about the electronic infrastructure necessary

to acquire patient-reported outcomes, organize

them in a way that helps us make meaningful

interpretations for the patients' progress as

well as the care and also have a way to begin

to aggregate for the institutions' awareness.

This afternoon we have a report

from the Commission paper. I'm happy today

David Cella is back again to speak with us.

And we'll have a reactor panel of three

individuals.

I am going to be introducing

everyone now to coordinate the time best.

It's just about 3:30. We are going to go
until about 20 until 5:00. So we'll have plenty of time for comments on this session. And then we'll do a wrap-up between about 20, a quarter of 5:00 and 5:00 o'clock.

But today in this next session, we're going to be talking about the key considerations for incorporating PROs into electronic health records. And let's think about those with small EHR, not large EHRs, because the electronic health record you know today is going to look very different within two or three months, in fact, if not two or three years.

David Cella will be giving us a report from the Commission paper. And then we will have three reactors' groups set up. Uma will be speaking with us about the experience at Cincinnati Children's Hospital. Kevin Larsen from the ONC will be discussing meaningful use and some of the indicators linking quality and measurement of outcomes. And then Ted Rooney from Maine will be
discussing the experiences in their outreach
group.

David?

MEMBER CELLA: Hello again. I am
happy to say I don't know if I am happy or you
will be happier to hear that this will be my
shortest tee-off. It's a par three session,
at least for the tee-off.

Okay. So let's try to broaden the
correlation and talk about the health or
health-related internet applications that
deliver a range of content, connectivity, and
clinical care, keeping in mind that most
e-health is driven by the electronic health
record that the provider offers to the
patient, although certainly not the only.

So what does that include? It
includes health information, individual and
group; online formularies, prescription
refills; -- I use it -- many of you probably
use it -- appointment scheduling and test
results; -- I've used it -- advanced care
planning and health care proxy designation, not used that yet, but it's available in some systems. And these applications tend to focus because they were built for them on the needs of the health care providers and the health care organizations.

And this has I think been a theme today that I didn't necessarily anticipate, but it is in the paper that there is also little evidence regarding whether these services offered are those that patients actually desire.

Next slide. So how about the side of integrating PROs into electronic health records and personal health records? PROs will likely constitute and important aspect of future stages of meaningful use.

Now, you could tell me I'm wrong and maybe I am wrong, but three years ago I went to a hospital administrator who will go nameless and said, "Hey, I think you really need to pay attention to PROs because you're
going to need to for meaningful use."

And he's a very nice guy. He said, "You know, I think we can kind of comply with meaningful use without worrying about PHO. So no thank you."

I don't think that's true anymore, but you tell me. And if it's in the future of meaningful use, then we ought to try to make it meaningful.

So critical features involve interoperability, as you know, widespread health information exchange, automated real-time quality and cost measurement, which the PROs can fit into, and sophisticated analytic capabilities that we certainly could engineer into the PRO site.

Next slide. So some important issues. First is the patients' perspective. Patients do want to be involved as partners in the flow of information, in health information exchange. They're certainly in their own personal health record.
You know, as we talked about in the last session, the clinical buy-in is really essential for good equality data over time especially. In order to get that buy-in, one important component that has to be compatible with clinical flow, I mentioned meaningful use.

And I suggested earlier eventually an issue of patient privacy, the actual physical transfer of patient-based PRO measures from the patient/the provider, there is a privacy, a possible, concern there. And then the electronic transfer of data or unauthorized access to patient-quoted data can cause privacy concerns or alerts.

So some key design principles in integrating PROs are to fit the measures into the flow of care, as I mentioned. The design, the system with stakeholder engagement -- it has got to be relevant and meaningful -- to merge the PRO data with other types of data so it's not a set-aside, stand-alone system,
although it is easier up front to build the stand-alone system, as we and others have, and to engage in continuous improvement of those systems based on user experience and technology, so sort of basic principles of getting it integrated; relevant; and, therefore, getting the buy-in and long-term sustainability.

Next slide. One of the things we developed with PROMIS, which I haven't mentioned yet today, is an assessment center, which is a web-based portal. I just mention that we and others have done these sort of set-aside systems.

So there is a web-based assessment system that anyone can use. You can go to NIHPROMIS.org. And you can take some CATs yourselves this evening when you have got nothing better to do other than watch the Olympics and see how your depression and fatigue levels are after this meeting.

So NIHPROMIS.org. And you can go
test yourself on a demo. But it's a
disconnected system. We don't ask you who you
are. You don't tell us who you are. You can
just try this out.

But if you wanted to use an
electronic health record environment, you
would need to link it to the electronic health
record. So, to that end, we've got SNOMED and
LOINC codes and have been working with
different her vendors to integrate assessment
center or the data capture and scoring
algorithms into the electronic health record.
So that is work that is ongoing and should
help maintain a truly sustainable integrated
system.

So I think that's the last slide.
Those are my introductory tee-off comments.
Thanks.

CO-CHAIR BRENNAN: Thanks very
much, David. Are there questions before David
sits down for clarification and more
information? We don't want him to go just
yet. Comments or questions specifically for David's section?

(No response.)

CO-CHAIR BRENAN: Okay. Thank you very much, David. I appreciate it.

I am going to ask Uma if you are ready to begin your remarks. Please?

DR. KOTAGAL: Good afternoon.

Thank you very much for inviting me. Before I make my comments about the PROs in the EHR, which is the topic we were asked to focus on, I would just want to give you some context.

So this will be in the context of a large academic health center. About 50 percent of our patients come from outside our region. A lot of our work is very specialized care. And all of it is pediatrics. So our patient-reported outcome conversation is in the context of that.

Secondly, put the focus in terms of our bias towards action and the fact that we are really interested about
patient-reported outcomes in the context of improvement, as opposed to thinking about it in the context of performance measures, so that our believe is that we will need to do a lot of work on practical applications of patient-reported outcomes and improvement before we could really look and say how this is going to fit into performance measures. So I just wanted to talk about that a little bit.

We have been engaged with patients and families for a long time, starting in 2002 with cystic fibrosis, families that taught us how to think about patient engagement. And all of our chronic disease teams have patients sitting at the table to help us think about what is important to them. And we accompany that with a fair amount of internal clinician engagement and training that enable us to look at patient-reported outcomes in the context of improvement.

Our work on patient-reported outcomes began with a five-year plan that asks
us to think about transforming care for children in chronic and complex disease for 60 conditions in the next 5 years. So we are looking at patient-reported outcomes in the context of clinical outcomes, patient-reported outcomes, and evidence-based processes, all of which we are looking to embed into our her with a combined reporting in the form of small multiples that allow us to look both at clinical outcomes and patient-reported outcomes at the same time.

A major interest, of course, is in scale. And so we are beginning to build our registries at this point primarily in Epic, in the content of Epic, and then using our reporting systems to be able to derive population-based data so that the individual patient data is within the system, the population-based data is sort of ad hoc or on top of it.

Our expectation is to build all components of Ed Wagner's chronic care model
into our Epic registry, which, of course, relates back to my earlier comments about the role of self-management and self-confidence and how the use of the patient-reported outcomes can adjust that balance between the provider and the patient, which, of course, remains a gradient at this point in time.

Our early lessons have told us that when we incorporate patient-reported outcomes into the registry and into the patient encounter, that provider behavior does change. But it starts with engaged clinicians. So our process does require us to have engaged clinicians.

For the 60 diseases, we currently have patient-reported outcomes in about 10. We'll have another five at the end of about ten months from now. And then we'll be building our way up to the next 40.

So the clinicians with patients choose the outcomes. And that's a really important consideration. They choose both the
clinical outcomes of interest and the
patient-reported outcomes of interest. And we
are not at this point requiring every chronic
disease team to have patient-reported
outcomes, but we are offering the invitation.
If they are interested, we would work with
them to do that.

So then our process really allows
us to engage with the clinicians, consult with
them, with the patients, decide what the
outcome of interest should be, go back to the
methodologists, and ask them the right
questions to figure out what is the right
measure, what is the evidence, what measure is
out there, what needs to be modified so we
come up with the right reported measure.

And all of this is at the
disease-based level. In addition to the
organization level, we are piloting with
measures of care coordination; measures of
patient experience that are not embedded or
related to the ER; and in some cases PHQL as
a broader generic measure of quality of life,
as opposed to a disease-specific measure; and,
finally, patient self-confidence in the
context of self-management.

So that's the context in which we work. The process -- and I want to just go to
my first slide here. I just used an example.
This is data from some time ago. These are children with obsessive compulsive disorders.
And I picked this one, as opposed to depression or functional outcomes or PROMIS scores, just because, even in the area of OCD in young children, we find that patient-reported outcomes are and can change the way it works.

This graph shows you data on four segments of patients that the clinicians self-classified. So we begin with a measure called the Symbyax, which is a gold standard of measure for patients, for OCD, but it's very lengthy, back to our comments about how dense the measurement can be.
We have modified it, proved that it is scientifically valid. And we began with paper and pencil, then with a self-standing kind of electronic system. And now it is incorporated into the electronic health record.

So the child comes in and fills a question, answering the question, "Did it get you or did you get it?" answering it in the context of burden of disease. This is then printed off.

Actually, the patient is tagged as sort of requiring a PRO measure at the clinic so that when they check in, the registration person gives them a confidential password as well as the information. We're currently doing it both using kiosks and tablets.

The patient fills in the data. The data is printed off. The patient has the data along with the family and the provider has the data.

So the conversation begins similar
to some of our discussions at the last break that say what is going on since the last time you see and what is happening.

And then this is obviously measured at each visit. The central concern here is how do we improve outcomes for patients. And when we do that, we find that the encounter gets to be very interesting. So it is for the individual patient that it matters, which we then roll up into a system or into a whole population level there.

Therapists find that they can stay focused on treatment goals. The patients and families find that they have a clear endpoint in mind so that they may say that when your symptom score gets to this level below 13, that's the end of our treatment. The care is more efficient because all of this information is available and the conversation can be more focused.

And, actually, what we have found is that by putting these evidence-based
practices into place, giving providers and clinicians a conversation to have with the her and the patient that we are able to detect these patterns.

So on the top right-hand side, you will see the response rates are pretty fast in some kids, high being not so good, low being better. On the left-hand side, you see a slightly different response rate. In the third graph on the bottom left, you'll see no changes in response. And the fourth is the exact pattern.

This was really done by looking at the data and segmenting it, sitting down with clinicians to do that. When the clinicians get together using this data to figure out what is going on, they concluded that their primary factor was actually a dose effect.

In other words, with children with OCD, the primary treatment is exposure to the compulsive disorder. And in the different groups, they were exposing them to different
dose effects, which they then through discussion were able to modify. And then on the next slide, you'll see I think the percent of children that have at least a 40 percent reduction in symptom scores. In this case, I think the number -- they have a target number that they use but find that they can make a big difference.

So we are using this idea of a kiosk or a tablet. We are not using portals at the present time for entering data, teeing it up to the clinician, a conversation by clinicians agreeing both on the measure and the target, and then a collaborative group that together decides how they are doing, why some people are doing better, why other people are not with a goal, then, for improvement in the functional measure.

So I'll stop with that and take any questions at the end.

CO-CHAIR BRENnan: Could you just speak a little bit about how this gets into
your electronic record now? Is there a
special section or does it go as a narrative
note?

DR. KOTAGAL: So currently it goes
into the electronic record as a pulled-up
document that shows the graph. The graph that
we get at Epic looks more like the one you see
on the last one, not like the previous one,
which is a little bit more interesting. But
it does get that way.

And the population-level data also
gets pulled out and is reported that the team
looks at on a regular basis.

CO-CHAIR BRENNAN: Thank you.

Are there clarification questions
you would like to ask Uma before we go on to
our next comments?

(No response.)

CO-CHAIR BRENNAN: Okay. Thank
you very much. Our next speaker will be Kevin
Larsen. Kevin joins us from the ONC.

DR. LARSEN: Hi. I am the Medical
Director of Meaningful Use for the ONC, a general internist by training and was doing that up until March, when I moved here and will hopefully be doing it again soon here in D.C.

I would start with a brief story. I grew up the brother of a type I diabetic. And so he was diagnosed with type I diabetes when we were both young children. And I lived through the change, the paradigm shift of doctor-directed care to patient-owned care through the eyes of my type I diabetic brother.

And it was really about 20 years ago that we stopped calling the doctor for every insulin injection. For the first 15 years of his diabetes, every single time we were going to change insulin, it was a phone call to his doctor, sometimes in the middle of the night. And that shift to self-management was a radical, wonderful shift that my brother undertook.
He is a software engineer and very bright and able to do this stuff. He recently got a continuous blood sugar monitor. And he now is in the best control he has ever been because he has the best real-time instant feedback.

That virtuous cycle of feedback to patients and tools to help them self-manage is really where I think we are headed. And that is really where I want to think about, how do we support that infrastructure much more broadly than just my brother and type I diabetes.

For meaningful use, for those of you who don't know, it's an incentive program, a CMS incentive program for hospitals and eligible providers, largely doctors but dentists and nurse practitioners as well, to implement in electronic health records. And they don't just get paid for plugging a computer into their office.

They actually have to demonstrate
that they're using it in a way that improves patient care. And that way is measured as clinical quality measures. And those clinical quality measures that have a secondary goal is that we start to make electronic and automate the capture of quality measurement that has historically been done as either claims or chart abstraction.

So many of us are living the life of innovating around how to take one type of measurement and now put that into a new world, a new way, the clinical quality measure. And I see many of my e-clinical equality measure friends here.

The program has three stages. The second stage is just about to -- the policy is about to be launched, but it won't actually happen until 2014. And we're in the process of building requirements ideas for the third stage.

This is really the infrastructure.

This is the chance to really influence the
infrastructure of electronic health records across the country because this is a giant cash infusion with quite a bit of incentives from CMS.

We are thinking about how can this infrastructure contain the building blocks that we want for the future of patients, the future of health, the future of health quality. And so this is our opportunity to say, "Here are the things. Here is the focus, few things, that we should invest the country's money in so that we can have the infrastructure that we want to assure that we are able to let people live healthier lives."

In the process of this, we have to think of things in a new way. And one of those ways is measures now become software. Instead of a measure being a piece of paper that lives in a document for someone to read through, the group of us that is working on these e-measures think of it as software.

So if a measure contains 15
different possible survey instruments,
somebody has to encode 15 different survey
instruments into whatever computer system
there is, keep those up to date, make sure
they all work, make sure that they all provide
the kind of results that they provide.

That is completely possible. The
question is the cost. Is that the right way
to spend the money?

And also, just like you can have a
proliferation of Apple and Microsoft and
everybody else with their own kind of
platform, the more that the electronic health
record measurement world has its own way that
each group wants to do things, the more it
costs everybody to keep up, maintain, and use.

So we have really been working
towards this building blocks idea. How can we
take tools that we can use as we build
infrastructure and use them over and over and
over again for multiple different purposes, as
opposed to building software that's only
specific for one little, tiny purpose and
can't be leveraged for other things.

An idea around that is something
like an app store. So you could imagine, any
of you that use an iPhone, Apple built the app
store. And then lots of people can put an app
in there as long as it conforms to all of the
rules to fit into that app store.

And so we are starting to think
about what we call a flexible platform, which
might be something like PROMIS has stood up.
It might be any number of things. But the
ideal world, at least to my perspective, is a
world that the infrastructure exists that can
very easily accommodate new instruments, new
tools, new ways without us having to stand up
something specific for each new thing we want
to use.

And we had a lot of talk about
work flow. Work flow is actually very
important in the implementation of these
measures, as anyone who has done it will tell
you. And I would say you have to think about
not just the work flow of the provider. You
have to think about the work flow of the
patient because now the patient is engaging
with the system and the patient has a work
flow. And their work flow should be paramount
as we think about patient-reported outcomes.

And articulating that and having
an ability to accommodate lots of different
styles of patients in lots of different ways
they might want to engage in the system is
going to be very important. So we need to be
really flexible.

I'll leave with another example
from where I just came from. So in Minnesota,
the Patient-Centered Medical Home Program has
a very specific set of requirements. And one
of those requirements was actually that we
have patient-defined care goals and that we
track the patients' response to their own
defined goals. And they didn't have to be
care. They could be anything the patient
wanted.

So we had to build an electronic system that could track response to any goal that a patient could articulate. And we did. We built a system that could take a free text thing. And it built a five-point scale. And the patient got to determine what that thing was "I want to be able to dance more."

"How well can you dance now?"

would be the item on there and they would get to score them one to five. And then each time they came back, we would rescore and say, "Are you dancing better this time than you did last time?"

This was a home run hit with the patients. They got to frame the whole discussion, not around our scientifically validated survey instrument. They got to frame the discussion about "This is what is important to me right now. This is what I want us to talk about." And then it helped the provider also frame that visit around this
is where the patient wants to go.

So I think that it is really important that we have goals that we can use across different organizations to compare one provider type to another provider type in different scientific frameworks.

It is also really important that we have things that patients say, "This is of value to me." And thinking about how we can do that might lead us down a very different path than thinking about what makes the most sense in the context of a specific disease state or specific category.

Ideally, they are complementary, I think, rather than exclusive.

CO-CHAIR BRENNAN: Thank you very much, Kevin.

Any comments or questions for Kevin before we go on? Yes, Steve? Steve, you have to pick up the microphone. Thank you. And make sure it's red. It's red. Green, green, green. These are red. Those
are green.

MEMBER FIHN: Well, not to be
outdone by my curmudgeonly colleagues, I want
to sort of pursue I think a potential line of
discussion that emanates from this. And bear
with me for one second because I think it is
important when we think about the EHR.

So I really appreciate the
comments, particularly of Kevin. So the VA
has been engaged in a very tortuous dance with
one of our other federal partners for the last
couple of years to build a joint medical
record.

And the upshot of that is that in
order to do that, actually, we have had to
take an approach which Kevin is suggesting of
reusable code. Actually, I'm told next week
we'll have an agreement with Apple for an app
store.

But if you take that and you think
about PROs, so we talk about apps where people
put stuff into the record. These apps
ultimately patients are going to want to get
stuff out of the record, too.

Once you have established that
paradigm of patients being able to right the
record, get back from the record, we no longer
own the records anymore. In fact, one could
then go down the road and think about we
actually don't have an her anymore. The
patient has an EHR. If you want to talk about
patient empowerment, actually, that's where I
think we're going, that we will cling
selfishly for a lot of economic reasons for
the record, but it's going to be an uphill
battle.

So if you then sort of say -- you
know, I mean, I think one of the points I
wanted to make is we are putting -- this
really may be the foot in the door or the
camel's nose under the tent that actually
leads to medical centers having spent tens of
millions, hundreds of millions of dollars on
records which no longer actually may not be as
relevant to them anymore. We could talk about that, in which these patients will actually own their records and have full access to them, which I think if you start thinking about patient recording, I think it is really important.

I think Kevin's other point about patient-identified goals, actually I kind of mentioned that before you came in. We have a program with the severely wounded OEF/OAF veterans who come in. Actually, they have something called a federal recovery plan for a person, you know, work for the government. You know, they're really cool. And, actually, they ask the patient, you know, "What is your goal?"

It's "I want to walk in five years." And that becomes a plan. And all the care gets built around that, those sets of goals. And that's incredibly powerful from my perspective.

But, you know, I think, actually,
you want to throw some bombs here. One of the
bombs may be that, you know, we don't have a
medical record anymore.

CO-CHAIR BRENNAN: I appreciate
that as a starting point. We are going to
hear from Ted Rooney in just a moment, but I
want to see if Kevin has a quick comment he
might want to make on that.

DR. LARSEN: So the advisory
committees that really inform our work for
meaningful use have been thinking about that
exact same thing for quite a while. And what
I'll say from being an organization that did
invest a lot of money and had a great big EHR,
I think it becomes like do you own the stuff
on your computer anymore.

And you start to know which stuff
you do own and which stuff you don't own. And
you kind of know where the interface is. And
so there's some stuff you really do own and
you know you own it. And there's other stuff
that you know that you always go back to the
same website. And that might change, but you
don't really care. And there's some stuff
that's really somebody else's.

So I think that the more you live
it, that that becomes less of an anxiety
provoker for you. But I think when you start,
when you have only ever had a computer not
connected to the internet, the internet is
scary.

But when you are actually there,
you start to figure out, "Well, the internet
isn't that scary. I still can control my
stuff. And I can get other things and share."

CO-CHAIR BRENNAN: We are going to
have a very lively discussion I'm sure about
this one.

(Laughter.)

CO-CHAIR BRENNAN: And the other
people in the room, don't worry. You didn't
waste your money.

Let me move to Ted Rooney's
comments. And then we'll open for the whole
discussion. I remember the people on the
phone and on the internet. You're welcome to
pose questions also. Ted Rooney from Maine.

MR. ROONEY: Hi. I am Ted Rooney,
obviously. I work with two groups in Maine.
One is the Maine Health Management Coalition,
which is an employer/union-led regional health
improvement collaborative that focuses on
performance measure, public reporting, payment
reform, benefit design, consumer engagement.

I also work for a provider of that
multi-stakeholder group, Quality Counts, which
really focuses on quality improvement and
getting consumers involved in the partners in
care. Together they run the Aligning Forces
for Quality initiative in Maine. So I'm sort
of in the middle there.

We also have an active health
information exchange. So everything has to be
electronic anymore. We do our performance
measurement through NCQA and Bridges to
Excellence. And we have got about 500
different physician and PCP offices in Maine.

We define a practice as a physical office
location. So even a group may have seven
offices. We count one practice.

And there are probably about -- I
don't know -- 60-70 percent of those having
reported got recognition from NCQA or Bridges
to Excellence in something, whether in office
assistance survey diabetes or heart disease.
And the diabetes are clinical outcomes
embedded in there.

And one of the things we realize
is we have these steering committees that help
run the program, you know, 14 docs and 6
employers and 3 health plans. And a year or
two ago, we were talking about getting
recognition. And there is this doc over here
who talks about having spent 20 hours trying
to get data out from his chart and even from
the EMR in order to apply to NCQA to get
diabetes recognition.

And so I turn to Frank Bragg, who
is a straight doc with Eastern Maine Medical Center. And I say, "Frank, how long does it take you to get recognized?"

He said, "Well, I don't do anything" because he happens to have Centricity. Centricity works with MQIC. Every night his data gets uploaded, and he gets reports every month. He's just looking at increasing reports. And he gets recognized as a byproduct of giving great care.

And then he takes that data. And he works within his own group of docs, like a performance assessor. You know, they aggregate the data. They work together. They drive performance up.

And we have been fans with the D5, which is the Minnesota measurement that looks at the five elements of diabetes and they track it over time. And we looked at that like five years ago. And there were people doing maybe five or ten percent of getting all their patients the goal of all five areas.
And we like that because that is a patient-centered measure.

Typically you measure how many of the doc's patients have a hemoglobin at this level, how many of the doc's patients have an LDL of that level. But they changed the paradigm by saying, "How many patients are for all levels?"

And, all of a sudden -- I haven't looked for the last month or two, but some of the practices are up in 50, 60, 70 percent of their patients echo for all those measures. They dramatically change the paradigm, I think, in delivering good care.

And that's what we began to see happen when people have data electronically that they control, they believe, and they use.

Now, I'm absolutely convinced that there's no way in hell Frank would have had all the time and resources to do that if there wasn't a performance measure, that his practice wasn't rated publicly on it, he
wasn't paid on it, there's no way he would
have had the time to do it.

So we talk in Maine a lot about
doing the head and the heart. I mean, the
heart is the right thing. It gets the
physician energized. This is good care. But
the head is the business case. Every
physician needs resources. You can't get
resources without a measurement.

And I actually think that CMS has
got it -- well, they've got a lot of things
right, but I have been trying to figure this
out. I think I finally figured out today
where they have this list of 33 measures that
ACOs are to be measured on. And I am
particularly interested in this because we
have one pioneer and three shared savings
models in Maine. So I know they are going to
be measuring it.

And they have this functional
status, health status section that I guess was
going to be part of CAHPS that has got about
5 or 6 SF-36 questions that I could see as well as some other neat health risk questions and everything.

And, whereas, every other mechanism is reporting the first year and to pay the second year, pay for the third year, this one is reporting all three years.

Well, I know for a fact that when I go back to Maine, I am going to start talking a lot about this. And even though these providers won't get paid on this for three years, they are going to pay so much attention to it because they know that's where it's going.

We've got at least providers who used to pay for performance. So you just have to threaten pay for performance down the road.

(Laughter.)

MR. ROONEY: And that gives time for the clinicians to work the heart because most of this stuff is the right kind of thing to do.
And so if you can set the system up in place where you know it's going to be reported, you know it's going to be paid for eventually, but you give the docs the right measures at the right time and let them work with it. I think we're going to see tremendous improvements.

And I think, at least for me, we're big fans of the patient-centered medical home concept or advanced primary care. And it's no longer the docs by themselves anymore.

And so when I think of the clinical data, I think you want the doc or the clinician getting some of the data from PROMIS or health risk appraisal stuff, but maybe in that practice, it's not the doc who is the care manager or maybe it's an aide. I mean, how does the practice take that data and get the right data and the right time so in the clinician in that encounter with the patient, they do the right things?

Like, do we want our PCP spending
five minutes of every visit counseling someone
to stop smoking? No. But we probably want
them to spend 30 seconds telling the
importance of it and giving the warm handoff
to a care manager, who spends a half-hour with
them, refers them to a stop smoking clinic,
and follows up to see how they are doing.

So I think the more we can get
this data electronic, the more we can sort of
have it as being part of a practice. And the
more we can think about who needs to see that
data in order to make right decisions, I think
some of this stuff will work itself out, but
we've got to start because if we don't start,
what we find is our clinicians are our best
innovators.

If you wait until you have it
perfect and then hand it to them, it's too
late. If you give it to them in a safe enough
space, like -- again, maybe we'll publicly
report it next year or the year after -- it's
going to happen, but give them some time to
work with it and refine this up. I think we'll see tremendous innovation.

And if we think of the system of care, it's not just about the doc. You know, I'd like our practices to tell us. You know, even in the docs I talk to, it's funny. You know, we're all individuals. Some docs really want to see the data themselves. They want to go in there, and they want to strategy it and look at it.

And another doc doesn't care about that. He wants his care manager to do it. He just wants to know which patient should I pay attention to? What are the key points? And what should I do with them?

We've got to begin to think about that in a team-based environment. Different members of the team will do different things and do it well. And we talk a lot about practicing the top of your license. So you get the least paid person doing the most. In fact, that's probably what the VA and the feds
have done for a long time.

How do we think of a system of care? And then how do we think of the patient-reported outcomes? What do we measure in the right way? You know it has to be electronic, but who needs to see it when? How often do we do it?

And how do we know or how do we point to -- how does a person know what to do about that? So, for example, if we do a yearly survey of everybody in a practice with a functional status measure and you get an aggregate data, how do you know -- of the 2,000 people who made up that panel, how do you know who to do what with?

So you've got to do it on an individual level, but at an individual level, you know, if you do it on an individual level, can you roll it up hopefully? Because one of the things we're big fans of is that performance measure should be a byproduct of giving great care, like Frank does.
I mean, we don't want our clinicians, we hate it when our clinicians, start to worry about which measurement program to follow. It's a waste of time. We say, "Well, just give great care. You know, let us try to figure out how the measurement and reward systems help you give great care and reward you for it. And do it in such a way that they know they have to do it, but give them enough time so that they have a chance to do it right, at least the early majority folks."

And so that's sort of my take on it. If we don't get it in the electronic health record, no one is going to pay attention to it. But we've got to do it in a way that is helpful and that we find this balance between improvement and reporting and accountability. And I think it's doable. It's just not easy.

CO-CHAIR BRENNAN: Thank you very much.
Comments or questions for Ted?

Yes, Kathy? Go for it, then.

MEMBER LOHR: I am stuck. I am one of those Luddites who sort of doesn't do a lot with computers and that sort of thing, really. But I haven't heard anybody say about whether much of this stuff can be stored on and the software placed on the cloud.

And it's as if people are still talking about storing all this stuff on servers someplace or something like that. I may be wrong, but that was kind of the sense I had. That was one question.

Another, Kevin, is really for you. Is there a time frame for the specs for meaningful use 3? Because would that help to put in some time perspective what NQF is trying to do with this whole effort? And I don't know what the dovetailing is of the time frame.

And the third question, which is maybe for everybody, is, but perhaps picking
up on what Ted was saying about, "We'll start now. And then three years down the road, people will be ready," are medical school or, let's say, stick with physicians. Are medical schools trying to teach this stuff?

Is the AAMC on board? Are the specialty boards and so forth on board, if you will, with moving in this direction? And if they're not part of the conversation soon, should they be? And I don't know the answer to that one either.

DR. LARSEN: So I can tackle the first two. I'll start on the third one. The question about the cloud, we really have a high bar of performance we need to have for patients. And that is we need to be flexible and adaptable to what patients want and need. And we need to give them access. And they need to be in complete control of their own data.

So they need to be able to tell us when and where that data should go and need to
be able to tell us which data they want to go 
where. So that is a lot of the work that ONC 
does: building the standards for how that 
works, certifying that systems can do that, 
and starting to provide a regulatory 
environment that allows all of those sort of 
challenging concepts to live together and make 
sure that happens.

Can it happen? Of course, it can 
happen. Your bank does that, right? So your 
money flows all around electronically. And 
you have trust in your bank that somehow 
through a combination of incentives and 
regulation, that they keep track of your money 
and it doesn't just slip out electronically 
through someone else. We have to do that same 
kind of thing in this more complicated 
landscape of health care.

So, to the cloud question, yes, we 
could use the cloud. Currently the regulatory 
environment and the certification environment 
are a little bit new for people to trust the
cloud to say, "I'm going to put my data there because I believe that it is going to be protected in the same way that the bank protects my data." We have to get to that level of trust in order for people to want to use the cloud for their data to be stored.

It's part of the reason hospitals are a little bit nervous about sharing with each other. It's not because they don't think it's the right thing to do. It's because they're nervous that they'll be held accountable if the information gets shared with somebody else and somebody else doesn't have the same level of security that they do.

Back to the sort of easiest question of yours, which is the meaningful use 3 timeline. The rule, preliminary rule, will probably be out in a year to 18 months or something like that. And so right now the reason I had to step out is we were on a call with the Quality Measures Workgroup. And they're busy framing up what their initial
thoughts will be for how meaningful use 3
looks for doing it over the next couple of
months.

And then that comes out as sort of
a series of questions to the country as a
whole, "Hey, what do you think about this?
What do you think about that? Please send us
comments." This is a very, very public
process. And we are really open. We really
welcome and open everybody's opinion.

But this fall we're starting to
try to frame what will be the major themes for
meaningful use 3. And then over the course of
the winter and next spring and summer, we'll
be refining those, sometimes directing them to
specific committees or task groups or other
places to say, "Hey, give us some really
specific thing that we can go on with this."

And the last to this start-now
question, interestingly, I was at a meeting
the American College of Surgeons put together.
And it was a fantastic meeting with surgery
representatives from a lot of the major surgical societies. They were all on board about this. This was not someone from outside driving this. This was the surgeons saying, "We want this. We think that the outcomes of our patients are what is really important to care."

And they wanted to measure safety. They thought that safety was really important. Surgical outcomes as measured by patients were really important. I can't necessarily speak to all of the other places being on board.

I know that there is a lot of culture change that we will likely need to have health care undergo to really move this, but from the leadership standpoint, I can see it happening.

DR. KOTAGAL: I can speak to some of the other societies. I think all of the boards are certainly working together around this question of quality outcomes. The American Board of Internal Medicine, the
American Board of Pediatrics are all doing that.

And it's all linked now to recertification. So that if you want to be recertified, you have to sort of show data and show evidence around improvement in the process and in the outcomes. And there are some standards for that that have to be met. So that's happened.

CO-CHAIR BRENNAN: Are there other questions? Yes?

MEMBER TORDA: Phyllis Torda from NCQA again. We're working with Kevin and our colleagues from Dartmouth -- and we're almost all here as well -- on three measures that use patient-reported outcomes with some portion of those measures designated for meaningful use stage 2 and were included in the draft reg and then some for stage 3. So I thought it might be helpful to just share what we have been able to accomplish and what we haven't.

So the three measures are
measurement of functional status before and after hip replacement, before and after knee replacement, and for patients with CHF.

In each of those cases, the goal for meaningful use stage 2 was what we called a building block measure. And that was to get some agreement on a defined set of tools, not one tool but a discrete set of tools, and to begin to be able to document that those tools are being used to measure in the case of hip and knee replacement before and after and CHF periodically.

So we have gotten that far. We have applied for the codes, the LOINC and the SNOMED codes, to be able to document that. There are some barriers in terms of the sort of current standardized structures for quality measures and being able to accommodate all of that data, but that is probably more important for the future.

Our goal for meaningful use stage 3, which we're really hoping to start
beginning virtually tomorrow, is to then be able to say what improvement would be.

So the task that we have before us, which is a little bit scary in light of the paper and all of the discussion today, the task that we have before us, is we have three measures. We're hoping to sort of be able to do hip and knee replacement at the same time, multiple tools.

And for each of those tools, we need to be able to define in a standardized way what would clinically meaningful improvement be so that we can then come up with the algorithms and test them to measure it.

So that's where we are and where we are not.

DR. LARSEN: I will make just a quick comment that I look to the NPRM2. We've got about five other measures that, in some way or another, you could talk about or think about their patient-reported outcomes, so
things like the PHQ-9 for depression, asthma control test scores, fall risk assessment, suicide risk assessment. Depending on how you phrase something like a suicide risk assessment and you think is at a patient-reported outcome or is that just an assessment, it kind of depends on your perspective.

The systems think of those as, hospital systems think of those patient-reported outcomes because they have to administer a survey, which feels to them like something else that other people would call a patient-reported outcome. I think that from the measurement community, they think of those as an assessment, not as an outcome.

But, anyway, there are a number of proposed measures that take in account this standardized collection of patient data into the EHR.

MEMBER BASCH: A quick comment again. Sorry, folks. Focusing a little bit
on Kevin's comments, the first is I believe in meaningful use, I think in phase 1, there was engage patients and their families --

DR. LARSEN: Right.

MEMBER BASCH: -- in health care, which I think was more focused around providing a copy of the record to the patient, but the spirit is sort of there. I mean, the functionality wasn't. It probably was developed early on.

I remember a conversation with David Blumenthal way at the beginning. I said, "Well, what about patient-reported outcomes?"

He says, "Well, you know, we're really far from that." But, you know, we've got to focus on the nuts and bolts first. But, you know, it seems that the spirit was there.

One area where this is perhaps immediately applicable is in the review of systems. So if you think about it, you know,
physicians do review of systems all the time. That's a part of the EHR. But they are symptoms, right?

So the way that we think about these things now, 2012, is those things that the patient is in the best position to report, right, should be reported by the patient. A review of systems is nausea, vomiting, fatigue, shortness of breath, chest pain, leg swelling, skin rash, all these experiences that the patients have, but we as clinicians report that.

And we know from abundant research that we as clinicians systematically underreport both the incidence and the severity of those phenomena.

And one could argue that there should be the immediate capacity for an her to collect PRO data for the review of systems as a part of functionality of EHR, some interesting thoughts.

DR. LARSEN: I think it is a great
idea. I think the kinds of things that we are trying to get to apply to lots of patients and lots of providers. And we can actually certify them that the system can do this.

And so those are exactly the kinds of ideas we are looking at because we could apply that to lots of patients. We could apply that to lots of doctors. And we could certify that an electronic health record could do that.

Then it's up to people like the people in this room to figure out what to do with that information once it's entered in.

As far as I know, there's not a particular national standard review of systems other than what CMS requires for E&M billing. And CMS is pretty prescriptive about what they want for evaluation/management billing. So that could become the standard by which we capture these symptoms, but you might imagine a different standard that had more meaning to patients.
CO-CHAIR BRENnan: Ted has a
comment.

MR. ROONEY: As far as meaningful
use, two examples that just make you want to
cry. You know, I mentioned before what we're
doing, NCQA and Bridges to Excellence, which
are really the only real outcome measures out
there for primary care.

And the story I told about Frank
Bragg with Centricity being able to report
through Bridges to Excellence, we have been
working to Bridges to Excellence and all the
major EMR vendors for a year now. We can't
get any other EMR vendor to do it because it's
not a meaningful use.

The second story, we're working
with Gene Nelson at Dartmouth. He's helping
us with some ACO measures. And we have this
wonderful practice right in my hometown called
Emerges of Family Doc. And he's totally
bought into this stuff. He wants to take
PROMIS and integrate it into his practice,
among other things.

And all he wanted to do was get it in a way that makes the work flow. He's got Centricity. He's got a hospital that supports him. And for four months, he's been trying to get his IT folks to get some help to go ahead and implement the PROMIS into the work flow. And he hasn't been able to do it.

I was just talking to Melanie about, you know, being able to do that. And why hadn't he been able to do it? Well, the hospitals are so focused on meaningful use that they are not doing this thing.

And so here you have this really dedicated clinician and his team who wants to do the right thing who can't. So I'm going to talk with Kevin about what I can do to get this stuff into meaningful use. But if we don't get it in now, more and more I think -- there are clinicians out there in the field, maybe not in the Beltway, in the field, who want to do this. This is the right thing to
do. And we've got to help them.

CO-CHAIR BRENNAN: Mary had a
comment first and then Rita.

MEMBER TINETTI: I just wanted to
begin by addressing Kathryn's questions about
medical schools. I am hitting our patient
care curriculum for Yale University, which is
certainly not known as a forward-thinking
medical school. Nobody is here listening.

(Laughter.)

MEMBER TINETTI: But everybody is
on board. And I'm hitting it. We have ED,
surgery, et cetera. And what we're trying to
do is decide, throw out the curriculum that
was great to training doctors for the
Nineteenth Century and make them for the
Twenty-First Century.

And we're all about this stuff.

And, in fact, it's kind of the chicken and the
egg. We're waiting to see what are the
measures we should be training our docs in.
everything. And we're finding a lot of other
medical schools are doing it as well. The
problem is the faculty are so far behind where
their students are. But I think it is going
to happen.

I want to sort of address Ted's
point. I think, like everything, this seems
to be sort of a top-down approach. We're
going to be telling docs what to do and
patients what to do, but I think your point is
well-taken. I've spun that over and over
again.

There are docs that are just doing
it. You know, we machinate about it. We have
meetings. And there are people that are just
doing it. They don't report it. They don't
publish it. And I know it sort of challenges
the entire group to find those people.

And we can talk about the
barriers, why we can't do it. Let's talk to
the people that are doing it. And I think
this would be sort of a great form to sort of
begin to get that to happen.

What is useable? What is really happening out there that's really informing clinical practice as well as policy. And I think that is something I hope this group won't forget about.

CO-CHAIR BRENNAN: Thank you.

Rita and then Lori. Yes? Am I calling you the wrong name? I'm sorry.

MEMBER GAGE: It's all right.

CO-CHAIR BRENNAN: Thank you.

MEMBER GAGE: There is a lot going on, actually, even within the Medicare program, not so much -- well, I can't speak to the physician. And the hospital end is kind of coming along. But when it comes to the area of post-acute care, the rest of the delivery system, the skilled nursing facilities, the rehab hospitals, the specialty long-term care hospitals, the home health agencies, CMS has been pretty busy the last five or ten years in running a consensus-based
approach to identify what types of assessment items or clinical items ought to be measured on the different populations so that they can monitor quality in the Medicare program, set payment rates, all of that.

And, in complement to that, they have been working on a lot of these interoperability issues to develop the HL7-associated terminology to be able to transfer data across the different settings.

So, as this change is going on in all of the ACOs and the A4FQs and the CVEs, out here on the other part of the system, they are kind of moving ahead. And, you know, they have the LOINC codes, and they have the interoperability. And they are working with the standard committees to move it forward.

So, going back to Mary's point about building on what is going on, there's quite a bit going on outside of the physician community and perhaps even within -- I just don't know the physician community.
CO-CHAIR BRENnan: Thank you.

Comment here? Okay. Lori? Wait a second. Kevin has a comment.

DR. LARSEN: The only thing I will say is that the national surveys show that the penetration for electronic health records is much higher in the acute care and ambulatory care and that the long-term care penetration and home care penetration of electronic health records is actually much, much lower. The adoption is quite low depending on how you define the electronic health record.

MEMBER GAGE: Thank you. Very important point about language. And what we found in traveling across the country and speaking to all of the different communities is that language differs.

So the electronic health record, the electronic medical record, you have the little slice over here with the personal health record, not completely duplicative.

And you have the existing system, which the
provider is using to manage their patient care, manage their billing, submit the bills to the payers.

So yes, I think we are just at the start of it, but there's a lot in place on which you can build.

MEMBER FRANK: Hi. Lori Frank, PCORI. I was quite taken with this notion of goal attainment scaling. When you put into Pub. Med. "goal attainment scaling" and "performance measurement," three articles pop up, one of them from 1984.

I mean, that's not necessarily an index of how widely this is used, but I am interested in your opinion and others' opinions about the advisability of GIS for performance measurement.

And then perhaps David could comment on whether there's an opening here for some form of adaptive testing that could be used for the goals.

DR. KOTAGAL: I can't speak to the
her pursuit, but I think goal attainment is very much a part of the self-management work. And, as providers are trained in really working with patients on self-management, posing the question really starts with asking what your goal is. As part of our registries, we are incorporating those into our models for measurement and reporting.

There is a two-part process to it, as we talked about. And the very important part of it, as I said earlier, is flattening the gradients between the provider and the patient so that the conversation about meaningful goals is real and not just simply something that we pass on into the electronic record. But if we are able to do that, I think we will see much bigger shifts in outcomes than we have to date.

DR. LARSEN: And I can speak to the Minnesota perspective that I lived through. So our patients in the medical home program -- in Minnesota, it's actually called
the health home program. And it's not NCQA.

It's actually an 85-item Joint Commission-like certification process, where they do site visits and require a whole number of system organizational culture and processes be in place.

It's a fairly new program in Minnesota. And I led the implementation at our site. And it was really through that process that they would come and assess, "Were you living out the spirit of what we had in our program, rather than looking for just numbers?"

So it was a sort of accountability program, but accountability, much more like a Joint Commission certification accountability, rather than a 60 percent of your patients reached 5 or more.

They didn't really care about that. They cared that when they watched our clinicians talking to patients, our clinicians really asked about the patients' goals. And
we really had a system to track the patients' goals.

That's a fundamentally different sort of qualitative approach to measurement that there is some experience, especially through the Joint Commission, in doing, but it's really the approach that Minnesota took in certifying its health homes.

CO-CHAIR BRENnan: What you are hearing is that the electronic health record doesn't exist in a vacuum. It exists in a context, in a system.

I see David. And did you also want to speak? Yes? So David and then you.

MEMBER CELLA: Lori actually asked about whether there would be an opportunity for measurement or use of item response 3 with goal attainment scaling or goal setting.

Actually, I used to work with Amy Peterman, who is now at University of North Carolina, Charlotte, developed one of these goal attainment measures. They're very tricky
to score because you have to make an assumption that, despite the variety of goals, that the underlying thing that you're measuring is attainment of that goal. And it's difficult to pull that off.

However, I do think there is potential here when I think about a goal of being able to dance. You know, where I would go with that is so -- so if I'm this patient's doctor and this patient says that she wants to be able to dance more, well, why can't she dance as much as she wants? Is it fatigue? Is it pain? Is it limited range of motion?

Now, we measure fatigue, pain, and range of motion. We don't measure dancing. But we measure the things that are interfering with that person's ability to dance.

So I guess I would go down that direction and say to the extent that the goals can be broken down into clinically intervenable elements that are PROs, then you're back to the point where you're to help
the patient dance.

I assume that the doctor is not giving a dance lesson.

(Laughter.)

MEMBER CELLA: You know, they're treating the problem that the patient -- you know, it could be depression. It may not be something physical.

So that's a great idea. It's very personalized. Of course, that's what means something to patients. I guess I would work toward breaking that down to the components that are the clinical intervention points.

CO-CHAIR BRENNAN: Up here in the front. And then we'll also check on the phone line in just a minute to see if there's anyone waiting to ask a question.

MEMBER WASSON: Just to carry David's point further, frankly, very few people have dance as a goal. And I think that is an important point that when you look at tens of thousands of people who do
problem-solving online -- and we have done that -- they have pretty much come into patterns, four or five patterns, for which then you can ask the secondary question of, how confident are you that you can reach this? And you can norm across all the sites immediately. So it's really not a difficult issue technically.

CO-CHAIR BRENNAN: So there are multiple approaches to getting that level of a targeted, patient-focused --

MEMBER WASSON: Yes. I mean, you just picture, for example, adult population. You already know that about 80 percent, it turns out, right now of adult Americans, either they are overweight, they are not exercising, they are smoking, or maybe they are having a few too many drinks. You're up to that number already.

And that can be fed back online in real time "Which one do you want to work on?"

Which is their goal? They'll pick one of them
for starters. And then "How confident are you right now?"

And they'll say, you know,

whatever the number, however you rate it. And then they go into a registry, and you follow them up. And you can see change over time, however you decide to measure it. It's not technically difficult just because, again, you are going with the 80 percent rule. You're not going with the dancing, which is probably one of the one percent rules.

CO-CHAIR BRENNAN: I think what I am also hearing is that there are emerging information tools that might be external to the clinical care facility that might be useful in measuring, identifying achievement of our accountability and quality improvement and goals under NQF that may be useful to examine.

MEMBER WASSON: Just to carry Steve's point, Steve Fihn's point, further and Kevin's point, as we move toward more and more
patient control of their information, it then
can be moved into the electronic health
record, but it need not be strangled by the
electronic health record, which is the problem
we have when we have 200 electronic health
records. And you just articulated an issue in
Maine.

CO-CHAIR BRENNAN: Let me check
and see if there's anyone on the phone. And
then we can go with the gentleman from
Hopkins.

MR. CUNNINGHAM: Excuse me.
Operator, would you check to see if anyone is
in the queue for questions or concerns?
THE OPERATOR: If anyone would
like to ask a question, please press *1 on
your telephone keypad.

CO-CHAIR BRENNAN: And while we're
waiting, please --

MEMBER WU: I would like to ask a
question. I guess I didn't get the exit
right. So I agree that, John, if you ask
people sort of what they want to work on,
     they'll think of sort of -- most of them will
think of four pretty medicalized things, which
have been repeated to them over and over and
which the media repeats them over and over.
     But I think that in the spirit of
it getting late in the afternoon, I would say
that we're not being very imaginative about
goal attainments. And, in fact, most people
have other goals that they want to achieve
that don't have to do with any of those
things. And so I still think there is some
room to develop this.

     There's one tool that you're aware
of I'm sure that's called the C-Qual. It was
developed by an Irish group. And it basically
asks people to identify activities that they
would like to -- that they find important.
And they might include dancing or might not.
They might include going to church or all
sorts of things. But, in any event, they then
rank them in importance and so forth.
I think that there is more room to come up with goals that are really more like people's actual goals.

MEMBER WASSON: Just one point. The tens of thousands of things that I've alluded to are all open-ended. So I wasn't medicalizing it in that sense, but I was using the medical example as one to relate.

CO-CHAIR BRENNAN: Thank you.

Is there a question in the back and then one up here in the front? Yes, sir?

MR. YANG: All right. So I am going to comment on this from a health IT vendor perspective.

CO-CHAIR BRENNAN: Okay.

MR. YANG: I know there has been some mention about data standards. Kevin, you mentioned some of those. And there are several levels of data standardization. And one of the things, actually, as a vendor that we try to implement to some sort of types of systems is to code whether,
for example, each questionnaire, you need to have a code for it, right? And then maybe down to the item level, you have to have a code.

For that right now, LOINC and SNOMED are the two places that have been doing that. However, for LOINC, there is an issue with the copyright. So they couldn't code out of the questionnaires that are out there. They can code only a few of them.

So I wonder if like an organization like ONC is going to be able to bring this issue up a level to see maybe what LOINC -- because the way LOINC implements this solution is assuming the instrument that they are all treated as laboratory results. And I'm not sure that's actually the right model for it, but I just wonder if ONC is assisting in that sort of a matter to maybe working with LOINC or maybe HL7 to see how they can incorporate that into the HL7 standards.

DR. LARSEN: Yes. Thank you. A
lot of us have been involved in this, Phyllis from NCQA. We've talked to NQF about this.

There is this fundamental question that if you're a researcher or a corporation that develops an instrument and you have invested a lot of R&D time and energy into that instrument, how is it that we keep a landscape of people willing to keep those up and to keep those coming?

So intellectual property has really been part of the way that that works, right? It's just like drugs. However, as we try to put these into federal incentive programs, it's a hidden cost for a federal incentive program if now everyone that uses this instrument has to pay a licensing fee to whomever is the holder of that intellectual property.

And so I won't necessarily speak for CMS, but I think I speak for CMS in that they are looking for things that don't have hidden costs. And ideally CMS would like to
pick public domain instruments, things that
don't have a hidden cost.

As we have been working through
some of the building out functional status
measures for the HR, we run into these
questions that if an organization or a
university or an individual holds intellectual
property and now we try to put it into the
measure for a federal incentive program and we
can't really resolve the intellectual property
questions, do we just eliminate that
instrument from an intellectual property and
hidden cost point of view and just go with
others instead?

If we do that, then we don't
necessarily have a landscape where people
develop those things because now they don't
have an incentive to develop those things
anymore.

So I don't know that I have an
answer to that question, but I think it's a
great thing for a group like this to think
through. I think it's especially good for NQF as a measure endorser to think through.

We are taking tactical approaches to figuring out how to build data standards around those. And so far it has been pretty easy except if the IP owner is in Europe because they have no sort of reason to move at our timelines here in the U.S. And although they're nice enough, they don't feel the same kind of pressure to get into a U.S. program that a U.S. IP owner might.

CO-CHAIR BRENNAN: Thank you, Kevin.

We are going to have to have the last question now or comment here from the table. And then I'll let the panelists have the final remarks. You are all right. Okay.

Let me ask the panelists if you have a closing remark. And if you don't, if you could address one of the questions we've left unaddressed, which is, what needs to happen to get patient-reported outcomes into
an electronic health record? What would be
the first thing you would do?

And while you're thinking, I'll
start and say mine is I would resolve the
issue of at what point is it appropriate and
at what point is it risky for patient-reported
outcome to be released into a clinical record
system, where it enters back into the care
process in ways they may not understand, the
patient may not understand?

For example, experience data put
into the clinical record seems to me to be
inappropriate. But other patient-reported
outcomes might be actually quite useful.
However, there may be things patients will
refer to control.

MEMBER WILKINSON: I would be very
curious eventually how one would determine
what was important for the patient to know and
what the patient would actually understand.
The reason for my question is
probably obvious. It is that so many times
experience shows that, just as clinicians can
underreport something that the patient views
quite differently, as was shown on one of the
slides, I think that principle could be
extended to a lot of other settings in which
measures and others presume they know what is
understandable or important and they leave out
of the equation communication.

CO-CHAIR BRENNAN: Yes.

MEMBER WILKINSON: I am
oversimplifying for the purposes of time.

CO-CHAIR BRENNAN: No. We will
have a whole day for more conversation
tomorrow. So let me go down the road and let
Ted start and then Uma and then Kevin.

MR. ROONEY: I don't know the
right lever, but getting PROMIS and a general
health risk appraisal into every EMR so a
clinician could use it would be huge.

CO-CHAIR BRENNAN: Thank you.

DR. KOTAGAL: I think engaging the
clinicians through incentives in a way that's
meaningful to shift from clinical outcomes to
patient-reported outcomes.

CO-CHAIR BRENNAN: Excellent.

Kevin?

DR. LARSEN: So I am with Ted. I have a little bit of mantra of just start.
Don't get worried that you don't exactly know the whole ending goal. You've got to start someplace.

And so then I am working very tactically. What can I start with with meaningful use 3? What is the right thing that we can use that lever to build infrastructure that gets us the furthest the fastest in the best way?

And so I am open to your input. Please help me and help our FACAs figure that question out.

CO-CHAIR BRENNAN: And, David, you have been so good to us all day I'm going to let you have the very last word of the panel. Thank you.
MEMBER CELLA: I like the idea of just starting. You know, the kind of work that we do and others of us, others of you who do similar work, we can get very bogged down in the detail and worry a lot.

And there is a lot of good information that we can already capture. And we are ready to go. And we'll get better along the way.

CO-CHAIR BRENNAN: Thank you very much. It's been a really excellent day. And I want to just remind you, first of all, that the panel would be here mostly for tomorrow. Speak up more. All right. The panel will be here mostly for tomorrow.

We're going to have a couple of closing comments before the day is over. I will be at -- I want you to just reflect for a few minutes about the key points that you heard through the day since tomorrow's work, we'll be moving into small workgroups.

Now, several of you have noticed
that there is a small dot on the back of your name tag. That identifies which workgroup you're going to be in. And for those of you who need a reminder, there is a paper on your table that summarizes that. And there is also a screen up here.

It is color-coded. If you don't have a dot on your name tag, please see one of our staff. The NQF staff will make sure you are in a workgroup tomorrow.

The workgroups will be focused tomorrow throughout the morning on, first of all, breakout session for selecting individual-level PROs for performance measurement. And then there will be a discussion on the next steps for what characteristics should be used for selecting PROs.

I am going to now turn to Karen or Karen Adams or Helen to see if you have closing remarks that you would like to make for the day.
MS. PACE: I will just make a quick note. As Patty said, if you don't have a colored dot on your name tag, see Gene Cunningham. And if we didn't receive your request, we tried to honor first and second requests. There is a limited availability to switch, but one of the groups is totally full. So, again, you would need to see Gene Cunningham about that.

We really appreciate all of your engagement. It's been very nice to see everyone involved, including our audience.

And I'll just give it back to Patty. And also, Joyce, do you want to make any comments?

CO-CHAIR BRENNAN: So on behalf of Joyce and myself, we want to thank you for taking the time to be here. We want to thank you to all the speakers today and particularly to David for a great deal of work and provocative thinking.

We want to thank all the
curmudgeons and the provocateurs in the room.

Please come back tomorrow and be provocateurs.

Thank you very much.

(Whereupon, the above-entitled matter went off the record at 4:45 p.m.)
modeling 283:6
models 13:19 98:15
189:1 283:2,16
354:18 382:7
moderate 142:9
moderator 4:8,17
4:22 5:6,15 37:17
moderators 42:19
modes 163:18
171:20 193:10
255:14 256:1
modified 194:12
329:15 331:1
modify 334:2
modifying 213:12
modules 77:8
momentarily 434:1
momentary 287:9
monday 77:8
modules 434:1
moment 502:6
momentarily 242:20
moments 8:1
momentum 98:16
MONDAY 1:6
Mondays 14:14
money 72:18
133:12 169:8
294:18 339:12
340:9 348:14
349:20 363:11,14
monitor 98:7 287:9
304:10 313:7
337:3 379:4
monitoring 304:16
306:12
month 74:7 182:20
352:8 353:10
months 11:22
20:18,19 45:13
123:6,7 130:12
139:12,12 153:9
182:21 309:12
312:16 318:12
328:18 364:18
365:3 375:5
mood 90:17
morning 6:5 7:21
40:22 42:20 64:15
69:21 95:14 96:7
117:22 128:11
209:18,22 210:18
248:21 400:12
MORSELL 3:16
Mort 2:5 4:18
110:12 114:1,14
122:13 128:7
132:5 149:10,18
155:8 178:17
197:1
mortality 219:14
225:9,10 232:4
Moseley 2:6 4:12
40:16 51:22 91:21
103:6
Mother 292:16
motion 385:13,15
move 10:19 15:4
27:15 37:16 40:4
40:7 72:11 79:1,7
89:7 102:7 124:18
165:12 183:8
184:18 186:11
188:11,19 202:17
203:16 210:20
228:1 250:20
262:17 303:3
349:21 366:15
379:17 388:22
395:7
moved 336:3 389:2
movement 17:8
98:17
mover 131:15
moves 15:6 28:13
50:3 109:1 136:20
298:22
moving 6:17 14:11
165:13 183:6
202:12 207:1
224:6 262:2
263:13 302:5
303:18 362:8
379:14 399:21
MPA 1:21
MPH 1:18,23 2:8,9
2:13,14,15,16,18
2:24 3:9
MQIC 352:6
MSC 1:14
MSN 3:8
MT 3:7
multidimensiona... 190:21
multiple 39:4 73:10
171:8 218:1
285:17 340:21
369:9 387:10
multiples 327:9
multiplied 67:15
multiply 84:5
multi-center 68:8
multi-stakeholder 350:12
MUNTHALI 3:17
MUP 1:13
musculoskeletal 16:20
myth 44:12
name 99:11 122:5
218:13 299:15
306:16 378:9
400:2,8 401:3
nameless 320:21
Nan 41:14
nanosecond 143:3
narrative 335:2
narrow 81:18
278:18
narrowly 53:19
263:14
national 1:1,8 2:3,6
2:16,18,20 4:5,12
5:10,17 11:18
40:16 55:3 57:18
58:17 20 99:12
103:22 113:14
137:18 227:3
246:2 373:15
380:5
nationally 220:6
naturally 14:19
nature 33:15 53:14
53:16 59:5 60:14
nausea 232:18,20
372:8
nay 107:18
NCI 185:4
NCQA 100:16
350:21 351:7,20
367:13 374:6
383:1 393:2
nearly 70:6
near 171:4
259:19 271:18
neat 355:2
necessarily 16:1
67:17 72:11 94:20
199:2 230:4
239:21 268:11
273:12 274:5
279:6 296:3 304:6
304:8 305:9 320:8
366:11 381:13
393:19 394:16
necessary 8:20
91:8 191:13
192:19 193:7
194:7 220:10
317:9
need 8:21 16:3
18:17,22 19:6
20:5 25:17 26:19
29:12 36:7,20
37:20 39:15 47:8
50:5 52:7 55:1
56:1,4,6,8,10,13
59:14 62:4 64:5
71:3 77:2 83:19
84:12 85:14,18
87:9,21 88:6
91:11,14 105:3
106:4,11 109:21
110:18 137:9
148:21 150:4
158:16 161:2,15
165:12 194:7
197:7 200:16
204:10 205:10,15
205:21 206:11,15
210:19 211:6,13
214:20 215:15
217:1,17 221:16
225:14 227:19
230:17 231:3,10
231:12 232:15
233:4 244:17
253:12 255:7
259:15 261:11
275:15 279:9
305:11 307:3
310:4 320:22
321:1 324:7 326:4
342:12 362:15,16
362:17,18,19,21
362:22 366:14
369:11 389:3
392:1 400:4 401:8
needed 9:5 55:7
63:14 181:7
189:17 272:9
needling 85:21
neendn’t 294:2
needs 19:22 44:15
48:2 60:19 69:10
76:11 85:11
108:21 137:7
161:4 188:19
193:12,20 194:19
194:20 207:14,20
255:19 263:16
266:20 273:14
294:1 307:13
320:4 329:15
354:8 357:11
359:6 395:21
negative 177:10
187:2
Nelson 2:8 4:19 5:9
110:13 114:2
132:22 152:1,20
153:15 156:13,16

Neal R. Gross & Co., Inc.
202-234-4433
resources 112:14
167:7 189:17,20
277:6 278:7,13
279:7,10 305:12
305:14 353:20
354:8,9
resource-intensive 191:15
respect 61:7,18
91:6 92:6,9 142:7
206:4 208:18
231:6 310:14
respectively 140:6
respiratory 239:2
respond 37:8 51:1
51:3 84:2 90:14
93:1 100:17
165:22 221:14
234:18 260:20
263:12 281:6
287:19 293:7
308:8 310:21
311:2,6
responded 260:17
respondent 67:13
120:13 164:1
172:5 255:3 308:4
respondents 63:5
respondent’s 307:22
responding 165:22
179:17,21 191:1
261:15
responds 84:1
response 4:21
95:18 132:18
154:10 168:15
171:16 175:8
176:6 180:13
199:15 221:6,7,11
221:15 222:3,4
233:18 234:14
238:18 253:8
284:22 293:17
294:3,12 297:11
298:5 310:8,19
325:3 333:6,9,11
335:18 342:20
343:3 384:17
responses 63:5
92:4 153:8,18
159:13,16 199:8
200:8 201:9 260:9
260:12
responsible 202:21
231:21
responsible 46:8
171:7 252:15
253:5 256:4 260:5
311:22
responsiveness 252:12 253:7
rest 246:6 313:10
378:17
restoration 137:11
restrooms 37:20
result 15:1 19:17
56:2 129:5 239:21
253:2 271:3
resulting 287:11
results 47:18 57:12
71:7 78:11 129:5
130:20 139:15
140:16 141:6,13
141:19,21 142:6
142:11 144:9
152:2,11 153:19
153:21 190:10
193:20 194:11
311:21 319:22
340:6 392:16
results-driven 65:4
resumed 110:10
240:21 316:11
retention 275:19
276:6 300:2
retrospect 121:7
return 204:4
returning 29:12
returns 237:8
reusable 345:17
revascularization 101:14,17
revascularizing 101:15
revealed 82:15
215:19
review 10:22 82:17
90:4 153:18
162:19 208:15
216:21 249:5
252:6 271:11
371:21 372:1,8,19
373:15
revisions 10:10
revising 280:8
revolutionary 306:2
reward 360:7,8
re-read 84:21
rheumatoid 134:7
144:8 267:19
311:11
rich 218:10
Richard 1:13 5:1
162:4 195:3
217:13 236:21
richer 267:3
riddle 133:6
ride 135:3
right 6:3 7:20 12:8
29:8 32:21 33:6
43:3 54:18 62:20
63:13 69:13 78:10
78:19 79:12 80:16
105:11 110:6
112:11 121:1
126:3 131:14
138:7,9 140:12
141:9 142:6,12
145:13,18 147:12
150:3,4 151:4
158:4,14 173:17
179:12,17 183:6
197:11,13,13
202:19 206:17
209:5 218:3
222:11 224:4
239:15 264:18
292:21 294:18
295:13 298:9,19
329:12,13,16
340:8 343:20
346:4 354:5,12
355:21 356:4,5,19
356:19,21 357:12
359:5 360:11
363:10 364:10,19
371:4 372:3,7
374:19 375:16,22
378:10 387:15
388:2 389:22
391:12 392:2,5,17
393:12 395:17
397:17 398:12
399:14
right-hand 333:5
rigor 157:13 229:2
rigorous 12:20
13:10 107:19
229:1
rings 313:17
risk 25:14 57:15
66:16 134:22
138:5 141:22
142:16 158:22
159:3 185:8
198:13,15 201:13
266:9,13 268:12
268:15,18,19,20
268:22 269:3,6
286:10,13 294:14
355:2 356:15
370:2,3,4 397:18
risksy 396:6
risk-adjust 286:10
risk-adjusted 94:14
Rita 376:3 378:8
RN 1:17 2:11,14,17
road 130:19 346:7
355:17 362:2
397:14
ROB 2:22
Robert 105:7
robust 121:17
role 39:10 89:19
206:9 207:2
CERTIFICATE

This is to certify that the foregoing transcript

In the matter of: Patient Reported Outcomes

Before: NQF

Date: 07-30-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

________________________
Court Reporter

NEAL R. GROSS
COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
WASHINGTON, D.C. 20005-3701
(202) 234-4433
www.nealrgross.com