NATIONAL QUALITY FORUM

+ + + + +

RISK ADJUSTMENT AND SOCIOECONOMIC STATUS OR SOCIODEMOGRAPHIC FACTORS EXPERT PANEL MEETING

+ + + + +

WEDNESDAY

JANUARY 15, 2014

+ + + + +

The Expert Panel met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Kevin Fiscella and David Nerenz, Co-Chairs, presiding.

PRESENT:

KEVIN FISCELLA, MD, MPH, University of Rochester, Co-Chair
DAVID NERENZ, PhD, Henry Ford Health System, Co-Chair
JEAN ACCIUS, PhD, PMP, AARP
ALYCE ADAMS, PhD, MPP, Kaiser Permanente Division of Research
MARY BARGER, PhD, MPH, CNM, FACNM, American College of Nurse-Midwives
SUSANNAH BERNHEIM, MD, MHS, Yale-New Haven Hospital/Center for Outcomes Research Outcomes
MONICA BHAREL, MD, MPH, Boston Children's Hospital
MARY BETH CALLAHAN, ACSW, LCSW, Dallas Transplant Institute
LAWRENCE CASALINO, MD, PhD, Weill Cornell Medical College
ALYNA CHIEN, MD, MS, Boston Children's Hospital

MARSHALL CHIN, MD, MPH, University of Chicago*
MARK COHEN, PhD, American College of Surgeons
NANCY GARRETT, PhD, Hennepin County Medical Center
NORBERT GOLDFIELD, MD, 3M HIS
ATUL GROVER, MD, PhD, FCCP, Association of American Medical Colleges
DIONNE JIMENEZ, MPP, Service Employees International Union
STEVE LIPSTEIN, MHA, BJC HealthCare
GENE NUCCIO, PhD, University of Colorado Anschutz Medical Campus
SEAN O'BRIEN, PhD, Duke University Medical Center
PAM OWENS, PhD, AHRQ
NINEZ PONCE, MPP, PhD, UCLA Center for Health Policy and Management
THU QUACH, PhD, MPH, Asian Health Services
TIA SAWHNEY, DrPH, FSA, MAAA, Illinois Medicaid
NANCY SUGG, MD, MPH, University of Washington, Harborview Medical Center
RACHEL WERNER, MD, PhD, University of Pennsylvania

NQF STAFF:

CHRISTINE CASSEL, President and CEO
HELEN BURSTIN
ANN HAMMERSMITH
KAREN PACE
SUZANNE THEBERGE

* present by teleconference

C-O-N-T-E-N-T-S

C-O-N-T-E-N-T-S (CONTINUED)

C-O-N-T-E-N-T-S (CONTINUED)

C-O-N-T-E-N-T-S (CONTINUED)

C-O-N-T-E-N-T-S (CONTINUED)

C-O-N-T-E-N-T-S (CONTINUED)

C-O-N-T-E-N-T-S (CONTINUED)

P-R-O-C-E-E-D-I-N-G-S

8:31 a.m.

MS. THEBERGE: Good morning, everyone. We're going to get started.

Operator, can you open the phone lines?

THE OPERATOR: Okay. One moment.

Okay. I am going to go ahead and read the introduction. One moment, please.

MS. THEBERGE: Okay. Thank you.

THE OPERATOR: Welcome to the Risk Adjustment and Socioeconomic Status meeting. Please note today's call is being recorded and all public lines will be muted during this broadcast.

Committee Members, please note your lines will be open for the duration of today's call. Please be sure to use your mute button when not speaking or presenting, and please do not place the call on hold.

If you need any assistance at anytime today, please press *, then zero. An

operator will assist you.

For technical support for today's program, you may send an email to nqf@commpartners.com.

Today's meeting will include specific questions and comments period. However, you can submit a question at anytime during today's presentation using the web conference window. To do so, simply type your question to the text chat box area at the lower left corner of the window. Be sure to click Send to send your question directly to our presenter.

During the designated public comment period, you will also have the opportunity to ask live questions over the phone by simply pressing *1. These instructions will be repeated later in the program.

And now, it is my pleasure to welcome you to the program. Let's get started.

MS. THEBERGE: Good morning, everyone, and welcome to the National Quality Forum's Risk Adjustment and Socioeconomic Status or Other Sociodemographic Factors Expert Panel Meeting.

Before we begin, I just want to make a quick technical announcement. The webinar number for day one that's on the agenda, there's a typo. The correct webinar number is 400441. That's 400441 for anyone who is on the line and having trouble seeing our slides.

So, with that, I am going to go ahead and get started, just do a quick overview of the purpose and scope of the project.

My name is Suzanne Theberge. I'm the Project Manager here for this project at NQF.

The purpose of this project is to identify and examine the issues related to risk adjustment outcome and resource use

performance measures for SES and other sociodemographic factors. And we're looking to this panel to make recommendations regarding if, when, for what, and how outcome and resource use performance measures should be adjusted. And we are focusing on outcome performance measures, outcome performance measures for accountability applications and consideration of SES or other sociodemographic variables as factors for risk adjustment.

But we are not going to be looking at specific performance measures, although we do have a panel this morning that will use a couple of specific measures for illustration purposes. We are also not going to focus on adjustments for determining payment for services such as capitated payments, and we are not going to be selecting a particular risk model or approach today.

Before we dive into the project, I just wanted to go over the project schedule very quickly.

Today is the panel's in-person meeting. We will have two followup conference calls in the next month, one on February 10th and one on February 18th. We will bring the panel back together continue the discussion, to review the draft report that the project team will be working, and to just really continue the discussion.

Once we have completed those two calls and written up the draft report, we will go to the NQF member/public comment period, which is a 30-day period from late February through late March. We'll take comments.

Following the close of that, we will send all those comments out to the Committee and bring you back together on another conference call in April to discuss the comments, make changes to the report, changes to your recommendations as necessary.

NQF's Consensus Standards Approval Committee will review the project on May 13th during a conference call. And then, the NQF

Board will review the recommendations in June and put the stamp on approval on those. And we expect to finish this with a final report by June 30th of 2014.

All right. Now I am going to turn it over to Karen.

MS. PACE: Good morning, everyone.

Just before we get into our setting the stage -- and we will -- after our Co-Chairs speak to us, we are going to do introductions of everyone, but we wanted to kind of set the stage before we got into the individual introductions.

So, I just want to introduce who is up here. We have Kevin Fiscella and David Nerenz, our Co-Chairs, that we are delighted to have working with us.

Next to them is Helen Burstin, our Senior Vice President for Performance Measurement, and Ann Hammersmith, our General Counsel, who also will say a few words when we get into introductions.

But I will just ask, before we turn it over to our Co-Chairs, Helen, if you want to make any opening remarks.

DR. BURSTIN: Just to say good morning. We will have a chance to discuss this further later, but certainly we recognize how high-profile and how important this work is, and we really thank you for taking the time to review the scads of remarkable materials that all of you suggested. And we really hope to have something that comes out of this that really helps push the issue forward. In a way, it feels like it has been sort of stuck for a while. So, really, thanks to everybody.

MS. PACE: And as you have seen from the agenda, we have a very packed agenda. We think we have set the times to allow for lots of discussion and interaction with our expert panel members. But we will try to stay on time and appreciate everyone helping us with that.

So, with that, I am going to turn it over to David to start us out. And then, we will go to Kevin after that.

CO-CHAIR NERENZ: Well, thanks and good morning.

I will make one observation. We are already ahead of schedule. This may be the only time. So, enjoy it for the moment while it lasts.

(Laughter.)

Thank you for being here. I think we are addressing a very significant issue. You have all taken time from very busy scheduled to be here and we do appreciate it.

All of us are here because we have some important perspective on the issue in front of us, some domain of expertise that NQF wanted to include in the discussion. So, you're here because you're special, and we appreciate that.

Just a few things about how we think about today's work in the larger

context.

The overall task in front of us has already been described. I think, as we go along, we will probably develop a little clearer sense of exactly what is within the target of our discussion, what might be outside, and I suspect once in a while we will probably deviate a little bit out to make a point or two, and then, come back in. Our job is to try to keep those deviations at least generally on point. We will try to do that.

But a little more specifically, our main task is to develop a set of recommendations. The recommendations are important. They, presumably, will influence NQF policy in this domain. NQF, in turn, I think as you all know, has a very important role, particularly with regard to CMS and what CMS does in its various performance measurement programs.

Much of this is set in law in the Affordable Care Act. But, even if it wasn't,

still, the policies and positions of NQF and the technical specs of the endorsed measures have great influence throughout CMS and elsewhere. So, we are taking up an issue that is part of that process.

Our task in front of us is to examine, first of all, associations between various SES variables and performance measures, either as individual examples or as a class, and then, to develop recommendations about, first of all, whether some set of SES variables should be included in adjustment models. And then, if the general answer to that would be yes, then there are all the detailed questions of how, when, which measures, which variables, which models.

In two days, we cannot possibly work through all the details. As Suzanne said, our task is not to go measure by measure, detailed model choice by detailed model choice. We really are being asked to work at the level of general principles, so

that details can follow, either after this in-person meeting to some extent or even after that, in the hands of the measure developers and the other NQF panels who look at individual measures and their parameters.

So, in our two days here together in person, much of the agenda today is on some expert presentations, where we will get some information. Much of it will be new to many of us or at least it will be designed to illustrate certain points, either about the relationship between SES variables and some set of performance measures or, then, about some of the technical details; for example, about where these data might come from or about how they might be included in an adjustment model.

The object here is not to immediately point to something and say, "Yes, this is right. This is the answer," but to say, "Here are a range of things that we think are important and that we should consider."

So, I think our task through much of the first day is to listen, to ask clarifying questions, perhaps to challenge a bit about interpretation. But, as we go through much of today's agenda, we are not seeking to decide or specifically endorse something that a person is presenting.

Okay. I guess we move to the next slide.

So, rules of the game. I think just about everyone here has been in many groups of this types. Although some of us know each other, we are barely acquainted and some of us are just, frankly, strangers to each other. We don't have a lot of time to get to know each other.

For those of you who have been on IOM committees that meet five or six times over a year and a half, you can appreciate that that pace is quite a bit of different. Often, in that time you spend the first meeting just basically getting to know each

other. You develop some trust. You develop a sense of where people are coming from. We don't have a lot of luxury for that. We are going to have to just move right through into the more substantive agenda.

So, a few requests I think are straightforward. First of all, listen with an open mind. Understand that those people who are here are here because they are experts of one type of another. When we may disagree, let's disagree in a polite, professional fashion. I don't know that we can get the police up here rapidly to break up fist fights. I hope none occur.

(Laughter.)

So, we are not really set up for that.

I think, again, we are here as a group of colleagues. I think what we seek to do is to identify the common ground that can serve, then, as the basis for a set of recommendations that, presumably, all of us,

then, at the end feel like we can endorse. Again, there may be details in how this all rolls out where we still have some differences of opinion, but I think our task, at least from the front of the room here, is to try to find those areas of common ground, try to help work through some areas of disagreement, if we think we can, and then have that be the core of what we report out.

Okay. Is this our --

MS. PACE: We just wanted to remind people about some of the definitions that we are using in the project. We talked about these on the conference call, and they are in your materials. So, we don't have to, obviously, read through these, but just wanted to kind of keep us grounded on some of our definitions. So, I think we can quickly move through those.

The next one.

CO-CHAIR NERENZ: Okay. Well, yes, clearly, again, the task here is not to

read these slides. That's a horrible abuse --

MS. PACE: Right.

CO-CHAIR NERENZ: -- of people in a room to read what's --

MS. PACE: Right.

CO-CHAIR NERENZ: -- on a PowerPoint slide.

But if we could just actually back up to the previous slide, I think a point to make about that, our charge is not specifically about disparities. It is really about measurement and, a little beyond that, it is about the accuracy, I'll call it accuracy and informative nature of the measures, and how SES plays into that.

The current NQF position really does speak, though, to disparities. It says that adjustment for SES can have the effect of masking disparities and, therefore, that is a concern. So, I think we just need to keep that in the back of our mind as we go through the discussion, recognizing that there are

both health and healthcare disparities that are somewhat different in nature, and that the context that we enter into about SES adjustment includes concern about the effect of adjustment on disparities or, presumably, the effect of lack of adjustment on disparities. And we will certainly discuss that. Okay?

MS. PACE: Next slide.

CO-CHAIR NERENZ: Now the word "outcome" was in Suzanne's discussion. I think that's worth a little special emphasis. There can be some fuzziness around the edge of what's outcome; why do we even talk about outcome?

In general, my sense in the domain of performance measures is that we tend to think about adjustment, including potentially SES adjustment, a little more in the domain of outcomes than we do in some of the healthcare process measures. That is not universally true. It is not strictly true. But it tends,

all else equal, to be true.

And that is because there are some process measures where the thing being measured is so mechanical that the right thing can be done virtually all the time for essentially any patient. And so, you just don't think a lot about adjustment. Or whatever is necessary for adjustment is just defined by how you define the denominator population.

But, more frequently in the domain of outcomes, other factors, other causal pathways play into the eventual measure performance. And some of them have to do with comorbidity and clinical severities, but some of them may have to do with these SES variables that we are talking about.

So, that is the reason why the word "outcome" is put in the framing. I suspect that we may in our discussions bring up examples of performance measures that may not be strictly outcomes, at least in the

Donabedian sense, but we will at least try to remember that this is the domain in which what we talk about probably matters most.

Just at the bottom of this slide, we point out that there are some outcome measures that may be expressed as cost. And this something where people can debate the semantics, but at least we make the point here that there are some performance measures that get expressed in dollar terms that at least are in the scope of what we are talking about.

I don't think we need a lot of definition here, except maybe to point out that, even though the general concept of risk adjustment can have different meanings, there are different statistical methods, there are, say, stratified reporting approaches that some people may quibble and say, well, that is not really adjustment at all; I mean, you haven't really done any mathematical manipulation of anything. You've just taken a measure and sort of broken it out.

But, essentially, we are talking about a conceptual issue here of accounting for or working with variants that may otherwise create some distortion or some inaccuracy of interpretation of a performance measure.

And I don't know; again, you can read faster than I can talk, and these were all discussed in our conference call, anyway. So, I guess we would just say, if there are any points of lack of clarity or questions, this might be the time. But I think that we covered this already in the first phone call.

MS. PACE: All right? Okay, we will move on to Kevin.

CO-CHAIR FISCELLA: For me, this question has had my head spinning. I mean, it really is a daunting task.

And so, to try to clear up some of the fuzziness, I began mapping out the two different pathways in my head. And I thought I would share it with the group.

So, the task before us is the question placed very simplistically: adjust the quality metrics for SES or not? So, if we go yes -- yes, these are going to be very fast slides -- we worry about a two-tier system, which, in turn -- next slide, please -- brings up some of the concerns that David mentioned: masking disparities, setting a lower bar for those with lower SES, and either masking or even fostering institution or individual bias.

This, in turn, could cause healthcare disparities to drop off the public agenda, after they have been on there for more than a decade now, and create lower expectations of care, ultimately, resulting in worse care and, in turn, either healthcare disparities persist or worsen.

So, I think the pathway at the top, and I think later on, if folks have additional concerns, I think these should be added to the top pathway. The bottom pathway, the no pathway, is a bit more complex. You

know, at a very basic level, there is the potential that the quality measures could be perceived as invalid, which, in turn, could undermine -- next slide, please -- undermine confidence in the metric itself.

But a bigger concern really is that providers and organizations serving large low SES populations might, on average -- and, obviously, we can all think of exceptions -- have worse performance metrics. Well, why might this happen?

Two reasons. One is greater population need, and the second is fewer organizational resources. In organizational resources, we mean things like resources to fund quality improvement or Lean Six Sigma, funding for patient enablement services and language translation, low health literacy, et cetera. Training needs, level of IT, and electronic health information technology, and time needed to train and transform systems.

So, if we think about population

need and we think about resources, the risk is that in providers of the disadvantaged that there is a potential mismatch between need and resources. And so, that residual, that difference, when you don't have the resources to address a need, some type of healthcare disparity may often show up and some of it will be captured through metrics.

Next slide, please.

So, some of the unintended consequences -- I think a lot of us have thought about this, and it was raised in the first call, of course -- is worse payment with P4P, pay for performance, which, in turn, can undermine -- next slide, please -- fewer organizational resources, which, in turn, can affect quality of metrics.

So, one has a cycle, a feedback cycle, that actually can worsen. A similar feedback cycle would be contractors and patients. If they see that the performance is worse for safety-net providers, they may avoid

contracting or avoid obtaining care there, which, in turn, could lead to fewer organizational resources, which could set up a cycle back with patients and contractors avoiding it because the resources are fewer, as well as affecting the quality metrics itself.

Next slide, please.

And a third potential consequence would be -- next slide -- would be that the providers avoid lower SES patients and populations, which, again, could result in fewer organizational resources. If you have, for example, disciplines from a certain provider who are avoiding an organization because they don't have the equipment and resources needed to do that, now you have a resource problem and a difficulty in addressing that need.

Okay. And again, we have got the same concern here. Healthcare disparities persist or worsen with this pathway.

So, let's briefly decompose need and resources. Next slide, please.

So, what do we mean by need? One is -- and I think this came up on the first conference and we will have some talks that will address this -- worse health among patients with lower SES. Why worse health? Worse health based on a number of different factors.

Next slide, please.

These include early life factors, actually, beginning oftentimes even before birth; epigenetic factors many of you are quite familiar with, the cumulative lifelong effects of stress, material deprivation, psychological and behavioral factors.

Next slide, please.

Another key pathway, in addition to the health pathway, is what I have labeled here as access and adherence factors.

Next slide, please.

And these are factors that we are

all familiar with among low SES populations: the ability to afford care, levels of patient activation, health literacy and numeracy, limited English proficiency, differences in culture, lack of social support either within the family or in the community, homelessness, and even the stress and skepticism of providers in medical institutions.

And these factors are compounded because, by definition, safety-net providers serve and have a concentration of low SES patients within these providers, basically, for two reasons.

Next slide, please.

One is a concentration geographically to the socioeconomic residential segregation. And the other is organizational mission itself to serve the underserved.

Next slide, please.

In terms of organizational resources, we are all familiar with it. More

uninsured and, also -- next slide, please -- the payment distribution, lower payments through Medicaid and Medicare relative to private payers.

Next slide, please.

Some additional issues include systematic measurement area. This gets -- and I think some of the talks will address this -- unaccounted disease severity through traditional case mix adjustment, which often doesn't directly measure severity, but rather diagnoses.

And the ICD and CPT coding bias, meaning that oftentimes the underserved have fewer visits and, as a result, may have fewer codes for that visit. They may utilize healthcare less, undergo fewer procedures, which can introduce, if you use ICD-9 and CPT coding, you may underestimate either the disease severity or even comorbidity.

Next slide, please.

So, very briefly, mapping this

out, one can begin to see -- and I am not necessarily proposing all of these, but these are just examples of opportunities to begin thinking about "both/and" rather than an "either/or" to the basic question. One is organizational stratification; another is -- next slide, please -- individual-level patient stratification. I think David alluded to some of these.

Next slide, please.

The issue of whether the methodology should be different or the same, depending on the purpose; whether for payment should be the same; for reporting, whether it should or should not be.

Next slide, please.

Whether the type of payment should be different; for example, pay for improvement rather than pay for performance.

Next slide, please.

Minimize concentration of disadvantaged.

Next slide, please.

Improve adjustment for disease severity, which accounts for some of the variance in SES.

Next slide, please.

And beginning to develop intermediate measures for access and adherence barriers.

Next slide, please.

And ultimately, at the back-end -- and we are not going to focus on this explicitly -- is improve equity in payment. That is actually leveling the playing field.

Next slide.

So, I thought I would put this out there, just as background for people to think about some of the potential pathways, and perhaps, as we move on, later on people can add or subtract to some of these pathways. But, hopefully, there is a common language in the potential consequences of going down each path and at least some preliminary thoughts on

them, on where one might seek a "both/and".

MS. PACE: Okay. Thank you very much.

Sorry about the technical glitch there.

What we wanted to do next was to really go through and have all of the expert panel introduce themselves. And as part of that, we will be asking you a couple of things.

One is who you are; you are a physician, an organization. We will be asking you tell us about your disclosure of information that you submitted when you were in the nomination process. And Ann Hammersmith will tell us a little bit more about that.

But, then, we really want you to also -- we have enough time here that we want you to tell us what your perspective is on this issue, not in great detail, but kind of if you have a current thought or position, we

really want to get those out on the table. But, as David said earlier, we know everybody came with thoughts about this, so we might as well share those and be aware of those.

But to ask people, also, to kind of suspend final judgment until we collectively examine the issues and think about potential solutions.

So, before we start, I am going to ask Ann to go through our usual disclosure information.

So, Ann?

MS. HAMMERSMITH: Good morning, everyone.

As Karen said, those of you have been on our committees before are probably familiar with this process, but it is very important. So, we go over it every time, remind you of a few things, and give you some guidelines for the disclosures that we look for you to make this morning.

If you recall, you all filled out

probably the electronic version of our new disclosure-of-interest form, where we ask even more detailed questions. And we thank you for doing that.

We do not expect you this morning to recount your resumes. In fact, we wish you wouldn't because we will be here all day. So, we are only looking for you to disclose things that are relevant to the work of the Committee. So, if you have engaged in research, consulting activities, speaking engagements, if you have received grants, if you have done any advocacy or lobbying work that is connected to the topic before the Committee today and tomorrow, we would look for you to disclose that.

The other thing I want to remind you of is, just because you disclose, it does mean you have a conflict of interest. Part of the point of this process is to get things out on the table, so everybody understands your background, where you are coming from, and

that we are completely open and transparent in our process.

Many times people will say, "I have no conflicts," which is great. You may not, but, then, they seem to think they don't have to disclose anything. And you may not need to disclose anything, but what we don't want you to do is look at your disclosure narrowly as only I have a conflict; I don't have a conflict. We are looking for you to disclose openly.

The other thing I want to remind you of is that you sit as individuals. Sometimes people, entirely well-meaning, will say, "I'm Susie Smith and I represent the American Society of" fill in the blank. And actually, you don't, not on a Committee like this. You sit as individuals. You're here because you're expert. So, you don't represent the interest of your employer. You don't represent the interest of anyone who may have nominated you to serve on the Committee.

And then, finally, we're not just interested in things that you do where money has changed hands. People will often say, entirely in good faith, "I have no financial disclosures" or "I have no financial conflicts," which is great. But because of the nature of the work we do, we are also interested in your disclosure, if applicable, of things that you did where no money may have changed hands. You may have done it as a volunteer. You may have sat on a committee. And all those types of things are relevant to the work that you will do on the Committee.

So, that is my two-minute summary of the conflict-of-interest process. Again, I want to remind you that you only need to disclose things that are relevant to the work that the Committee will be doing.

And, Dr. Nerentz, did you want to add anything before we go around? Okay.

CO-CHAIR NERENZ: Well, again, good morning.

I'm David Nerenz. I'm from the Henry Ford Health System in Detroit. I have a couple of job titles there, but the one relevant is I'm Director of our Center for Health Policy and Health Services Research.

I have spent much time in the last 20 years working on issues of healthcare disparities, particularly racial and ethnic disparities, in the context of managed care plans, basically, seeking to reduce and eliminate disparities in HEDIS measures. I think I've got a strong record, concern about disparities and advocacy for work against them.

And from that platform, I am very concerned about the possible effects of risk adjustment, of SES adjustment, on disparities. But I am concerned about the effect of the absence of adjustment on disparities. So, both of them are present in my mind through the past that Kevin described.

Our organization has hospital,

medical group, health plan, home health, other components. There is probably not a performance measure on earth that does not affect that organization in some way. And so, I at least have exposure to a lot of the domain.

I am also a MedPAC Commissioner and was part of the approval process for the MedPAC recommendation about stratification of the hospital readmission measure that is part of our background reading material. So, at least in that sense, I am formally on record in favor of that particular approach in that particular measure.

CO-CHAIR FISCELLA: Good morning.

I'm Kevin Fiscella. I'm a family physician/researcher at the University of Rochester. Actually, all my clinical work has been in federally-qualified health centers. My research has largely focused on healthcare disparities and strategies to mitigate them.

In terms of conflict of interest,

I mentioned, possibly related, I am on the National Commission for Correctional Healthcare -- I'm on the Board of Directors -- because they do certify the healthcare quality in jails and prisons.

I can't think of any other conflicts of interest.

Oh, yes, as you can tell from the slide, I am very torn. Like David, I am concerned about both potential pathways.

I am really excited that we have really such an esteemed panel and really smart, creative people here to come to help resolve this dilemma.

MS. PACE: Before we move on to the panel here, I just want to check in. We have two people that are on the line, to make sure they are there.

Marshall Chin, are you on the line?

MEMBER CHIN: Yes, thank you.

MS. PACE: And why don't you go

ahead and introduce yourself? And then, we will go to Mary Beth.

MEMBER CHIN: My name is Marshall Chin. I'm a general internist at the University of Chicago, a health services researcher. I do mostly disparities research in the safety net.

One of our grants that just ended came from the Merck Foundation, which is a philanthropic foundation funded by the Merck company on business disparities. As part of that, we made some Hill visits to congressional staff and Members, educating multi-business disparities.

Some of my collaborative work involves groups that will be affected by, essentially, risk adjustment, including the National Association of Community Health Centers and America's Essential Hospitals. At the University of Chicago, about our third of our patients are Medicaid patients.

Sorry I can't be there. The D.C.

airport is closed because of fog last night. So, we were rerouted back to Chicago. But I am looking forward to participating.

MS. PACE: And do you want to share anything about your current perspective about adjusting for sociodemographic factors?

MEMBER CHIN: Yes. Thank you for that.

First, I thought Kevin's summary was brilliant. It will cover, I think, the vast majority of issues we are going to discuss over the next couple of days.

One thing is that I'm the disparities representative on the NQF Measures Applications Partnership Coordinating Committee, which is one of the umbrella organizations within NQF. One of my concerns throughout has been that, especially for disparities, the issues that are importantly potentially go beyond -- I know David and Kevin mentioned what is the scope of NQF. So, in particular, many of the issues I think that

Kevin put on the slide really start getting at the implementation end of things, as opposed to purely measure selection or risk adjustment per se.

And so, I guess one of my concerns going into the meeting is that we truly address the important issues, because I think that some of them will start getting into this gray zone about what/where within our scope. But, unless we do address them in some way, the Committee's work really won't be fruitful. And so, I do think we need to explicitly think about how can we make sure that we do address the issues that are important within the constraints that we have.

MS. PACE: Okay. Thank you.

And, Mary Beth Callahan, are you on?

MEMBER CALLAHAN: Yes, I am. Can you hear me?

MS. PACE: Yes.

MEMBER CALLAHAN: Good, good.

I am a nephrology social worker and have worked in dialysis and transplant for about 30 years. I guess in terms of disclosure, nothing financial, but I have worked with the National Kidney Foundation and Counsel of Nephrology Social Workers in terms of their KDOQI guidelines, two of them, in particular.

I sat on the panel for preemptive transplant. And in that regard, there are certainly some considered financial disparities in who can get on a transplant waiting list because, within insurance, of course, a person can't get on a transplant waiting list. And the other KDOQI guideline would be hypertensive and anti-hypertensive agents in chronic kidney disease.

And then, also, I worked with a task force of the Texas Medical Foundation on a Dallas Advisory Group to improve kidney testing among African-Americans with diabetes. And the idea there was to improve

microalbuminuria testing for patients identified in certain zip codes that were thought to not be able to get testing who had diabetes.

Additionally, I work with the Society for Transplant Social Workers and Public Policy, and have worked with public policy with the National Kidney Foundation, so that as a disclosure.

From a social work perspective, I think a couple of things. Of course, we all probably know that, when insurance is limited, whether that be because of lack of insurance or because premiums are too high or because deductibles are too high -- last year I had someone with a $15,000 deductible -- decisions are going to be made between healthcare and food or medicines. And so, that is just out there.

Sometimes it has been my experience that somebody who has Medicare and Medicaid, whether that is Medicaid Q and B or

regular Medicaid, has better access to healthcare than someone with an employer group health plan, because with Medicare Q and B everything is paid for. Now there could be issues with access in terms of who will accept that, but I just wanted to get that out there in terms of my thinking.

And then, the last thing is that I think that, when there are multiple socioeconomic risk factors, cognitive capacity becomes a more significant issue in managing the whole situation.

There is one more disclosure. I wrote a chapter in a book titled, Kidney Transplantation: A Guide to the Care of the Kidney Transplant Risk Event, and the chapter title was "Socioeconomic Issues and the Transplant Recipient". And I wrote that with Dr. Connie Davis.

So, thank you.

MS. PACE: Okay. Thank you.

So, now we can go around the room,

and maybe, Nancy?

MEMBER GARRETT: Good morning, everybody.

I'm Nancy Garrett. I'm the Chief Analytics Officer at Hennepin County Medical Center, which is a safety-net care provider in Minnesota. So, Hennepin County that Minneapolis is located in. And so, I lead analytics and information technology at Hennepin County Medical Center.

I don't have any disclosures.

And in terms of my thinking on th is, it has kind of evolved over my care, I would say. So, a lot of my career has been at health plans. I'm a sociologist and demographer by training. And so, like David, I have worked on issues of disparities for a lot of my career. I helped create a system for collecting race and ethnicity data at HealthPartners, which is a provider and a health plan, and also worked at a health plan to try to start collecting that data as a

first step toward starting to address disparities.

And I served on the Board of Minnesota Community Measurement, where we also set up systems for developing quality and cost measures in the State. And I was really concerned at that point of masking disparities, as Kevin so nicely outlined those risks.

And now, with this new perspective of working with a safety-net population, I am really concerned if we don't start to address it. And one of the reasons is because of the fact that payments are increasingly being tied to performance. And I feel that if we don't address it, we are really missing an opportunity.

And just having interacted with our patient population and really seeing examples of a patient, for example, who was in our hospital recently with community-acquired pneumonia and was medically stable, ready to

be discharged, but we were going to be discharging him to a homeless shelter and it was January, and they kick them out in the middle of the day. And so, he would be wandering around and it's 10 below. Is that a good medical decision?

And so, how do we get the resources to be able to manage those social conditions that are so interrelated? So, I am concerned that, if we don't address this issue, we are not going to be able to take care of the populations adequately.

I think the one additional thought I will throw in about the way we have structured the day is, you know, specifically, we said our scope is not to talk about payments and capitation. But the thing is that the cost and resource use measures are so related to that.

So, in Minnesota the cost and resource use measures that NQF endorses are actually used in shared savings programs, and

they are directly connected. So, it is neat to kind of sort of divide that, but I don't know if we really can because that is all part of what we are doing here.

So, thank you.

MEMBER SUGG: Hi. I'm Nancy Sugg. I am an Associate Professor of Medicine at the University of Washington and I am a primary care internist working at Harborview Medical Center. I am the Medical Director for the homeless programs there.

So, the first disclosure is I work for a State university. And then, I also sat on the Seattle Council to look at outcomes for grants done by Seattle specifically to the safety-net population.

So, a few thoughts. One is, when I look at disparities, I definitely like to look at non-adjusted data because I think it does shine light on important things. But when I look at quality measures, I really want to have socioeconomic adjustments made to

that.

I think it is really difficult to look at quality and not really be able to compare apples to apples. When I think about looking at quality, though, I want to make sure that we know the measures are not going to be perfect, but I think they can be somewhat accurate. But I think it is very important to be able to say why they will or will not be accurate.

So, I think of it much like why I order a lab test. I know that there are certain things that will make it a false-positive and there are certain things that will make it a false-negative. And so, I would like downstream of this process to be able to make sure that, when we come up with measures, that we are able to clearly say these are things that you can use these measures for and these are things that you cannot use these measures for, and these are the limitations and this is why.

And that will make me feel better that downstream, when these things are looked at for things like pay for performance or physician quality program records, where they actually put you on the web and say this is your quality performance, that we can really say, no, that it is not a legitimate use of this measure or, yes, this is a legitimate use of this measure.

Thank you.

MEMBER BHAREL: Hi. Good morning, everyone.

My name is Monica Bharel. I'm the Chief Medical Officer at Boston Healthcare for the Homeless Program in Boston, Mass, and Nancy and I don't know each other, but I think our souls must be connected because I happened to sit right next to her.

(Laughter.)

So, disclosures, so that I don't forget them. I serve on a couple of subcommittees that might be relevant,

including the National Healthcare for the Homeless Council of Policy and Clinical Committee on the Mass League of Community Health Centers Policy and Clinical Committee.

And I also, through my organization, do lobbying work at the Massachusetts Medicaid level on risk adjustment for homeless as a risk indicator.

In terms of my association, so I am at Boston Healthcare for the Homeless. My academic associations are at Mass General Hospital, HMS, HSPH, and Boston Medical Center. I am a primary care generalist by training and have worked extensively in different safety-net groups.

As for the question at hand, I do believe we should adjust quality metrics for SES. And additionally, I believe that we should look outside of traditional realms of how SES is defined, at indicators such as homeless status. However, I do believe that I think we could spend the entire time looking

at Kevin's slide, which really got to the meat of the information. I do believe it's the devil is in the details of how including things like organizational stratification and pay for improvement are critical.

Thank you.

MEMBER GROVER: Good morning, everyone.

I am Atul Grover. I am the Chief Public Policy Officer at the Association of American Medical Colleges and, quite frankly, familiar with most of the people and institutions that have already spoken because they're all our members, whether it is Henry Ford or University of Washington or Boston Medical Center. And so, these are the issues we hear about all the time.

I am, by background, a general internist and health services researcher. And certainly, from my own clinical experience, I know how frustrating it is, inpatient or outpatient, to have to deal with factors that

largely are beyond your control.

And I think, you know, my own sort of personal perspective, having lobbied on the ACA on behalf of med schools and teaching hospitals, who, by the way, do take a disproportionate share of the care of the underserved, both the charity care as well as Medicaid patients and medically-complex, is that it was really remarkable that we spent so much time thinking about how to push all the levers on the hospital to improve care and adjust for quality and outcomes, when so much of the care that we are looking to really improve and that is actually delivered in the community is outside those four walls.

And so, how do we come up with measures and metrics that adjust for what happens when that patient leaves, whether they are going to a homeless shelter or whether they have a dual diagnosis or other factors?

And we also spend a lot of time thinking about how we improve disparities,

certainly have a long history in many of our institutions of distrust by the communities that they serve. And how do we improve those relationships and improve outcomes for those communities?

But, again, I just sort of generally feel at my core that we need to find ways to at least adjust in some way without entrenching those disparities. And I think everyone has really reinforced what Kevin laid out for us; find ways to level the playing field while at the same time not keeping our eye off the ball of saying, you know, we don't want to really relegate anyone to the dust bin of the second or third tier of healthcare. And I don't come in with any answers of how to do that, but I do think it is very important we try to do our best to find a way.

MEMBER LIPSTEIN: Good morning.

My name is Steve Lipstein. I'm the President and Chief Executive of BJC HealthCare in St. Louis.

I'm a little bit intimidated in this audience because I'm not a doctor, I'm not a researcher, and I don't have any academic rank.

(Laughter.)

So, I'm sitting here, as we started the introductions, thinking, why am I here? And so, my disclosures are that I do serve on the Boards of Trustees at Emory University and Washington University. I serve on the Board of Governors of the Patient-Centered Outcomes Research Institute, and I think there's a PCORI-funded investigator somewhere in the room.

And I'm a previous Director of the St. Louis Federal Reserve Bank, which probably has some influence on why I am here because what interlaces with this whole issue of socioeconomic status risk adjustment is human behavioral economics and how people are going to respond to the presence or absence of risk that is or is not adjusted for.

So, those are my disclosures.

I think what gives me keen topical interest, as the head of a health system, is that with BJC HealthCare we have 12 hospitals, and two of those hospitals are located in rural communities, one of which is a critical access hospital. Four of those hospitals are located in suburbia. Three of those hospitals are large community medical centers, one in a very affluent section of St. Louis and one in a very -- I don't know if this is the right word, David -- unaffluent, inaffluent --

CO-CHAIR NERENZ: We get the point.

(Laughter.)

MEMBER LIPSTEIN: You get the point. And then, two of these hospitals are big teaching hospitals, and the rest are non-teaching hospitals. So, think large and small, teaching and non-teaching, urban and rural, pediatric and adult.

And as I was looking at Kevin's

diagram, the socioeconomic status, risk adjustment, yes or no, the yes pathway is a hypothetical; the no is what I'm living. Okay?

And since the no is what I'm living, and we study outcomes across all 12 of our hospitals in great detail, using health information management databases, which we used to call medical records, as opposed to claims datasets or the Medicare professional analysis review MEDPAR, we have come to learn that what you know through studying patients in MEDPAR doesn't give you a complete picture of the patient. You are working with incomplete data.

And when you work with incomplete data, you don't get an incomplete answer. In our view, you get the wrong answer. So, if you take three plus five plus "X" and you say, well, the answer is eight and what we don't know -- we know the answer is not eight. And so, trying to work with more complete data to

understand outcomes has become kind of a passion of ours.

And so, I will share with you and disclose that two of our twelve hospitals pay pretty significant penalties for high readmission rates and ten do not. And we spend a lot more money on trying to prevent avoidable readmissions at those two than we do at the other ten.

So, as you know, many hospitals are out there hovering for 30 days post-discharge over patients, and we are trying to understand. And so, one of the things that is interesting to us a little bit is that, if we don't know what happened to the patient, meaning they weren't readmitted within 30 days, if we don't know what happened to the patient, we think that is a good outcome. But if we do know what happened, they were readmitted, or we do know other things about them, we are not sure that is a good outcome.

And so, just to kind of finish

this up, since I know you don't want me to go on too long, one of the things we know about our 12 hospitals is that they are located in very geography, as described in empirical data by Census-tracked information.

And so, what we have learned is that the individual income of the patient isn't determinative of many outcomes, but where they live has great influence. And so, if you live in a community that doesn't have grocery stores or doesn't have drug stores or doesn't have laundromats or doesn't have taxicab stands or public transportation, that that plays a significant role in the outcome that happens after a patient is discharged from the hospital.

And so, we serve all those different kinds of communities, and we can compare with BJC what happens from an outcomes perspective. And then, what we try to figure out is, okay, once we know what happens from an outcomes perspective within BJC, what

happens across our State, what happens across our country.

But one of my aha experiences in all this was when I joined the Patient-Centered Outcomes Research Institute Board, at my first dinner I sat next to Harlan Krumholz, who is also on that Board, who actually knows a little bit about readmission rates and a little bit about socioeconomic adjustment, and kind of has helped me to understand that this is a much more complicated topic than just where your hospital is located and the patients that it takes care of.

But suffice to say that the perspective I hope I bring to this is that we know that where those hospitals are located and the resources that they have available to manage patients who have difficult life circumstances is highly influential on the patient's outcome.

And when you write public policy at the national level and you think East St.

Louis is the same as Chesterfield, or you think Detroit is the same as Scottsdale, Arizona, the public policy implications are pretty significant. And because we have linked that to pay for performance, there is now federal funds flow, in my view, leaving Detroit and going to Scottsdale because of absence of socioeconomic risk adjustment.

MEMBER SAWHNEY: It's hard to follow up on that.

(Laughter.)

I'm Tia Sawhney. I am the Director of Data Analytics and Research for the Illinois Department of Healthcare and Family Services, which is a long way of saying the Illinois Medicaid plan.

I'm a qualified health insurance actuary, a Fellow of the Society of Actuaries, and a member of the American Academy of Actuaries. I am active in both organizations. And those are long-term credentials.

In the more recent-term, I got a

doctorate degree in public health, and my dissertation was at the divergence of insurance and public health, specifically, risk selection by both insurers and providers and risk adjustment under the ACA. So, I spend a lot of time thinking through the issues, but more from a commercial insurer perspective.

I am a data person. So, what I will probably do from time to time in this conversation is ask things like, "Yes, but how do we do it, and how do we do it reliably, and how do we make the math work?" And some of you may want to throw your shoe at me, and that's okay. And it is not that I'm not sympathetic to the larger social goals, because I'm all for them, but how do you make it happen?

MEMBER COHEN: I'm Mark Cohen. I am a statistician. I manage the Cisco group for ACS NSQIP, which is the American College of Surgeons and their Surgical Quality

Improvement Program.

Until I joined this Committee and read the materials, I didn't know how easy I had it.

We not only are involved, essentially, in reporting, but it is private reporting. So, the consequences, there's no pay-for-performance compensation, no consequences for being held up to public scrutiny.

Our models are really pretty successful. We have maybe a hundred few models. We have 40 predictors usually available for selection, a lot of overlap. We include race and ethnicity, but because our variables are very correlated, they rarely are very powerful.

But after hearing this discussion of when you move from the position of private reporting to the position of pay for performance and public reporting, my sympathies are towards making the adjustment

as being essential.

But I am also a data person, and looking at that graph, there are a lot of questions about how you implement this and how you make the distinction.

I was also struck by the fact -- one thing clear in reading the documents was the issue about that NQF measures have to be used for both quality and accountability, which seems to say that they can only be used for accountability.

We are very successful in quality improvement, even though we may not attend to income disparities. For our purposes, it serves very well.

So, I just wanted some clarification about, is that really essential, that you can't get NQF endorsement unless the measure is intended for both purposes? Okay.

Thank you.

MS. HAMMERSMITH: Do you have any disclosures that you would like to make? You

don't have to if you don't have any.

MEMBER COHEN: I don't believe I have any disclosures to make.

MEMBER CASALINO: Larry Casalino. I'm the Chief of the Division of Healthcare Policy and Economics at Weill Cornell Medical College. Before that, I was at the University of Chicago. And before that, I worked for 20 years as a family physician in a small private practice in Half Moon Bay, California. So, it is on the coast just south of San Francisco.

Disclosures: let's see, these are all unpaid. I'm on the Board of Directors of the American Medical Group Association Foundation, of the American Hospital Association. I'm a member of the American Hospital Association Committee on Research, and I'm a member of the Board of Healthcare Research and Education Trust, which is closely aligned to the AHA.

I have also done a fair amount of research and some speaking related to the

STARS, some of that research actually with Alyna and Marshall, who are both members of the Committee.

First of all, fantastic materials for the meeting. Really, for David and Kevin and the NQF staff, I don't think I have ever been to a meeting that the materials were better for it, and not just that they were a very complete references in the articles handed to us, but also just the thoughtfulness. I mean, each sentence in the prep materials is very carefully crafted and really addresses the issues, and it lays them out in a thoughtful way.

In terms of my perspective, one quick point. I think somewhere in the materials it does mention that the usual view of process measures is that they don't have to be adjusted for anything really. And that I believe is true for inpatient measures, such as preventing central line infections. Or I guess that's not really a process measure.

Better would be counseling before people are discharged.

But in my mind that is just a silly statement, and I can't believe that it still is the norm in relation to outpatient measures. In our practice in Half Moon Bay, because of our location and the fact that we were the only game in town, we had Silicon Valley executives and we have farm workers and everybody in between. And believe me, it's easier to get a high mammography rate for Silicon Valley executives than it is for farm workers, right, who have almost no -- many reasons; I don't need to go into the details.

The other perspective I'll mention, just to summarize that process measure, I think if the process measure depends on the patient doing something, which the inpatient measures don't, then you have to think about whether it has to be adjusted. I think that should be the maximum.

In terms of the overall issue,

though, I think that both points of view, the concerns for and against doing some kind of adjustment for SES are valid, right? So, we are not going to find some pure principle that requires no compromise that we can just move forward on.

You know, the philosopher Isaiah Berlin, his whole life really was spent saying you cannot create a Utopian society. There will be conflicts among equally-valid principles and people of equal goodwill, and you need to have a system that can accommodate that. So, he is very anti-Utopian ideologies. And I think we will need compromise as well.

So, in this context I think it is important -- and I won't go on much longer -- I think it is important to differentiate among accountability applications. So, one might want to do things differently, for example, for pay for performance and for public reporting.

So, for public reporting, my view

at this point -- and I am, I think, quite open to change, based on what happens these two days -- is it is complicated, but I would report both unadjusted and probably stratified results. I would report them both. There's no particular reason that that can't be done. But I would report both by type of organization. So, X kind of organization compared to other X kind of organizations, but also for all organizations how they do for different types of patients. So, that is what I would do for public reporting.

For pay for performance, you know, we can't report it both ways. So, there I do think -- and I'm sure others agree with some version of this -- that we need some kind of blended measures. My thinking now is I probably would pay based on a blend of unadjusted and stratified performance and, also, pay on a blend of ABSTA scores and improvement in scores. And that would go far to deal with the issues that we are talking

about.

The problem is, I think it is two problems. One quick one, and then, I am done, is that the stratification is nice to talk about, but in many cases, except for the very largest organizations, the "N" may be too small for some of the cells, for critical cells. I would be interested to hear people's ideas about what to do about that.

And the other point is I think -- and Marshall alluded to this, I believe -- it is one thing to identify an SES measure and say we think it ought to be used. It is another thing to talk about how it ought to be used.

I do think that the comments so far and in the materials, and I think in a lot of the rest of the meeting, are really necessarily about how the measure ought to be used. The reasons why are obvious, I think.

So, I'm not sure if I am correct about this, but I seem to remember from the

conference call and from some comments in our packet that this goes beyond what NQF usually does, is to talk about how things ought to be done. And so, this is a problem I think we may come up against again and again.

MEMBER ADAMS: Hi. Alyce Adams. I am a research scientist as well as Chief of Healthcare Delivery and Policy at Kaiser Permanente's Division of Research in Oakland, California.

In terms of disclosures, I'm also on the Kaiser Permanente Work Group on Healthcare Disparities. It is not a decisionmaking group, but we are grappling with these issues every day, about reporting, what do we report, how much do we trust the data.

And so, I am particularly interested in this issue of data quality because it is a big problem. We've just gotten to the place where we feel like we trust our race/ethnicity data. So, SES data

is another realm for us, and I am sure other sort of healthcare organizations, as we try to assess whether or not to adjust.

We have talked a lot about adjusting in terms of the process measures as well as outcome measures and what that might mean. And I concur with what has been said so far, both in terms of the materials that we were given and the conceptualization we got this morning are really spot-on. I don't know the answer. I don't feel like I have a specific preference to go in either direction. But a lot of our work does deal in unintended consequences.

And I think it is incredibly important to look at it through that lens because anything we do is not going to be perfect, but as long as we can talk about the consequences of each of our choices and the reasoning behind that, I think that is going to go a long way in terms of helping us take that next big leap to this question of

adjustment.

MEMBER BARGER: Hi. I'm Mary Barger. I am an Associate Professor at the University of San Diego. I am a certified nurse midwife and perinatal epidemiologist.

My disclosures are I do research on racial disparities related to perinatal outcomes. I have had grants related to that. I was one of the representatives for the American College of Nurse Midwives to the first Healthy People Work Group, where we published target measures by race. And then, we felt, oh, that was a big mistake. And then, we said, no, there should just be one target measure.

So, having taught in a school of public health, where I think -- you know, I pointed out to my students that racial disparities are important; if we don't collect data on it, it will be put under the table.

I have feelings on both sides of the coin, having worked in community health

clinics as a nurse midwife, realizing that a lot of the things we do are completely outside our realm as healthcare providers, especially if we look at the life course perspective. You know, we can't go back to things that happened at birth to people and change that.

And so, that's a concern. I think I am leaning towards that, yes, we should have some adjustment for quality measures because of that. However, as an epidemiologist, I worry about these are really, really rough measures for whatever proxy we think it means. And so, when you adjust, are you just putting in more confusion. Instead of getting closer to the truth, are you get farther away from the truth? And there are certainly theoretical models in epi that show that, if you do that, you get further from truth than closer to the truth when you use a rough proxy measure.

So, I have a concern about that, in that large variability is a way for us to

highlight opportunities to improve care. So, that is sort of where I am at the moment.

CO-CHAIR NERENZ: Karen, if I can just make one pace observation --

MS. PACE: Yes.

CO-CHAIR NERENZ: -- as we switch to the other side of the table?

In principle, on a minutes-per-person, we are running a little behind, except I think people have done a wonderful, wonderful job of speaking clearly and concisely. And I think this has been really good so far.

Am I correct we do not have a CMS presentation this morning?

MS. PACE: Right.

CO-CHAIR NERENZ: So, we can let this run a bit longer than the agenda says?

MS. PACE: We have until 10:20.

CO-CHAIR NERENZ: Fine. Good.

MS. PACE: So, we have plenty of time.

And I should have mentioned that Kate Goodrich from CMS is planning to join us this afternoon.

But we have until 10:20. So, we're good.

CO-CHAIR NERENZ: We're doing fine? Okay. Good.

MS. PACE: We're good.

MEMBER ACCIUS: Jean Accius. I work with AARP. I am the Director of Health and Long-Term Service and Supports. In that capacity, I work with our National Policy Council, which is a group that advises the Board of Directors on health and economic and consumer issues.

I am extremely interested in this topic. In fact, I have done work looking at racial disparities, particularly in access to care among Medicare beneficiaries, as well as doing some work around the implications and the variation across racial groups as it relates to retirement decisions within the

context of Social Security and potential reforms to the program.

So, I come to this from the perspective of, how do we, as Steven articulated earlier, look at policy implications, both at the national level, but also at the local level, to fully understand how can we incentivize behaviors regardless of what the unit of analysis is.

From the perspective of AARP, clearly, there is a great deal of interest from the perspective of a life course perspective, that the disparities that we see in old age just did not trigger at the age of 50 or 65 or 62, whatever marker you define as old, but that it had pretty much life course implications over time.

So, that being said, my goal here is to really kind of raise some of the questions around what are the policy implications and how do we really incentivize behavior in a meaningful way, to really try to

reduce the disparities that we see across a lifespan.

MEMBER O'BRIEN: Good morning, everyone.

I'm Sean O'Brien. I'm a statistician at Duke University. Part of the work I do there involves development and evaluation of performance measures, in particular, working with cardiovascular registries, national clinical registries.

So, in terms of conflicts of interest, I have been involved with several NQF measure submissions, some that are currently in the pipeline, especially with the Society of Thoracic Surgeons. They have three databases that I'll submit measures to NQF.

I was recently involved in development of a 30-day readmission measure that was contracted by CMS that was for patients undergoing bypass surgery, and a couple of other miscellaneous measurement projects.

So, the issues of how to address socioeconomic factors and case mix adjustment have come up in basically every project I have been involved with that involved NQF measure submissions, and I was also in other projects where I have served on NQF panels.

So, in terms of my perspective, as a statistician, I think I usually let other people do more of the setting the agenda and saying what questions should we be attempting to answer, and ask and answer with data. I usually limit myself somewhat to basically, given some particular set of objectives, what's methodologically the appropriate way to address those. And I am willing to make a statement about what we should or shouldn't do, you know, a blanket statement. But I am more a stickler about, given this particular set of objectives, how should that be done?

So, currently, when I think about performance measures, the things you showed us distinguished a couple of different

perspectives. And one is they are used to incentivize, I think, behavior and effecting behavior change. And the other is basically, when you are reporting data, you're trying to answer questions with the data. You're trying to answer maybe a "What if?" question. What would outcomes look like if case mix was different? Or what inferences can we draw about the processes and underlying quality of the providers?

And when you're trying to report measures that have a particular interpretation, and you are also trying to incentivize a certain behavior, you are trying to do two objectives at once. And in my experience, unfortunately, sometimes when you have multiple objectives, it is hard to do either one really, really well.

And when it comes to incentivizing behavior change, I don't think it is a requirement that measures need to be valid or have any particular interpretation for them to

be successful at incentivizing behavior.

So, in a P4P context, you could reward units that have good outcomes, even very small numbers of patients, and maybe too small to get reliable estimates of some underlying performance. So, it may be driven by chance, but that still can potentially drive people to improve. And in sports competitions there's a lot of random variability and people get very motivated in those contexts.

So, I don't think validity is necessarily a requirement, but what I am kind of stickler for is that, if people are interpreting measures in a certain way and they are going to draw inferences, I think the methods need to support those inferences. And so, when you don't adjust for certain variables, I think it is relatively hard to say what the correct interpretation is. And that is a problem for me.

I have heard a few things repeated

that I have never been able to quite wrap my mind around as a statistician. And one is that we shouldn't risk-adjust for associated factors; instead, stratify. But, in my mind, stratification really is a form of risk adjustment, and they can have some of the same problems in terms of reducing incentives to improve or masking disparities.

So, I think, well, what do you mean by stratifying? Typically, we're starting with a performance measure that is trying to measure and compare the performance of different units, such as hospitals or physicians. When we say, "Well, we're going to stratify now," what are we talking about? Are we talking about, within different subgroups of patients, compare outcomes of different units within these subgroups of patients?

So, for example, it may be useful to compare how hospitals do among lower socioeconomic status patients. Well, having

adjustment in your risk models doesn't prevent you from doing that type of comparison. There's actually no problem with that.

If you are interested at a population level being able to look to see, well, which groups of patients have better outcomes/worst outcomes, you get that exact type of inference from having those factors in the risk model. So, at the point in time when the risk model is developed at a population level, you can actually get a good insight in terms of disparities by putting these variables in models and seeing what their effect is after adjusting for other models.

And if you are interested in kind of comparing how disparities change over time, not at the point the risk model is developed, there are approaches you can use and still use these risk models that include adjustment for socioeconomic factors and still make the types of comparisons you're interested in.

For example, when you comparing

across groups of patients, if your model adjusts for that factor that you're interested in comparing, you are adding something to the model in the denominator of your observed expected ratio at the same time you are trying to compare differences across observed, and you subtract out the effect of interest.

What we can do is you can evaluate your risk model algorithm as if all patients in your population you're interested in studying, as if they all have the same socioeconomic factors, and just arbitrarily treat all patients as if they were the most common race or the most common socioeconomic status. And then, you can still apply that model. So, a lot of the recommendations I hear, when I think about them, they are not very necessary.

And finally, in terms of the concern about taking away the incentive to perform, I just think the incentives need to be designed and addressed explicitly. And

right now, the measures people pay attention to these overall global summaries of performance where basically you're measuring a hospital or a unit's performance. Implicitly, it is weighting their outcomes across all the different subgroups of patients in proportion to their prevalence of those populations in the population.

And instead, you could provide measures that upweight different groups of patients or you could report outcomes separately for different groups of patients, which is stratification, and explicitly build those incentives into the measurement process.

So, those are some of my perspectives.

MEMBER JIMENEZ: Hi. I'm Dionne Jimenez. And I guess my disclosure is I'm employed by the Service Employees International Union. So, we're the largest union of healthcare workers, representing over a million healthcare workers.

I mean, it is great to be part of such a wonderful group with so many areas of expertise. Just hearing Sean speak, I'm like, "Wow." I should have paid more attention in my statistics classes.

One thing I wanted to say from my own personal -- so, I am a Research and Policy Coordinator for the Union. And so, in that role, basically, I help inform our leaders in terms of what positions we should advocate with both the state and the federal levels regarding various healthcare policy issues.

So, from my personal perspective, I think that it is very important to address healthcare disparities. But, as we are seeing the results of the first few years of the accountability applications, especially in the Hospital Value-Based Purchasing Program, which I spend a lot of time looking at, and the Hospital Readmissions Reduction Program, you know, I think there is definitely a pressing need to address for SES, because we are not

necessarily eliminating disparities. It is actually exacerbating them, especially for seeking the institutions.

And I think when we are looking at real-world consequences, you know, you could see the provider perspective, but we also have to think about the people who are working in the hospitals and the institutions. I mean, anecdotally, we are seeing our represented providers starting to come to us in negotiations, thinking about, well, where are we going to have start making cuts, when the purpose of these programs is actually to improve quality.

And you have to remember that a lot of these workers are actually in lower SES categories, too. So, I want to keep in mind.

But I think from the research perspective, I think I agree, though, it is very important that there needs to be both the raw and the unadjusted as well as the adjusted data. So, all of our work could be used. But

when it comes to the implementation programs, adjustment is very important.

So, that's it.

MEMBER PONCE: Good morning.

I'm Ninez Ponce. I'm a Professor in the Department of Health Policy and Management at the UCLA Fielding School of Public Health.

My disclosures are I'm a health services researcher with economics training, and I am a disparities researcher. I am also on the Board of the National Health Law Program, which champions the rights for the low-income population, particularly those on Medicaid.

I was nominated here by the California Pan-Ethic Health Network, which is an organization of several racial/ethnic disparities organizations.

I guess my research disclosure is I lead the California Health Interview Survey, which is a population-based survey. So, I

definitely have a bias towards population-based measures and social determinants. I recently a chapter on multi-level social determinants of health.

So, my perspective on adjustment -- I also teach econometrics, and I have a fear of meta-variables. So, not having social determinants of health as structural indicators of the complexity of the patient is really important.

And some of the articles that I read said disease severity helped, but there was one article that swamped out income which was really disconcerting for me. So, I think if we are going to use social determinants of health or SES as a way to adjust, then it really is trying to get at incentives and rewards for providers. So, not to identify those who avoid sick patients, but to reward those who seek sick patients and complexity and the structural determinants. So, that is my perspective.

I also work very closely with Thu. I think that has to be disclosed. We have worked on a project on social determinants of health and risk adjustment using her clinical data. I have worked with her also in a HRSA-funded project called -- and Kevin is on it, too -- on Community Health Centers. And I have just agreed to be on a panel for the National Association of Community Health Centers -- and Michelle is here -- on social determinants of health and risk adjustment.

MEMBER QUACH: Hi. I'm Thu Quach. I am the Research Director for Asian Health Services, which is a Community Health Center in Oakland Chinatown, California. We serve about 24,000 patients. Most of them are Asian-Americans and Pacific Islanders. The majority are immigrants. And so, we are really big on the issue of language and culturally-competent services.

And so, my disclosure is that I'm

also a research scientist with the Cancer Prevention Institute of California, where I am funded to do a lot of different environmental health research. I'm an environmental epidemiologist. A lot of my research is focused on Vietnamese nail salon workers and their disproportionate exposures in the workplace.

My big disclosure is that I do a lot of work around environmental justice, both in my professional and personal affiliations. I am involved with a lot of advocacy organizations that promote health and equality for Asian-Americans and Pacific Islanders. I'm really big on immigrant experience and how to get at that, and sort of the complexity of patients who are immigrants and sort of some of the social barriers and the cultural stress that they face when resettling here. I think Ninez mentioned some of my other disclosures.

In terms of my perspective, as an epidemiologist, I do believe that you have to

adjust. I think the big issue is how well the data captures the complexity of the patients and the communities that we served.

In a lot of my work, I do a lot of community-based participatory research. And so, it is hard to say that you don't do advocacy because a lot of your research should inform social change. So, I really believe that a lot of the data that I am collecting, oftentimes with the community, really is informing, should inform more of the health policies. And so, risk adjustment in this work, while I am really new to the healthcare field, I do believe that -- someone said it best -- the devil is in the details in terms of how you capture some of these complexities for communities.

MEMBER NUCCIO: Good morning.

Gene Nuccio, University of Colorado Anschutz Medical Campus, I'm a faculty member there.

In terms of disclosure, I work in

doing home health risk adjustment. That is, the OASIS instrument is the instrument that my Division promulgated and has revised twice. I am personally in charge of doing all the risk models to risk-adjust about 41 healthcare outcomes and about 10 adverse event outcomes for the home health world. These outcomes are recorded both privately by CASPER Reports and on the Home Health Compare site that we also helped design back in 2003.

My work is primarily funded out of CMS and MedPAC. With MedPAC, I helped design some alternative outcome measures and looked at a 30-day both hospitalization and a 30-day rehospitalization measure based on claims data and risk-adjusted through the OASIS instrument.

The OASIS instrument is one that home health agencies use to assess a patient at the beginning of care and at every 60 days, should they be on care that long, and at the end of care, whether that care goes back to

the hospital or if they are discharged to the community.

As such, I'm very much interested in the whole concept of risk adjustment and distinguishing between creating the prediction model that is used to risk-adjust or to predict the outcome and whether or not we include or don't include some measure of sociodemographics in that model, but also how you end up applying that information to adjust.

That is, currently, many of our adjustments are based on the idea of using a national reference to adjust the value. That is, everybody is held to a national standard. And my belief is, why? As many of you have pointed out, there are huge differences between Arizona and Minnesota or South Dakota and Boston. Why don't we use something else, like maybe a CMS regional value or a state value, as our reference point in terms of adjusting?

With regard to, if we decide that it is appropriate in some instances to use SES to adjust an outcome based on accountability issues and quality issues or other perspectives, do we represent that information using a patient-level value, that is, the individual's racial SES, and so on and so forth, or do we use some sort of geo-value, a Census value, for the various population areas that that healthcare provider looks at or typically serves?

So, I mean, you know, all healthcare is local. I mean, you might need to go to Mayo in Minnesota, but, more likely, you're going to go to the hospital that is down the road.

So, those are my particular issues. So, I guess right now my perspective is perhaps not yes or no, but sometimes. And then, the criteria or the important variables is how are we going to use the information and, then, how should we adequately represent

the information.

MEMBER CHIEN: Hi. My name is Alyna Chien. I am an Assistant Professor at Boston Children's Hospital and Harvard Medical School.

The way I usually describe what I do is I work on how we pay doctors, but I think about incentives broadly. There are payment incentives, reputational ones, what your organization is doing to help you.

Clinically, I'm a general pediatrician, and I recognize all of the issues that all of the providers have spoken about. I share all of the concerns that have been raised about data quality and being torn.

I guess where I fall is mainly that, even though we want to stay focused so that we can have a productive conversation, I'm worried about oversimplifying the situation, and that we do need to recognize that how quality measures are used is very complex. And I think that we should use all

the tools that we normally use to understand the situation, as researchers and healthcare providers in the broader quality realm.

MS. HAMMERSMITH: Do you have any disclosures that you would like to make?

MEMBER CHIEN: I mean, just that I make a living doing this.

MS. HAMMERSMITH: Okay. That's good.

(Laughter.)

MEMBER CHIEN: But I honestly wouldn't have gone to medical school or chosen a research path if I wasn't passionately interested in vulnerable populations. So, take that for what it's worth.

MEMBER WERNER: I'm Rachel Werner. I am an Associate Professor of Medicine at the University of Pennsylvania, where I am a health economist and I do research related to the use of quality improvement incentives, specifically financial incentives on healthcare delivery and quality of care. I'm

also a general internist and I have a joint appointment at the Philadelphia VA, where I practice.

For disclosures, you know, I receive grant funding to study these issues and I speak nationally about these issues. But, beyond that, I don't have any other disclosures.

In terms of my perspective, I also am sort of torn. I think that David and Kevin laid out very nicely that there's two potential outcomes from risk-adjusting, however you go about risk-adjusting. I think they are both important.

I think it is, from my perspective, very clear that, when financial incentives are being tied to quality of care, we need to do something to level the playing field, so that providers who disproportionately care for uninsured or low-SES patients are not penalized for that.

But I think that in terms of

measuring quality, it is not so clear to me what to do.

And I want to just sort of lay one thing on the table, which is, as an economist, I'm also scared of metavariables, and I worry a little bit about what we are measuring when we measure socioeconomic status. In Kevin's diagram, he laid out a number of things that lead to low or poor health among low-SES patients or socially-disadvantaged patients. And those are things like early-life factors, epigenetics. There's also things related to access and adherence, language, patient activation.

So, I worry that, by simply adjusting for socioeconomic status, we are really trying to capture all of that information which we don't have data on, and that it may sort of dampen the enthusiasm for directly addressing those things which may improve equity in care. And so, I don't say this to let the perfect be the enemy of the

good, but just to be aware of sort of some of the downsides of trying to adjust for these things.

MEMBER OWENS: Hello. My name is Pam Owens. I am the Scientific Director of the AHRQ Quality Indicators. That's a project that has 92 indicators for the patient safety indicators, the patient quality indicators, prevention quality indicators, and pediatric quality indicators.

The other hat -- I have been at AHRQ for 12 years -- the other hat that I wear at AHRQ is that I am the Coordination of Outpatient Data on the Healthcare Costs and Utilization Project. We have 44 states participating with their discharge data for both the hospital side, inpatient hospital, and then, 33 states participating on the outpatient side.

So, I am saying this in the context I'm a data person, and I have the policy piece as well as the technical

specifications on the quality indicators, which, as you know, many of the quality indicators are being used in CMS programs. And I work closely with CMS as those get translated.

The other things you should know, I do sit on the NQF MAP Hospital Workgroup. I also am going to be sitting on the NQF Population Health Workgroup. I sit on the Interagency Committee from the Department of Health and Human Services on Measurement and Measurement Policy. So, there's a couple of different influences that you may hear that might have subliminally got into me.

From an analytic standpoint, one of the projects that I am a Task Order Officer on is improving the AHRQ Quality Indicators. And we are explicitly looking to see if hospital characteristics should be risk-adjusted for the Quality Indicators to improve their performance. And what I mean by that is we are looking at things like safety net.

And so, the analysis is not there enough, or I'm not confident that we have taken into account the clinical aspects of that analysis enough to bring it to this table, but it is really relevant because we do see differences.

The other piece is I actually started four years ago -- well, for four years, I took a break from AHRQ and I worked at Washington University and with Steve Lipstein at BJC. And we started the discussion around SES and readmissions. Or maybe my views are influenced by you, so there. But, at any rate, so that you have some context of where I am coming from.

In terms of what I'm coming to the table with right now, and it is very important, Ann, to reflect this is my view; this is not AHRQ's view. And you will see all that, all those caveats on all manuscripts, "This is an individual view; it's not the AHRQ view," because I have not vetted it with AHRQ.

So, I do think SES is a complex concept, and it is actually one I started my dissertation on many, many years ago, examining how SES relates to recurrent strokes. And the first part of that dissertation is let's delve into what do we mean by SES.

Now it's interesting because around the table I have heard many people reflect different statements of what SES is to them and what it is reflecting. And I think that's important, that we think about as we move forward, because it is both the concept of what we are trying to reflect as well as to operationalize it.

The other piece that has come up a number of times, and which I agree with, it does depend on its purpose. And I think we need to think both in its purpose from a research perspective and improving the measure and the specificity and sensitivity of the measure, but also from the implementation

perspective.

And these things have legs that you don't realize. So, whatever we recommend, things happen to them, and you should just be aware of that.

In terms of what I am currently sitting with respect to adjustment, I do see a need for unadjusted measures and I see a need for stratified measures. I am less clear how to do risk adjustment in which it is just an indicator, a variable in a model, and everybody is mooshed together.

I feel like we are masking some of those disparities. And as you know, AHRQ does a lot of work on the NHDR, examining those disparities, using the Quality Indicators, and we do use risk-adjusted indicators in them, but we haven't put SES in them. So, there is a lot at stake.

MEMBER BERNHEIM: Hi. I'm Susannah Bernheim. My main job is that I am the Director for Quality Measure Programs at

Yale Center for Outcomes Research and Evaluation. And in that role, I oversee two contracts with CMS where we develop outcome measures. So, we are the developer of the publicly-reported readmission measures and mortality measures that people are aware of. So, we spend a lot of time talking about this, from meetings in-house to dinner table conversations, to everything else.

I am also a family physician. I do my clinical work at a federally-qualified health center. I have always done my clinical work in underserved populations; trained at San Francisco General Hospital.

And I have the research background and the research training, and my research has always focused on the intersection of quality and socioeconomic status. So, I come at this from many perspectives.

And I am going to talk a little bit later just specifically about how we think about it as a measure developer to give that

perspective. But, as other people have said, I do not sit at this table as CMS's representative or Harlan's representative or anyone else's representative. I really do think about this from a lot of different angles, and I agree with everyone that it is complex.

I don't think I have any other conflicts to disclose. But I will say, just in terms of my perspective, I am going to show you some of the analyses we have done and how we have thought about them. I am comfortable with the decisions we have made in the current readmissions measures. They were done in accordance with NQF guidance. But I do not think that this is a one-size-fits-all situation. I think that, as I have thought about how we came to those decisions, I have tried to think a lot about sort of what are the criteria where, given that you are not going to get this right, right -- there's a mix of things going on -- I'm really coming to

this trying to think about how can we, as a group, articulate as clearly as possible what are the circumstances under which the risk/benefit of this kind of risk adjustment goes in one way versus the other.

So, I think that is really our job, to say, you know, what purposes of measures, what kinds of data analyses, what kinds of SES variables that are available that seem like the right ones. When do we have a setup where it is clear that we should do one thing versus the other? And I don't think that is an easy task, but I come to this with a perspective that that is kind of the job we have at hand.

The one other thing I will say -- and it has been echoed by Larry and Rachel and some other people -- is I know that it is hard to differentiate the implementation of a measure from the measure itself, but one thing I will say is that I start, because it is where I started, by thinking about first these

measures as how do we best reveal quality, knowing that this is going to play out in different ways.

And then, what if those measures get used in ways that are going to hurt the safety net? Maybe that gets dealt with in a different place. So, that is a little bit of a bias that I come to, that we should separate how we think about quality measures and implementation, and I know that's not simple. And I think it is a conversation we are going to have lots of times.

But what I have generally argued for is that the policy of how these measures are used should be changed, but not that the measures themselves should change. So, that is probably my predisposition, but I am very open to us talking about sort of criteria and specific measures where we would do things differently.

MEMBER GOLDFIELD: Again, I'm glad that I'm the last one, I guess.

(Laughter.)

Or maybe not. There's more. Or we could start all over again, I guess.

(Laughter.)

My name is Norbert Goldfield, and I'm the Medical Director of the Research Group at 3M Health Information Systems and a clinical internist.

I guess from my perspective, while everybody has emphasized these issues are complex, decisions are continuously made regarding SES payment and quality. So, we just need to acknowledge that upfront, that I think it is real important.

The main reason I'm interested in this is that I am hoping that we can push the process forward. And by that, I mean the pushing the process forward, as Medical Director of the Research Group at 3M Health Information Systems, we do a lot of work with CMS, AHRQ, and MedPAC in approximately 35 states and private insurers.

And by that, I mean specifically create case mix measures. I'm the lead clinical developer of the case mix measures that link payment and outcomes quality for each of the four types of healthcare encounters. And for me, they are ambulatory care; hospital care; year-long, person-based episodes, and long-term care.

As a clinical internist, I see patients two days a week at a health center. I'm particularly interested in programmatic innovation. We were the first site of implementation of a dual-eligible program in Massachusetts, where I live.

Just a couple of observations. I think it is important to distinguish between using SES as an independent variable versus a dependent variable. I think it is important to understand are we talking about confidential disclosure versus public disclosure or used as payment. I consider public disclosure tantamount to payment, for

any number of reasons.

As a consequence, my personal perspective is that -- and I do speak for our Research Group -- is that it is really necessary to have clinically-robust, detailed, severity-adjusted measures. And frankly, most are not.

The devil is certainly into detail. Implementation is key. You can have a great tool, and I would say there's still an 80-percent chance of poor implementation, which creates its own set of issues. That is not really part so much of this group.

I would say that, from a disclosure point of view, that, in essence, when CMS uses our work, it is public domain. When it is not, I have, frankly, the intellectual luxury -- and it is a luxury -- to develop tools that are proprietary that are, then, used in states and overseas by private insurers.

A good example might be New York

and Texas. New York and Texas, as far as I know, are run by different political parties, but have focused on paying for better outcomes.

And in that spirit, I want to say -- and David and Kevin and Suzanne and Karen know -- that I believe that there is a significant error in the charge for this group, which is to say that payment and quality have been separated. And, in fact, everybody, starting with yourself, has outlined, in fact, how the two are not separate. And I think we are being disingenuous when we make that separation.

And certainly, any number of states, and to a certain extent CMS, but many states -- and that's why I used New York and Texas, and there are almost 100 million people there -- are really focusing on paying for better outcomes.

And the intermix, the necessity for linking the question of SES is addressed,

had to be addressed in both of those. So, clearly, I hope that the charge can be changed. But, if not, whatever happens, I'm hoping that there be a clear set of suggestions on a timeline and an approach for including SES measures that are the least gamable. It is not a small issue in terms of its gamability.

I'm particularly interested in dependent variables. A clear example of what I am speaking about is the work that I am doing with several states to try to have a statewide consistent collection of the term "homelessness". And so, as a consequence, it is obvious -- I mean, for myself, I saw patients all day yesterday until nine o'clock last night -- that anybody who is homeless is a clear risk. At the end of the day, obviously, there is a lot of discrepancy over that definition. But if you can have a consistent, clear, statewide, probably HUD-reliant definition, I think that should be

absolutely useful for linking payment and outcomes quality.

So, to put it differently, the perspective I bring recognizes that at the end of the day healthcare is fundamentally an economic activity. And obviously, it doesn't requiring, from knowing what different people do here, it is clearly the poorest who are discriminated the most, whether it is in this country or the country I was born in, in Italy, where they have a much better healthcare system than the United States.

But the best way to devoid politicization from my perspective, because then it becomes whoever has the most power who is a safety-net institution, is to have clinically-robust models.

But we also need to acknowledge -- and I have been impressed by that big time by Karen, and so I want to acknowledge Karen's absolute correctness in stating that clinical models, we are dealing with human beings.

Let's get real folks, right? Which is why it is an honor to be a physician.

We will never explain 100 percent of the variance. And thus, we need to suggest a path forward and to maximize the probability of acceptance of use. So, that's important. And we need to understand that funding typically is a zero-sum game. If a safety net gets more, that means a wealthy hospital gets less. So, we need to be very robust in understanding, making the argument that a safety-net institution should get more.

And that I think can be very well-described, and I will try to show that, that if we have clinically-valid descriptions of human beings, that can be fostered. So, for me, it is not "whether," but "how" and "what timeline".

So, I am pleased to be here, and thanks.

MS. HAMMERSMITH: Okay. Thank you for your thorough disclosures and thoughtful

comments.

Do any of you have any questions of each other or of me, based upon the disclosures that you made this morning?

Okay, you do?

MEMBER LIPSTEIN: I have just one question.

MS. HAMMERSMITH: Would you turn on your microphone, please?

MEMBER LIPSTEIN: Okay. My question is, especially since I'm new to NQF and some people are not, there's an issue of whether we risk-adjust with the purpose of improving quality and outcomes or there's an issue of NQF taking the position that one of its measures, if not risk-adjusted, might not be fair to use in payment methodology. Is that on the table or is that off the table?

In other word, as part of our charge, is NQF willing to do something that I'm told, I guess, by people in the room that it hasn't done before, which is to tell CMS

what it should and should not do with the outcome measures that it endorses?

MS. HAMMERSMITH: I think that that is part of the substantive work of the Committee. So, I will let them address that after I am out of here.

(Laughter.)

You are all delightful, but I would like to leave.

(Laughter.)

Any other questions about disclosures that people made?

(No response.)

Okay. Thank you and good luck.

MS. PACE: And actually, we're moving into that right now in terms of NQF. So, we will do a little presentation and can have some further dialog about that.

So, we have one more set of presentations. Helen and I will present. And then, we are going to take a break. And then, we will come back after the break and continue

our agenda.

So, Suzanne, do you want to move on?

DR. BURSTIN: Great. Since you have all been asking for context, here it is.

(Laughter.)

So, appropriately named "Context" by Karen.

So, thank you for those incredibly-thoughtful introductions and perspectives. There's actually more commonality than I thought, walking in the room. I think as Kevin said earlier, the devil is in the details of how we sort of make this work, but more there.

So, I want to talk a little bit about endorsement, but also about what is truly potentially a significant change afoot. Some of you know Chris Cassel joined us as our CEO, President and CEO, about six months ago.

We are very much in a state of looking at what we do very differently, very

critically. And, in fact, a lot of work with our Board on strategic planning is, in fact, looking very closely at exactly the questions that all of you have raised today about our role. So, I can't give you definitive answers, but I will at least give you some context.

So, first, specifically about endorsement, and then, I will return to the questions several of you asked about how this relates to payment and selection of measures.

So, first of all, the current state is, in fact, that at least our current process is that there's an expectation that, if an NQF measure is endorsed, the Committee has decided that that measure is suitable for both performance improvement and accountability applications. We have not to date made a distinction between saying this measure is okay for quality improvement.

Oh, look, I say "Dr. Cassel," and she walked in the room.

Welcome to Chris Cassel, our CEO.

So, we have to date not distinguished endorsement for measures for different purposes. So, currently, there is an expectation that you would use performance results for a wide range of potential purposes, and they are listed out here. And to date, when we have talked about accountability, we are specifically referring to this wide breadth of public reporting, accreditation, licensure, certification, incentives, performance-based payment, network inclusion/exclusion, et cetera.

But we recognize the world has significantly changed from these early days when one-size-fits-all. And I think there is, clearly heard from our Board, as well as from the MAP discussions last week, a great deal of interest in potentially moving towards endorsement more fit for purpose.

So, I think part of what we will get a very good sense of over the next couple

of days is really beginning to understand -- and many of you brought this up -- what, in fact, differentiates measures for different purposes. When is a measure potentially appropriate, as somebody said, for private benchmarking or quality improvement? When is a measure appropriate for public reporting? When is a measure appropriate for patient-level selection/payment purposes?

Interestingly, some of you, you know, talking about payment is potentially the top of that hierarchy. And I will tell you that it is very interesting at NQF, where there are so many perspectives at the table, that we frequently hear, for example, from consumers and purchasers that their ability to select the right provider for what they need is equally high stakes for them, as the financial issues might be high stakes from the provider perspective. So, we are careful about language in terms of some of that.

But we do fully recognize that we

are at the point of needing to undertake this exercise of considering whether to endorse measures for different purposes, in addition to the fact that we have not to date dived deeply, at least on the endorsement side, into how a measure is implemented or how a measure is reported.

We have had some forays into this work. And, in fact, as part of our readmission project, that probably was the major impetus for this work a year or two ago now, the Committee did specifically recommend, as part of the use of the all-cause readmission measure that was submitted by Yale and CMS, that like hospitals be compared with like hospitals. So, it was the beginning of that thinking of how do you begin looking at least the reporting and the implementation of those measures.

The last thing is there is a part of NQF, the Measures Application Partnership that Marshall mentioned at the outset, which

has specifically been charged with helping to assist CMS and other agencies, CMS and others, in the selection of measures for different programs. And that's where the fit-for-purpose sort of thoughts have really been mostly concentrated to date, as opposed to the endorsement side.

And I think what we have increasingly seen is that the MAP is very dependent in some ways around a scientific review of the measures. And so, increasingly, we are considering better ways to begin integrating at least our internal work to potentially allow the endorsement function and the review of measures to provide a more granular assessment of the potential uses for which the scientific properties of the measures lend themselves.

So, naturally, it would be stepping into areas like this. So, we are very much at a cusp of thinking about how potentially to integrate or at least better

relate the work we do on the scientific review of the measurement properties, evidence, scientific validity, reliability, usability, and how that, then, relates to the selection of a measure for a particular program for a particular purpose. So, that is our current state.

Chris, I don't know if you want to add anything there?

DR. CASSEL: No. Well, just let me welcome everyone, and I'm sorry I can't actually be with you for the entire meeting because this is such an important issue for the nation, actually. And so, we really appreciate your contribution to helping us take this issue to the next step.

I mean, Helen has, I think, described very well what NQF Board and staff, I think, really think we are at the cusp of having to look at the endorsement process very differently, given how innovation, and some people would say chaos, but it is probably a

combination of the two, are helping in the healthcare world, much of it around measurement, measurement systems, ways of using data in this extremely data-rich environment that we now find ourselves in.

And the NQF process and single set of criteria for endorsement just isn't adequate for that, for all of that purpose. So, this may have seemed, when we put out the call for nominations, this may have seemed like a relatively-academic sort of methodological discussion, and I'm sure you'll have plenty of that, but it also is very consequential, I think, in helping us think about should we have multiple different approaches to measures that mean different things for different audiences, for different purposes, et cetera.

So, thank you again for your participation, and I just look forward to a real interesting process and report.

DR. BURSTIN: Thanks.

So, next slide.

So, I think on this slide we just give you a couple of examples of the current uses of endorsed performance measures, both public reporting -- for example, the measures on Hospital Compare, Nursing Home Compare, fill-in-the-blank "Compare" programs, as well as the way some of the measures have been used quite extensively, particularly on the hospital side, around the Readmission Reduction Program, Value-Based Purchasing, and the Shared Savings Program, again, just as an example.

So, we do have a very close working relationship with HHS and CMS, and that was brought up in some of the questions early on. And again, this is independent work. It is funded by HHS. You will have an opportunity to hear from Pam at the table. You will have an opportunity to talk to Kate later as well.

But, again, I think it is just

really important to emphasize this is very much an independent endeavor. There is no sense that we have at the start of this work how it will turn out, and that is really our intent.

And you probably have already heard that you come from such an incredible variety of perspectives and expertise, that I have no doubt this will be a great effort going forward.

Karen will go into a deeper dive to follow on our criteria.

And somebody had raised the issue earlier of how CMS -- oh, I guess it was Susannah -- how CMS, basically, and Yale followed NQF's guidance. So, it is, in fact, true that NQF's guidance to date has been not to risk-adjust for these variables for which there is potential for obscuring disparties and really having a preference for stratification.

So, in fact, this Committee's

findings may have a very significant impact on our criteria for measures, which will, then, obviously, have a pretty significant effect on measure development and potentially changes to measures going forward. So, I just wanted to put that out there.

Karen?

MS. PACE: Okay. So, next slide.

And we won't get into our criteria in great detail, but, basically, we have a section of our criteria about scientific acceptability measure properties, primarily reliability and validity. And when we look at validity, we also in terms of what we ask the measure submitters to do is to do some traditional validity testing, but we include under our validity criterion looking at things that could be threats to validity.

And generally, when you are looking at outcome or resource use performance measurement, if you don't risk-adjust, you're at risk of incorrect inferences or conclusions

about quality. And that is why we kind of put it with our thoughts about validity.

Next slide.

And as Helen mentioned, I mean, our criteria says we're looking for an evidence-based risk-adjustment strategy. It should be based on patient factors that influence the measured outcome, but not factors related to disparities in care or the quality of care and are present at the start of care, have demonstrated adequate discrimination and calibration.

And one of our notes is specifically risk models should not obscure disparities in care for populations by including factors that are associated with differences or inequalities in care.

And I should point out -- and I think Sean's and another people's comments earlier on -- that stratification is one way of adjusting. And we should clarify that this was really specifically talking about them in

a statistical risk model versus doing something with stratification. And the preference was, in the light of interest in identifying and reducing disparities, of not obscuring those.

So, any of you who have been working with NQF over the years, you know that we evolve as the field evolves. We are definitely open to revisiting this issue and really thinking through it with all of you as experts, and seeing where we come out.

As Helen said, the recommendations that you come out with will definitely impact how we state our criteria and how we implement that, which will have implications for endorsement.

Okay. Is that the last one?

Oh, so the other thing that I wanted to just talk a little bit about -- it has come up several times -- about adjustment for performance measurement versus adjustments that are done in terms of determining payment

for providing services, not the pay for performance that is based on quality performance. But, obviously, those adjustments are also made.

And some examples of those that are already in place are there some hospital payment adjustment for disproportionate share of certain low-income patients. We have just posted some of these things on our SharePoint page for you all. That is in a fact sheet about Medicare hospital payment.

And certainly one example would be in the inclusion of Medicaid status and case mix adjustment for Medicare Advantage plans. Sometimes Medicaid status, obviously, is tied to income and sometimes that is used as a proxy.

So, these things are happening, maybe not as systematically or to the degree that we want. And I think there's obviously a couple of things in regards to our work here. Our specific charge is about outcome

performance measurements, which is what NQF endorses, but we totally recognize that there is a linkage and overlap. There's not a bright line between these things because, as Kevin pointed out, the payment affects the resources of the provider, which, in turn, affects the kinds of care that they are able to provide. So, they are definitely linked, and we're not saying that we can't discuss those, but, ultimately, our charge is: what are we going to do about outcome performance measurement?

We certainly can include recommendations. It is not going to have the same kind of effect because NQF doesn't do anything in the adjusting for payments, that realm. But certainly we don't mean to cut off that kind of discussion or how they interact and questions about, well, what if the payment really did adequately adjust, so that providers were given adequate resources to care for these patients? Does that have any

implications for what we would do or wouldn't do on the outcome performance side?

So, I think we definitely will need to have those conversations, just in terms of kind of our realm of working in performance measurement in general.

Okay. I think that was the last -- okay, so this is just exactly what I was just saying. You know, similar issues; it's related, but in terms of being out of scope, as I said, it is really in terms of what we really can implement and some of the things that we have already talked about.

You know, certainly, these patients have greater needs. And does payment actually reflect the cost of caring for these patients? And then, if it does, does that have any implications for how we think about outcome performance measurement?

So, we have a few minutes. We can have some clarifying questions or some comments about this or anything we have talked

about this morning before we go to break, and just want to open that up.

MEMBER GARRETT: So, you gave a couple of examples, Helen, of some ways that NQF measures are used in national reporting programs and pay-for-performance programs. But, I mean, NQF measures are used a lot at the local level, too. So, I think the impact can't be understated here, the work.

DR. BURSTIN: You're absolutely right, and we have really just begun to sort of dive deeper into particularly some of the state-based issues, which I know you are very familiar with, Nancy. But, again, I think these have broader implications than just federal. But since there is such a strong focus around some of the federal measures we've been talking about, we thought it would be appropriate to just give those as examples.

MS. PACE: I misspoke. Actually, we have David presenting on the MedPAC recommendations. David, you may have been

ready to make a comment first, but I just wanted to --

CO-CHAIR NERENZ: No, actually, I was just going to do a very mechanical process. As we get to the point where it's a more open flow of discussion, we just need a rule for how do people indicate that they have something to say. My suggestion is that they do this (indicating). Because the trouble is this is such a big room, that a gentle and polite, this sort of gesture, we can't even see at the far end. And we'll try to keep track of whose things have gone into the vertical mode first. Okay. And I see this starting up. That's all.

MS. PACE: All right. David, we were going to have you talk a little bit about the MedPAC recommendations.

CO-CHAIR NERENZ: Can we have a couple of comments, though, on this?

MS. PACE: Yes, sure. Sure.

MEMBER LIPSTEIN: I wanted to see

if we could add a third bullet to your implications that derives from the evidence around the human behavioral economics. And you just took the slide away, which was important.

Because the two implications have to do with the cost of providing care. And what's obvious now, from what Dr. Cassel described and what Helen described, was that, when risk exists, okay -- I'm going to take this to a real high level -- in the world of behavioral economics, when risk exists, people will take on that risk if the rewards are greater than the risk. People will try to reduce the risk, and people will try to eliminate the risk. But, if they can't, they will avoid the risk.

And so, risk avoidance becomes a reality when payment methodologies are introduced when you introduce financial into either offsetting or not offsetting risk. And one of the reasons I sent in papers related to

a non-healthcare application, which was teachers, was because in the world of trying to improve public education and public test scores for fifth-graders, third-graders, and eighth-graders, what we have learned is that, if you don't risk-adjust test scores, and teachers, then, perceive themselves to be on an unlevel playing field, do they avoid the risk of working in high-vulnerable innercity public schools? And so, there is a lot of literature out there now about whether Race to the Top actually became Race to the Suburbs.

And so, I wanted to introduce the concept of behavioral economics in this because it is highly relevant.

MEMBER GROVER: Just a question, since I'm not as familiar with the NQF process. And you had mentioned the role of the MAP. Typically, does the MAP just say, "Use this measure" or does it ever say, "Use this measure, but adjust it in this way" or "Use it in this circumstance or not this

circumstance."?

DR. BURSTIN: Yes, so the MAP does provide some conditional support with some conditions put forward. Again, I think they often look towards the scientific review of the measure for that input, and I think that is where the issue has been, to make sure that we have got the appropriate input for them as they are making that. Do they have the right information to make those recommendations?

But it's an excellent question.

MEMBER CASALINO: Yes, I'm just delighted to hear that NQF is open to discussing at least the possibility of recommendations for how these measures could be used.

But I just want to highlight, once again, I think it seems to me that this is different than the average thing that NQF has had to consider in the past, right? So, it is one thing to think about what's a good measure of cardiovascular care, right? And that's not

easy to do, but one can use certain criteria to decide what is a good measure and, then, put it out there, and that's it, right?

Well, so we really have two tasks, as I see it. One is that, right? What is a good measure or measures of SES, for example, right?

But a lot of the energy in the room, however, appears to be about the other task, which is a less traditional task for NQF, as I understand it, which is how should the measure be used. And that's important for SES measures in a way that is not true for just measuring cardiovascular care, you know, what you could measure of cardiovascular care.

Because, I mean, if you asked me, I think, or probably, it sounds like, most of the people in the room, "Would you endorse X measure of SES?", you know, my answer right now would be, "Well, yes, that's a good measure, but depending on how it's used, it could make things much worse. It could make

disparities much worse or it could make them better."

So, I can't endorse that measure without also give me statement about the ways that the measure I think could be accurately use. Because, if it is just endorse, yes or no, we're put on the horns of the dilemma which we're all aware of.

So, I do think that most of the energy in the room that we have heard so far is about the second task, which is how it should be used, how measures should be used, the SES measures. But, you know, the first task is also a difficult one. And fortunately, there are some experts in the room, not including me, to help us decide, well, leave aside for the moment how should they be used; what actually are good measures of SES for our purposes?

MEMBER SUGG: To kind of segue on what you are saying, I think when we talk about socioeconomic status, the issue is, if

we try to get very granular, so we can really compare quality to quality, it takes a lot of money to do that. And I think, when we talk about cost, that is one of the things that I want to make sure is really out there: what is the burden if we want to collect a whole lot of socioeconomic things, so we have a really good quality measure? Great. But if we have overburdened organizations and actually taken away money from patient care to do this, then we also increase the disparities.

So, it would be great if we could put all these amazing things into our socioeconomic status, but I am very sensitive to how much is this going to cost medical care to do this.

CO-CHAIR NERENZ: Let me just make one quick response to Larry's point, and I'm going to look very much sideways for clarification because I am going to say something that I may not be the right person

to say.

But it had not occurred to me until you said it that our task or NQF's task was to endorse SES measures. It just hadn't occurred to me, with that phrase in it, that's what we were about here, at least in the sense that NQF endorses quality or performance measures.

I had understood our task to be that we were talking about the inclusion of one or more sociodemographic factors in the adjustment of quality or performance measures. I'm seeing some nods around the table.

So, I am checking here, too, because I worry we could be distracted if we somehow begin to think that NQF is in the business of literally endorsing SES measures.

MS. PACE: You are absolutely right. We are talking about use of these factors in adjustment of outcome performance measures.

Now, obviously, part of that

discussion is what sociodemographic factors and how that can be measured and the data burden, et cetera. But you're right on in terms of our mission.

DR. BURSTIN: And just to build on that a bit, we have been doing a fair amount of work in the population health space. We have, in fact, been looking at the question of, should we be endorsing measures, for example, of social determinants of health and this question of, is it at what level? Is it a community? Is it a state? So, these issues are on the table, but I think they are directly on the table for this particular discussion.

CO-CHAIR NERENZ: And then, having going that step out on a limb, I'll go yet again. In response to a comment Norbert made, it had also not occurred to me that we would be talking about situations in which SES or a sociodemographic variable was a dependent variable with, presumably, some healthcare

thing as the independent variable. It had occurred to me, I had assumed it was always in our discussion the other way around, that the SES or sociodemographic variable is an independent variable; a healthcare performance or quality measure is the dependent variable. And I am seeing a few nods about that as well.

MEMBER GOLDFIELD: I don't know who added that, because I think, Kevin, in your email, you know, actually brought that issue up. I mean, I will discuss that in one of the slides that I will show.

CO-CHAIR NERENZ: All right. I'll do this very quickly because I know we are up against our break time.

And first of all, a little context explanation. In the previous version of the meeting agenda, this made much more sense than it does now, just the sequence.

(Laughter.)

There was going to be a CMS presentation about their general policies and

principles about SES adjustment, which generally are of the nature not to do it. And then, in that context, it made sense to say, well, why did MedPAC do something different? And then, I was going to talk a bit about that.

We don't have the previous context, but here we are. So, what I will try to do just very quickly is to talk about the recommendation MedPAC made in the June 2013 report about the hospital readmission measure, which, in fact, we did perceive as being somewhat running against the grain of the typical CMS and NQF policy.

I tried to be just literal in the selection of tables from the report. They end up being a little busy. I hope people are close enough to one screen or another that they can see it.

The real technical trick here, though, is I have, will have shortly, a laser pointer that can only shoot at one screen at

a time.

(Laughter.)

I haven't learned how to do this. I can't even get to the -- oh, there we go. Okay. As long as I don't hit something.

All right. The issue in front of us at MedPAC back a year or so ago was what we started with, an observed empirical relationship between a particular measure of low income and the readmission measure.

So, what we start with here is a table in which hospitals are grouped into 10 deciles on the basis of the proportion of their Medicare patients who are eligible for supplemental security income. So, I will point out right here this is just one variable. There could have been others. This has at least some advantage of being objectively measured. It's hard to game. It's in administrative datasets.

And we clearly could, staff actually, since Commissioners didn't do this,

could have done quintiles, could have done quartiles. It just happens to be deciles here.

So, the deciles are defined by the percent of patients in the overall Medicare mix who are on SSI from low to high. So, relatively affluent patients, relatively poor patients.

The key point is that, if you look at the average readmission penalty -- and this is now a percent of hospital revenues -- it is relatively small here; it is relatively large here. If we, then, look at the percent of hospitals who got the max penalty, relatively small here; relatively large here, a factor of four difference.

And then, it is just the inverse or converse. Those percent with no penalty, relatively high here; relatively low here. Okay? So, that's the starting point.

And as a group, we felt this was wrong. And I guess I may as well just say

that explicitly.

Okay. Next slide.

So, there was an alternative method derived that would apply the penalty within strata. So, if you were relatively high in your readmissions, within your decile there would be a penalty. There are some technical details we don't have time to get into here.

But, basically, these are sort of the current state penalties as a percent of hospital revenues. It repeats what we saw in the previous slide. In a peer group comparison where you're compared to those in your decile, these are relatively even.

Now you see these are all relatively high, actually. That's because there was an arbitrary target set. That could be modified up/down. So, you could make these as big or small as you wanted. The point though, is that they are relatively equal across the deciles.

Okay. Next slide.

Now we recognize, of course, that this is not a static thing, that every year hospitals do work on this. We have evidence of national improvement. We have evidence of local improvement.

So, there was a question, so what happened if all the hospitals got 10 percent better? I won't walk through every slide, except just to point out that, under the current penalty, again, there's a relatively-light hit here, a relatively-heavy hit here.

And if you apply the adjustment model -- and here's where this table gets a little difficult to follow -- it becomes more even, but, again, it is a little hard to intuit this when it is expressed as a percent of revenue penalty. The only thing I can say is that it is relatively more even, which I think is the point of the adjustment.

Okay. Next slide.

So, again, this is just to

represent an illustration. The first thing is the stratification here involves a stratification of hospitals into deciles that are defined by this percent SSI. There are all sorts of other ways to do stratification. This is a way.

We did not talk about, nor recommend, the stratified reporting of readmission rates within each hospital for individuals either with or without SSI or some other variable. We didn't say anything against it. We just didn't speak to it. So, just to point out, that could be also part of this picture.

Percent SSI is an option, one option among many; could have done something else. The fact that it is a MedPAC recommendation does certainly not mean that it is law or that it is CMS policy. We advise. People either take the advice or they don't. So, it has that standing, but has not yet been implemented, to my knowledge, as actual

policy.

And then, we have this question that kind of gets into this interface between what's performance measurement and what's payment. We recognize this could be a question, well, do you need to do that if these high SSI hospitals are already receiving DSH payments? I mean, hasn't the problem essentially already been solved? And we just observe -- and this is just a verbatim cut-and-paste from the report -- that, in principle, the DSH payments go for the higher cost of treating patients while in hospital. They are not designed to cover any excess cost related to readmission or to offset a readmission penalty. And I see some nods around the table.

Okay. So, that's it. It's just to indicate that there is this recommendation out there. It would have made a little more sense this morning if we could have set that a little more clearly, sort of against where

CMS has been on this. But it is an illustration and it is a formal MedPAC recommendation.

I am not personally aware of any response. They would not come to me. Again, those tend to come to Mark Miller or to Glenn Hackbarth, not to me. I'm only one of the Commissioners.

Others actually may be able to speak to that.

MEMBER GROVER: I can tell you that, during both discussions with Senate Finance originally in thinking about this, the original legislative language had more explicit direction to HHS to account for factors such as SES. That was stripped in the end, but it was still an option with the Secretary.

And so, when this data and similar data that we ran and did by DSH percentile showed similar findings, I think the two responses we really got were, one, that means

somebody who is not losing as much money now is going to lose more money, and we're worried about doing that. Otherwise, we will get less money out of the hospitals as a group, and we don't want to give it up.

And third, they ultimately came back to, well, if Denver Health can do this, and they have a decent outcome, why can't the rest of you, right? So, if I can find you one example of a place that it works, then Henry Ford and BJC and everybody else should do it. That was their response.

MEMBER LIPSTEIN: So, you have introduced the subject of stratification, which I think is an important topic for our panel to think about and talk about. But it reminded me of something Gene said. Because there is literature that suggests that within strata you can still make improvements in readmission rates. And not everybody who is a high SSI decile necessarily has a poor readmission outcome.

So, we went in search of what else is going on. And having lived in Baltimore, East Baltimore, at the time they closed Baltimore City Hospital, and then, lived in Chicago where they have Cook County Hospital, and worked in Atlanta, where they have Grady Memorial, what we found -- and this is really important -- is that each part of the United States has a different tax base.

And so, in the case of Denver or Cook County or Atlanta, there's a local tax base that supports the regional safety net. And if you don't have a local tax base that supports the regional safety net, even if you have a high percentage of SSI or a high percentage of DSH payments, there are local circumstances that really drive outcome.

And so, what I am worried about here isn't BJC. In rural communities, they just can't increase taxes to support their safety net. In Denver, they do increase taxes or do have taxes. In Kansas City at Truman

Medical Center, it pays for a whole primary care apparatus when you discharge patients from a hospital into the community, if there is a safety net there that is funded by the local community.

What is happening in America right now is federal funding of the local safety nets is changing through reduction of DSH payments or through these penalties. And so, it comes back to, if you write policy at the federal level, MedPAC, without recognizing that there's real variability in local tax bases supporting safety nets, you get very variable outcomes.

And one of the purposes of NQF policy is to eliminate variation or eliminate disparities. But if you don't address those local tax base issues, you are going to again get to the wrong answer.

MS. PACE: Okay. One more comment, and then, we need to break and get back.

MEMBER SAWHNEY: We need to keep some different perspectives in mind. One is we all come from a public health background, but one of the major players is the commercial insurance players. And we're all very comfortable talking about race and racial disparities.

Commercial insurance companies have a history that goes back long before I was in the industry of acting very badly with respect to race and getting their butt sued. And after that, they officially became race-blind. Now it is not to say they really are, but they try very hard to keep up that pretense and to openly talk about race and race disparities in commercial context. This is very difficult for them.

And I'm just putting that on the table. They will talk about income much more readily, and they are players in this.

The other thing, they are also players -- I mean, Medicare and Medicaid have

clear social good objectives. So, when we talk about adjustment payments and talk about adjusting Medicare and Medicaid payments, that's one thing. But, then, when the same providers want the commercial world to adjust, that's going to be a different issue. And then, the commercial world is going to say, well, if we have to adjust our payments to providers, who is going to pay us for that? So, things to think about.

MS. PACE: Okay. So, we are actually now behind a little bit.

(Laughter.)

But great discussion.

And why don't we take a 10-minute break and be back here at 11:10 and we'll reconvene?

(Whereupon, the foregoing matter went off the record at 10:58 a.m. and went back on the record at 11:12 a.m.)

CO-CHAIR NERENZ: One thing I think most of us learned very early on in this

kind of context, there is no such thing as a 10-minute break. It's just not possible. You can say 10; you can try 10. Twelve is not bad. Twelve is pretty good. So, I think that is about as good as we can do.

We have a couple of presentations here. In terms of trying to catch up a little bit, I'll watch the time allocations pretty quickly and I'll start waving or flashing a light or something if we are getting close.

There is a time block for question and discussion. We don't want to cut that off entirely, but let's try to have the questions be clarification questions rather than broader discussion, because for the rest of the two days there will be much chance for some of the broader discussion. And I think, by doing that, then we can get this bit and, then, to the next panel, and eventually to lunch, which people will start thinking about.

All right. Susannah?

Let's do this: since there are

two in this section, if there is an absolutely-direct clarifying question for Susannah, let's do that before we move to the next, just because, otherwise, we'll lose track of it. But let's try, again, to keep it very much focused on what does this number mean; what did you mean when you said this, that kind of thing. Okay?

MEMBER BERNHEIM: You'll just advance slides for me? Okay, great.

Okay. So, I'm Susannah Bernheim again, and I, as I said, come with many hats. I am right now putting on my measure developer hat. My goal is to talk through how we have thought about this issue, particularly with reference to the readmission measures.

I am going to show a couple of slides that are pretty basic about sort of how we think about risk adjustment. People are coming from very different backgrounds. Forgive me if this is oversimplified, but I think it is an important baseline for sort of

how measure developers kind of approach this problem.

I am going to show a couple of slides that show analyses and try to say both, per Karen's request, whether we have done this in other measures, so people have a sense of how universal these findings are, and how we, as measure developers sort of think about these findings. And then, I have a couple of conclusions. And I am going to try very hard to go slow enough that you can understand me and quick enough that we are within 10 minutes.

So, when we are building risk-adjustment models for the purpose of measuring quality, this first one is a very important point, and it is often confused. We are not aiming to maximize patient-level prediction. I am not putting everything I can in the model to predict whether or not a patient is going to be readmitted. And there's a number of reasons for that.

But the most obvious is that I am trying to illuminate quality. So, conceptually and very simplistically, as a measure developer, we think about a patient outcome as being the result of the baseline status that they come in with, the quality of what -- this especially applies to hospitalized patients -- the quality of what we do, and some random variation producing the outcomes.

So, what I am trying to do is level the playing field for those baseline factors and not adjust for anything that is largely mediated by the quality of the care that is provided, because I am trying to illuminate those differences.

So, to make this really concrete, if I wanted to predict mortality, and I knew whether a patient had had a complication that led to them going to the ICU, I would certainly build that into my prediction model because those patients are more likely to die

in the next 30 days.

If I want to look at quality, a hospital that, once you have risk-adjusted for how sick the patients are when they come in, that has more patients having complications that lead to ICU stays should probably look worse on a quality measure, right? So, this is, again, oversimplified, but that's what is in the measure developer's head: what is baseline and somewhat unmediateable/mediateable health status? And what is potentially quality of care? And it is not always a simple decision, but that is sort of how we think about measures.

Can I have the next slide?

The other thing you need to understand -- and this applies to the readmission and mortality measures; some measure developers do this differently -- the measures are designed to be relative measures. We are comparing a hospital, actually, really not to the hospital down the street, but,

actually, to what the model says would be expected for an average caring for the same patients.

We use hierarchical modeling, which allows us to account for the clustering of patients within hospitals. It produces a ratio, which we call a predicted-to-expected. The predicted is complicated, and I spend a lot of my days trying to stand between hospitals and statisticians and somehow make these things make sense. But you can think about it in this context as being analogous to an observed-to-expected. We are creating a ratio that says, how does this hospital do compare to what a hospital with a similar case mix would do, an average hospital with a similar case mix would do?

And so, quite literally, when we are talking about risk adjustment, we are talking about the setting the expected for a hospital. We are talking about setting the standard. And people really don't like us

using the word "standard," but I think, quite honestly, that's true, that when you put something into the risk-adjustment model, you are setting what is expected for that hospital. You are saying, what is the case mix of this hospital that we are comparing it to? So, that is how we think about risk adjustments.

I'm just going to make sure I haven't forgotten anything crucial, but I don't think I have. Okay.

So, the next slide.

There are a few basic standards about risk adjustment. These are both consistent with NQF and with published standards about how you develop outcomes measures.

One is that you want to adjust for factors that are present at the start of your measurement period. That is also consistent with what I said about complications.

And again -- I sort of said this

earlier -- you don't want to have the factors that are clearly affected or mediated by quality, like complications.

SES is really hard. So, as a measure developer, how do I think about SES? I know that SES affects baseline health status, and our models reflect that.

So, I will tell you that, when we look at Medicaid versus non-Medicaid patients in our models, the expected readmission rates for the Medicaid patients are higher. They are sicker, and the model counts them as sicker. So, we are accounting to some extent for the fact that these patients are coming in. Part of the way SES plays out is that patients come in sicker, and the models accounts for that.

There are many ways in which quality of care can intercept with SES. We know from some of the literature on race, in particular, that Black patients are more concentrated in poor-performing hospitals.

They are literally going to hospitals that do less well. We know that there may be within hospitals differential treatment, and we know that hospitals may, in the context that we have been talking about before around resources, not be able because of resource constraints to give the same or I guess to give the quality of care we would aspire to, that some patients may require more resources, and hospitals may or may not be able to provide that. And that is a very complicated one.

And last, there may be pathways that don't fit into either of my nice boxes, which is sort of what happens on the back-end and how much can a hospital or a health system affect adherence or access. So, there's lots of ways that SES can play out. So, it doesn't fit into one of the boxes. And so, we're stuck sort of thinking about what is the risk/benefit of where to put it in our models.

The one thing I will say is this

concept of sort of the hospital's responsibility has been -- actually, let me show you some data, and then I will come back to that.

So, the first thing we do -- and I don't want to actually go to the next slide yet because I want to set the stage for it -- is we just look at the hospital level, at how hospitals that are caring for a large proportion of low SES patients compare on the measures to hospitals that have fewer low SES patients.

And I will tell you we have done this using many measures of SES. The easiest thing for us to do is to look at the proportion of patients that are of Medicaid status. And we can do that either at the hospital level or the proportion of patients within the measure that have Medicaid. We have linked patients' zip codes to the Census tract and looked at the median income or the poverty level from their Census tract. We

have identified hospitals based on whether they are considered safety-net hospitals, comparing them to the average Medicaid caseload in their state, because Medicaid is different by states, and whether they are public hospitals. We have looked at this percentage.

We have grabbed what variables we can and done these analyses many different ways. And the two key things to know are you identify different hospitals, right? And this Committee knows this, but when I said who are the 20 percent of hospitals caring for the greatest burden of low SES patients based on Medicaid status, they are not the same, if I say who are the 20 percent of hospitals caring for the greatest proportion of low SES patients based on the zip code median income. And I don't know what the right one is.

So, what I have chosen to show you today is actually the most extreme version of this because I don't want to oversimplify

this. So, what we have done in the next slide is we have taken hospitals and the heart failure readmission measure, and we have used Medicaid status, because that's where we get the most extreme differences, and we have taken the 20 percent of hospitals that have the fewest Medicaid patients and we have taken the 20 percent of hospitals that have the greatest percentage of Medicaid patients, and we have just lined up the distribution of their performance.

And now to the next slide.

So, what you see here -- oh, the slide doesn't work perfectly; I seem to have repeated my figure.

So, what you see here is what we call the Q1, which is not actually quartile. It is quintile 1. These are hospitals that on average have only 7.1 percent of their heart failure patients with Medicaid, and the Q5 hospitals are hospitals that on average have 55 percent of their patients on Medicaid. So,

a pretty big difference in the case mix I'm looking at.

And you see that the white and gray is those Q5 hospitals. These are the readmission rates, the distribution of readmission rates for hospitals caring for a greater proportion of Medicaid patients, and the Q1 is the distribution of hospitals' performance on the readmission measure for those caring for the lowest percentage. And they overlap a lot, and they don't overlap completely, right?

So, again, what I have given you -- and we have done this now looking at certainly all of the publicly-reported measures and a number of the measures in development, and generally, these curves actually overlap more. What you see is that there is a difference between those two curves and that they are largely overlapping.

And Steve loves it when CMS says this, but I will say it again: you see in

those white bars many hospitals with very large Medicaid proportions achieving low readmission rates.

MEMBER OWENS: Just a quick clarifying question. When you say "percentage of Medicaid patients," is that percentage of Medicaid patients who are also Medicare, or are you looking at a hospital characteristic overall in your definition, whatever definition, you are actually all patients --

MEMBER BERNHEIM: Right.

MEMBER OWENS: -- because you have focused mostly on --

MEMBER BERNHEIM: So, we have done it both ways without a huge difference in our findings. In this case, this is the heart failure readmission measure, and we are looking at the proportion of the patients who are all fee-for-service Medicare patients in the measures who are dually eligible for Medicaid. So, I should have been more clear about it.

But, again, we have done it by just looking at the hospital's percentage among all their patients as well.

So, what do we do with this? As a measure developer, we say these are a little bit different, and we see that there is a lot of burdened by low SES hospitals performing well on the measure. And it doesn't fully answer the question, but it is important.

I am just making sure I'm keeping up with my own notes here.

And we don't know how to interpret the slight difference that remains, again, whether this has to do with inherent patient factors or differences among what hospitals are capable of. But we do think that it is important that these hospitals are capable of performing well on this measure.

The next thing I am going to show you is what if we took this exact same measure with the Medicaid status as an indicator of risk and in the patient model risk-adjusted

for it. And we are often reticent to show this figure because I don't think conceptually whether risk adjustment makes a difference or not is the right rationale for putting it in the model. I think it is more of a conceptual thing. But people are really interested to see this.

So, if you'll go to the next slide?

Here I have on the X-axis the risk standardized readmission rates for all of the hospitals with the current measure that is in public reporting. I have now pulled into the model on the Y-axis adjusted for the same patient-level measure that I used in the previous one to define the hospital groups. And we have color-coded, and I know this is hard to see, but that is part of the point.

So, the Q1 hospitals are the ones who had the least low SES patients. They are in blue. You can see that they rise a tiny bit off the line, if you look really

carefully. We have lumped all the Q2 through 4 hospitals together. I think that is mislabeled as Q5. It is Q2 through Q4. They are the yellow hospitals. And then, you can see the red hospitals are the ones that have the greatest proportion of Medicaid patients.

And you see that, in fact, despite the slight differences we saw in the two distributions, when we add SES to the risk-adjustment model, it makes very little difference. And I'll give you some numbers to go with this, just so you have a sense.

Of the Q5 hospitals, those with 55 percent on average Medicaid patients, their RSRs change on -- the median hospitals' RSR among that group changes by 0.17 percent. And the ones that change the most, 5 percent of the Q5 hospitals by 0.5 percent. And you're now down to 20 percent of hospitals, and 5 percent of those changing by half a percent.

So, in this measure we don't see a big difference. And again, we have done this

with the income as well to see if it looks different.

MEMBER LIPSTEIN: Susannah, just because I don't know how the statistics work, the variability of eligibility from state to state, does that all just get mooshed in here or does this somehow take into consideration that some states are eligible up to 138 percent of federal poverty and some aren't eligible at all?

MEMBER BERNHEIM: Right. So, this doesn't, which is why we have done it other ways, because there is inherent problems with all these variables. So, then, we have linked to income; it looks the same.

What we can't do is -- well, what's too complicated to do in this room is to try to account for those things simultaneously. Maybe somebody could. I don't know if it would make a huge difference.

Do people have other questions on this?

CO-CHAIR NERENZ: I know it's a little tough. We're already a little past 10 minutes.

MEMBER BERNHEIM: Oh, are we? Okay, I can be real fast. I've got two more slides and that's all.

Okay. Next slide.

Okay. So, conceptually, now we are still stuck with this problem of sort of patient-versus-hospital influence. So, we did one other analysis. This we did in the hospital-wide measure.

And here's what we did: we took the hospitals and stratified them based on Medicaid patients, but, then, we ran the model only on their Medicare patients. So, they are now not being judged on the outcomes of the Medicaid patients; they are only being judged on the outcomes of their Medicare patients, but we know what group they are in.

What you see is that the high Medicaid hospitals still have slightly higher

readmission rates among their Medicare patients, which are used at least for some influence of the hospital.

I am going to do one last slide.

And we can talk about this later, but it is really important to understand with these measures that stratification is not simple. So, I am not going to talk about patient-level stratification because it hasn't been mentioned a lot.

But there's two ways to think about stratification at the hospital level. What we have to remember is that these measures are observed-to-expected-ish. So, you come up with a ratio, and then, you multiply it by a national rate.

So, if you separate hospitals into two groups before you run the model, you then have different rates. And so, it is going to be very confusing because Hospital A from strata 1 and Hospital B from strata 2 might both be 24 percent, but for one of them that

was against a 23.7 national rate. So, they are actually doing a little worse than expected. And the other one, it was 24.5, and they look the same, right? Big mess.

What you can do is run the model on everybody -- and this is what I believe MedPAC's recommendation is -- and then, stratify the hospitals and set a cutpoint that is different for the two strata, right? And that's my understanding of MedPAC's recommendation. That's not really stratifying the measure. It is stratifying the hospitals after you have applied the measure.

My last one is just a conclusion. It just says what we found. We don't find it determinative of hospital performance. There is a wide range: how SES defined changes, what hospitals are identified as low SES. Risk adjustment does not change hospital performance substantially, and we find that there's both a hospital influence as well as a patient influence on the outcomes.

CO-CHAIR NERENZ: Thank you.

Any other immediate clarifying questions?

Yes, Nancy?

MEMBER SUGG: I just had a question. When you said that you ran it also by income --

MEMBER BERNHEIM: Uh-hum.

MEMBER SUGG: -- was that zip code income? Okay.

MEMBER BERNHEIM: Yes, it is zip code income, which I think at a patient level may not be ideal, but at a hospital level, if you are trying to understand whether they've got a lot of poor patients, if you know all the neighborhoods they're coming from, it probably helps you understand the kind of patients. And then, we see hospitals with over 90 percent coming from low-income areas and other hospitals with 1 percent.

CO-CHAIR NERENZ: Thank you. Very helpful.

Our next person is actually on the phone, I understand.

MS. PACE: Yes.

CO-CHAIR NERENZ: Okay. So, I was looking around the room to find that person and said, wait a minute, I'm lost.

(Laughter.)

MS. PACE: Sajid?

MR. ZAIDI: Yes, I'm here. Can everybody hear me?

MS. PACE: Yes. Do you want to just introduce yourself quickly? And then, we will get into your presentation.

MR. ZAIDI: Yes. So, I am Sajid Zaidi. I'm a measure developer here at Acumen, and we are the measure developer for the Medicare spending-per-beneficiary measure contracted with CMS.

Yes, and so, we just went through the NQF endorsement process, actually, this fall. So, yes, we have been looking at these issues quite a bit.

So, should I just get started then?

CO-CHAIR NERENZ: Yes, please.

MR. ZAIDI: Okay, great.

Could we go to the second slide, please?

CO-CHAIR NERENZ: We're there.

MR. ZAIDI: Great.

So, I would just like to provide a brief overview of the Medicare spending-per-beneficiary measure for those who may not be familiar with it. The MSPB measure measures total Medicare-allowed cost for hospitalization episodes. So, it is a cost measure for hospitals where cost is defined as spending by Medicare. An MSPB episode includes all Medicare Part A and B claims, but not Part D, between three days prior to the index admission date up to 30 days after the hospital discharge date.

This measure includes all conditions. So, it is all discharges. It

applies to Medicare fee-for-service beneficiaries discharged during the period of performance, which is usually a one-year period, for hospitals paid under the inpatient prospective payment system which are located in the 50 states or D.C.

The measure is payment standardized and risk adjusted to allow for a comparison across all hospitals in the country. So, we remove things like IME and DSH and the effects of the wage index.

The MSPB amount is the average payment standardized risk-adjustment spending across all of the hospital's eligible MSPB episodes. And then, in order for an episode to be eligible, you know, the patient has to be enrolled in Medicare fee for service through the whole episode time window, and they can't have a primary payer other than Medicare.

And finally, the final MSPB measure is the ratio of the MSPB amount for

that hospital divided by the median MSPB amount across hospitals. So, this is just for ease of interpretation. So, a measure of 1 means that the hospital is performing, is at the median across all hospitals.

Next slide, please.

So, I would like to briefly describe the MSPB risk-adjustment model, and I would like to reiterate what Susannah said. Basically, we are trying to control for everything that we think is outside the provider's control at the start of the MSPB episode. So, we are using all the information we have which is present at admission to the index admission for the MSPB episode.

So, the risk adjustment uses an augmented ACC model and includes the following variables: age, the HPC variables and their interactions, ESRD status, disability status which is defined as whether they are eligible for Medicare under the disability provisions. We control for the MS-DRG of the index

admission, and we also control for whether the patient is institutionalized in the long-term care facility.

We don't control for gender or Medicaid status, and we did this to be consistent with NQF policy at the time that we were formulating our measure. But, for the purpose of NQF endorsement, we did test the effects of including Medicaid status as a risk adjuster. And those are the results I'll describe here today.

Oh, and one final note. We used linear regression because cost is a continuous variable.

Next slide, please.

So, these tables describe the episode level differences in spending between Medicaid beneficiaries and non-Medicaid beneficiaries. And, of course, this is just restricting to the Medicare population. So, we are looking at Medicare beneficiaries who are also eligible for Medicaid versus those

who are not.

So, in the first table you can see that, just looking at observed costs without doing any risk adjustments, Medicaid beneficiaries are more expensive on average than non-Medicaid beneficiaries, but the difference is the magnitude is not as great as one would think. It is a difference of about 2 percent.

After risk adjustments -- so, that is the second line -- you can see that the difference is still there, and it is around the same magnitude of about 3.1 percent. I wouldn't interpret the difference between the 2 percent and the 3.1 percent, I wouldn't interpret that as being significant. The magnitude is very similar.

The next table shows the coefficient if you include a variable for Medicaid status in the regression. And the magnitude of the coefficient is around $1,000, which if you express that as a percentage of

the average episode cost, that's around 5 percent of cost. You know, the p-value is zero. So, it is highly statistically-significant. So, Medicaid does, we can conclude that Medicaid status does have an impact on predictive cost.

Next slide, please.

So, this slide shows the actual impact on hospital rankings if you include a Medicaid status indicator in the regression. So, for the Value-Based Purchasing Program, the actual achievement and improvement points that a hospital received are based on the decile that the hospital falls in compared to all hospitals in the country.

And so, this first table shows the distribution of decile changes. So, if a hospital doesn't change the decile they're in, when you include Medicaid in the risk adjustment versus not including it, they would show up in the no-change row.

So, what you can see is 84 percent

of hospitals do not change the decile that they are in in the distribution, and over 99 percent have a change of one decile or no change. So, including Medicaid status doesn't actually have that much of an effect on the final distribution of MSPB measures across hospitals.

And the second table on this slide shows the improvement in the r-squared of the regression. You can see it is a very negligible improvement in the r-squared when you include Medicaid status.

So, I think the takeaway here is that Medicaid status does have an impact. It is a statistically-significant coefficient. But, in terms of explaining overall variation, it is very negligible, of course, with the major caveat being that is after controlling for all the other health factors that we have in the model.

So, once you control for the ACCs and DRGs and AH and all those other factors,

it seems that Medicaid doesn't contribute that much extra in terms of explaining overall variability.

Next slide, please.

So, yes, this just repeats what I just said. Our conclusion is that including Medicaid status in the risk model has a statistically-significant effect on spending, about 5 percent of average episode cost, but the change in r-squared is negligible.

In terms of final results, including Medicaid status has very little effect on final hospital ranking. And the final note that I would make is that Medicaid status may have a more limited effect for the MSPB measure than for other cost measures, such as total per-capita cost, because MSPB is conditional on being hospitalized. And we know that a big part of the extra cost for Medicaid beneficiaries is their rates of hospitalization in the first place.

And the MSPB measure also controls

for the initial DRG. And again, that is another margin of variation that Medicaid status could be having an impact on.

So, yes, that's the end of the presentation, and I would be happy to take any questions that people might have. Thank you.

CO-CHAIR NERENZ: Okay. Thank you. That was great.

We'll moderate questions, but because you can't see, I don't think, who is asking, I'll just ask people to start with their names, just so you know who is asking you the question.

Yes, go ahead.

MEMBER GROVER: This is Atul Grover.

Thanks for your presentation.

Two questions. One, in that shift of 263 or so hospitals that went down by one decile, I mean, these are fairly narrow differences in terms of the spend. Any characteristics that you could pull out from

looking at those hospitals in terms of region, size, public, private, teaching, that might help us get a sense of whether there is a cohort that is moving here?

And similarly, you know, when we look at Medicare spend, and if you look at how this is done with HRRs in general, you get very different maps of the U.S. once you adjust for wage indices and policy payments. But, then, if you look at Medicare total spend, which would include the Part D and out of pocket, all of a sudden, when you go from having the coasts light up as high-spend areas, you end with the middle of the country bottom-to-top as high-spend areas. So, was any work done to look at total spend on these Medicare beneficiaries?

MR. ZAIDI: Yes, that's a great point. So, we do include out-of-pocket costs. This is total Medicare-allowed cost, which includes out-of-pocket costs, but we don't include Part D, as you said.

And we did see that same pattern that you referred to. Once you controlled for geographic payment differences, the coasts become far less prominent in terms of overall spending.

But I didn't include that information here because it wasn't directly relevant to the question of Medicaid status, I think, unless I missed part of the question.

The first part of the question about the cohort of hospitals that moved down one decile, I didn't include a graph we did, but we did look at the correlation of these hospitals versus other factors, such as teaching status, number of beds, and whether they are urban or rural.

And what we saw is that, depending on what variables you are looking at, once you control for those other variables, it can have a large effect on these results. So, I guess the overall point is that, including Medicaid status in the regression, the results could be

very different if you also control for other hospital factors, such as teaching status or the number of beds or urban/rural status. So, there are a lot of interacting variables here. And, yes, they are highly correlated.

CO-CHAIR NERENZ: Okay. Thank you.

Mark?

MEMBER COHEN: I have a question for the previous presentation, if that's okay.

CO-CHAIR NERENZ: Larry, is yours on the second presentation?

MEMBER CASALINO: I just had one point about each.

CO-CHAIR NERENZ: Okay. Let's sort of, if you can flip order, because, then, we can switch back to Susannah, but at least let's stay focused on this one as long as we have the slides in front of us.

MEMBER CASALINO: Oh, sure, yes. I actually have a question, yes, about this one, just a simple point. I mean, 263

hospitals plus the smaller number that went down more than one decile, it doesn't sound like a lot, but just to kind of reiterate Atul's question, which wasn't entirely accurate, I think.

If those 263-plus, it doesn't sound like a lot, but if those are all hospitals that have high Medicaid proportions, then that's exactly the hospitals we wouldn't want to hurt. And if there is 300 of them, I think that is a lot, actually.

CO-CHAIR NERENZ: Okay. Just before we move back to Susannah, any other questions for Sajid?

I'm sorry, Nancy?

Okay, I'm sorry, I didn't pick up the question (referring to Mr. Casalino's question).

Sajid, is there an answer to that?

MR. ZAIDI: I'm sorry, I couldn't hear that question.

MEMBER CASALINO: I'm sorry, I

didn't phrase it very it very clearly the first time.

Those 263 plus some more that went down more than one decile hospitals, presumably, they were hospitals that had very high percentages of Medicaid patients, is that correct?

MR. ZAIDI: I didn't look at that hospital-level correlation specifically, but, yes, I would assume so.

But, again, I would emphasize that a one-decile change is a relatively small magnitude. We're not seeing any hospitals with -- we are seeing only two hospitals with more than one-decile change. And so, it's not like there are hospitals moving four or five deciles.

But, yes, I would agree with that point, that the hospitals that move down one decile probably have higher Medicaid percentage. But there are an equal number of hospitals which moved up one decile.

CO-CHAIR NERENZ: Okay. Nancy, why don't let's go with you, if you have one for both? And then, we'll come back and pick up Mark.

MEMBER GARRETT: So, this may be something that is going to be covered later, but I just would like a little more clarification about using Medicaid as a proxy for socioeconomic status and what evidence we have that you can actually do that.

My concern is hospitals that do a lot of unfunded care or undocumented care. And I'm also worried about rural hospitals in states that do not have expanded Medicaid, and that they're going to look worse and it is not because it is worse quality. It is, again, back to kind of local taxing, economics, politics. So, just in general comments about it.

MEMBER BERNHEIM: I am happy to respond to it, but, like everyone here, this is part of the challenge, right? You know, if

I think it is really about adherence, I don't have a measure for that.

Medicaid status, when you're on Medicare, is at least partially related to income status, but it changes by state, and it is about to change a lot. And so, I think we have to really think about that, right, because different states are doing different things with Medicaid expansion. So, it is going to differ across states even more.

Again, I chose that one. It is not actually my favorite, but it is very accessible data, and I was actually trying to show the place where we were seeing the biggest differences.

So, our group's approach has been to say, even if there is not a clear right variable, let's look at kind of everything we can get our hands on, to see if there is a different pattern. And there's different hospitals, but not a different pattern in terms of the relationship.

MEMBER COHEN: I just have a technical question. When you were doing this 30-day readmission, right, was there control for the length of the hospitalization?

MEMBER BERNHEIM: We don't control for the length of the hospitalization, again, in the spirit of sort of establishing the time zero for these measures in terms of risk adjustment at the entrance to the hospital, because after that, theoretically, the hospital is in control.

MEMBER COHEN: Right.

MEMBER BERNHEIM: But we do -- and I think it is important -- have a standard assessment period. So, some measures run into trouble when you're being watched for seven days versus 14 days, where the length of stay changes how long you're actually tracking a patient.

MEMBER COHEN: It does.

MEMBER BERNHEIM: So, these are all standardly from the time of discharge to

a non-acute setting --

MEMBER COHEN: Right.

MEMBER BERNHEIM: -- for the following 40 days.

MEMBER COHEN: But if one group is sicker, can they stay longer? That might be protective readmission.

MEMBER BERNHEIM: Right. So, you don't know, when people stay longer, whether it is differences in hospitals' approach, complication rates, sicker. I mean, we hope and believe that we're catching some amount of the sicker at the time of admission. And these measures have all been held up against chart measures, where you have medical record data to understand how sick they were and the profiling of the hospital.

MEMBER COHEN: You know, it could be correlated with the SES variable. They might be sicker going in, but in longer and, then, they are protected for readmission.

MEMBER BERNHEIM: Right. Right.

I mean, again, the models account, as best they can, for the severity when you enter, but they don't account for length of stay.

CO-CHAIR NERENZ: Larry again.

MEMBER CASALINO: It seemed to me that there was a contradiction between the MedPAC results and Susannah's results, and to some extent Sajid's.

So, first of all, is that perception correct? Is there a contradiction? And secondly, what's the explanation?

CO-CHAIR NERENZ: I guess you have to say more. What contradiction?

MEMBER BERNHEIM: So, I didn't want to bring this up earlier, but I think it is important. The measure is a risk-standardized readmission rate and has an interval estimate and hospitals have a point estimate, but also in terms of being identified as whether they are high- or low-performers, it matters whether their interval estimate crosses the national rate.

The way that ACA was written, it uses that ratio without accounting for the interval estimate. And hospitals that are on one side or the other of one for their ratio, that determines their penalty, and how far away they are determines how big the penalty is.

And there's three measures that you can be on one side or the other for one. So, many more hospitals get penalized because all you have to do is be on one side or the other of one. And so, part of what may be accounting for that is whether the accumulation of those three measures is differentially adding up on those.

Does that make sense? I'm trying to talk fast. But it is an issue of both the three measures and the fact that they don't take any account of the interval estimate around the measure when they assign the penalties.

MEMBER WERNER: I don't know if

this is what Larry was asking, but I am not sure how that explains why the MedPAC report seemed to find that, when you adjust or when you stratify by SSI category, it seemed to make a very big difference in penalties; whereas --

CO-CHAIR NERENZ: No. I can speak to that; actually, our expert staff person.

Essentially, that is by design, meaning that if you create a model that, then, applies the penalty based on a certain cutpoint within each decile, you have almost guaranteed that the number of hospitals or percent penalized within each decile is going to be the same. I mean, it is not an empirical finding in the same way that some of your analyses are.

So, in that sense, I didn't see a contradiction, actually, between the two. In your situations, and I think in this one, the analyses do not make any intentional change in the application of the penalty or a

calculation of, say, a per-capita cost decile. Those are just left as they are, and you say, how does the movement of hospitals across those deciles change if you add or don't add the variable?

But that is not what the MedPAC analysis did. The MedPAC analysis actually said, first of all, let's group hospitals according to this particular measure. Now, just intentionally and by design, let's apply the penalty in a different way. Let's apply it within decile rather than as we currently do it.

And you have essentially guaranteed, then, as an illustration, that the number of hospitals penalized, the percent hospitals where the impact as a percent revenue within each decile is now the same, or close to the same, not different.

So, again, I actually saw no contradiction among these.

MEMBER CASALINO: Well, I am still

not sure I entirely grasp. I'm closer to grasping the technical side. But, the contradiction conceptually, you know, in my mind, is based on what Susannah showed and to some extent Sajid. One wouldn't think that MedPAC should recommend what MedPAC recommended, and yet, you did. And so, why?

CO-CHAIR NERENZ: Well, I mean, part of it is that we were based on a certain set of findings and analyses, but I think it still stands as valid that in the current application penalty -- and particularly as Susannah just said -- it is not just the heart failure component; it is actually the way the measure is actually constructed with the three different clinical groups. It simply is a matter of fact, I think, unless there is some technical error, that hospitals in the highest decile, meaning the highest percentage of SSI, were four times as likely to get the maximum penalty as CMS currently applies as it as those in the lowest decile. It is just what

the data show.

And again, I don't think that essentially contradicts anything in your dataset. It is just a way of manipulating the numbers. But, as far as I know, the numbers are correct.

MEMBER BERNHEIM: So, was the confusion around, once there was risk adjustment, why it was -- once they were stratified, why it was equal, or more the preliminary findings where I am showing that, when you assess by quintiles, there is not a huge difference between two groups? And when they are dividing by deciles, they are seeing bigger differences among the groups? Which of those two things were the confusion?

MEMBER CASALINO: The latter.

MEMBER BERNHEIM: Okay. That's what I thought. So, it is not so much about why the stratification works.

So, I think if you take the two overlapping histograms, the way the penalty is

set up is that it essentially takes half the hospitals because the model is made in a way that about half the hospitals are going to have a ratio that is greater than one, and about half the hospitals are going to have a ratio that is less than one.

When you, then, look at high-SSI hospitals, which we didn't do exactly, on any given measure, those hospitals have a slightly higher than 50-percent rate. And I can tell you on the heart failure measure I think it is 57 percent, but I can go back and look. I have it.

But, then, there are three measures that are put into the formula. So, I suspect -- I haven't done this math -- but I suspect the reason they are getting numbers that are even higher than the 50-some percent is that, if you also have 50 percent of those hospitals getting penalized for the pneumonia measures, and those aren't the same ones, you start to build these differentials where I

think that it kind of adds up. And I think that that is why in that top decile you're seeing higher percentages in the original stratification than you would expect from my histograms, because I was looking at a single measure. Does that make sense? I mean, I don't know that, but that is my best guess about why those are different.

CO-CHAIR NERENZ: The geeks among us could enjoy a long discussion about this. We may have to take this offline and see if there is something that we can come up with that is relevant to the group. We can bring it back sometime in the next couple of days. But we probably should move on to our next panel and make sure that we don't delay lunch to the point that people faint and have other problems.

MS. PACE: Okay. So, we are going to move on to our panel. We are going to just ask you to present from your places there. You each have a microphone. Suzanne will

advance the slides.

And we really do need to ask you to stick to your five minutes. I know we said five to six, but we want you to try to do it in five.

Again, we wanted these presentations, again, to offer some information for us to be kind of thinking about. We are not really going to discuss each of the presentations. You know, this will carry us over into this afternoon's discussions, because some of these will be illustrations that we want to think about as we start really delving into the issues.

So, with that, I think we will go ahead and start with Monica.

MEMBER BHAREL: Great. Thank you.

So, in my four minutes, I'm going to try and express a lot of information, but there's tons of data behind this. I cheated; I put two slides in one. So, I am going to talk fast and go through a lot of information.

I am going to have to skip over clinical models, which are important to understand the context of this, and I'm going to have to skip over a lot of the ways Massachusetts is different than the rest of the country. But I can get back to that offline or in this presentation, or you may already know that in all ways.

So, one of the points of focusing on homeless is, one, to see if it, indeed, should be looked at independently of these other more traditional socioeconomic risk factors, but also because it is an extreme case, and extreme cases can teach us a lot.

So, with that, let me just go to the next slide, please.

So, abject poverty covers a lot of these SES measures that we are talking about. We are talking about the extremes of poverty in this case.

If you look at the left side of the slide, you will see that, in addition to

the standard SES measures that we are used to thinking about, we are also talking about, in addition to that, compounding that, the nexus of lack of consistent shelter, violence and trauma that is disproportionate to any other population. As an example, 96 percent of homeless women have had some kind of violence or trauma experience. And then, even more disproportionate absence of healthy food. Thinking about this in terms of health and healthcare, you move to the right side of the slide. All of those points that I list there have data behind them in terms of where this is specifically an issue for homeless individuals and, also, they fit into all the different categories that we are thinking about not just healthcare, but environment as well as behaviors, et cetera.

So, keeping that in mind, if you can go to the next slide, please?

So, let's talk for a second about the higher morbidity among homeless

individuals, in an attempt to look at pathways. So, if you look at the left slide, this is Medicaid data, Massachusetts Medicaid data, and this is looking at disease prevalence in 2010 of 6500 individuals who we care for at Boston Healthcare for the Homeless.

So, you will see profound burden of disease. Just to point out a couple, hepatitis C at 23 percent; the national average is 1.8; mental illness at 68 percent; substance use, 60 percent of the entire population. And even common diseases, such as diabetes, is 18 percent compared to a level of 8 percent in the general population.

So, when you look at that, I want to point out one thing that has come up this morning, and that is about data and where to get data. The reason we fed into the system our Healthcare for the Homeless patients was because it is not well-collected, this issue of how -- and this has come up -- who is

homeless. But I must say that it is possible to collect.

The second thing is we talked in some of the previous discussions about ICD-10 and coding. So, the hepatitis C rate, which is 23, so that is a quarter of our population, when we do more extensive chart reviews -- we took a sample of a thousand random charts -- their percentage was actually 40 percent. So, even this percentage I believe is an underrepresentation of what is actually happening in many of these categories.

If you move to the right side of the chart, this is to give you some comparison data. The statewide number is for patients in our Massachusetts Medicaid PCC program. That is basically non-managed care, full Medicaid. And that's compared to a cohort of our patients from that larger sample, about 4,000, who are in our primary care panel.

And what I want to focus your attention on here is the DxCG score in

relation to thinking about risk and risk adjustment. So, a DxCG score in Massachusetts is used by Medicaid for some risk adjustment. And you'll see that the statewide group has a DxCG score of 1.5, and this is averaged out in 2010 to about 1 for general Medicaid patients. In the homeless individuals it is 3.4. So, some, but not all, of the risk is captured in this DxCG score, saying there's 3.4 times as much disease burden for the diseases that we measure.

Next slide, please.

Let's look at mortality for a second. I don't need to draw on the fact that there's premature mortality. In our most recent, Travis Baggett from our group, looking at death data, found the premature mortality average age to be 51. I draw your attention to drug overdose in the youngest group, which is nine times higher than the general population.

Next slide, please.

If you look at the left side of this first, please, so if you look at etiology of premature mortality, some of the generalizations about homeless individuals and mortality are related to substance use, as our last slide showed.

If you look at this -- and this I ask you not to share; it is a manuscript in preparation -- using some techniques of population attributable fractions that I can go into later, if you would like, the etiology of these premature mortalities was 52 percent substance use. But, then, what is in this unexplained mortality gap of 48 percent and what is the risk associated with that, and where is that? It is not known, but there is something there in that 48 percent. So, about half of them not explained by some of the common beliefs of risk of death.

If you look on the right side, you know, is homeless independently associated with death, the real truthful answer is we

don't know. These are all assumptions that we are making. But here is a little bit of data from Steve Hwang, who is now in Toronto, from our group at the time, looking at a hazard ratio of almost two times for staying in a shelter. And that is when you attribute -- everything else is matched. So, that is just by staying in the shelter, about two times more hazard ratio.

Next slide, please.

Just looking at the first bullet point for a second, so, you know, we are talking about how to tease out the issue of homelessness. So, if we look for a second at resource use as a proxy for disease burden, then this is the cost data from that same Medicaid data. And in this part of the analysis, we matched DxCG scores.

So, for example, we are looking at an individual who has, say, schizophrenia and diabetes who is housed and has Medicaid versus schizophrenia and diabetes who is homeless.

And to care for those individuals with the same matched DxCG scores, so standard ways of looking at risk as we now have it in the system, there was a cost differential of $210 more monthly to take care of the homeless individuals.

Does that get at an independent variable? Not quite, but it is a roundabout way to get at it.

So, what I am showing you here is that the morbidity and mortality data is suggestive. The clinical experience is more than suggestive that, when you take homelessness on top of the other factors that we are speaking about, that there is something that happens when you put those all together that the homeless compounds that is not being picked up in the current system.

And, you know, truthfully, to get direct causal data will be challenging, and the methodology for that is not available. But I ask us to look at, in this context that

we are talking now, homeless in two ways. One, to think about it as a group of individuals who we know are greatly affected by all of these issues that we're talking about, but that there is something above and beyond that can be measured. That should be something we should consider.

Thank you.

CO-CHAIR FISCELLA: Thank you, Monica.

I think we are going to defer all the questions until the end because we have six presenters and we are about 30 minutes behind. And that will also help for those of you who have multiple questions to multiple presenters.

So, let's go on to Thu.

MEMBER QUACH: Okay. Again, I am from Asian Health Services, and we are a Community Health Center that serves mostly Asian immigrants. A lot of them are limited English proficient patients. And so, we want

to consider LEP in the risk adjustment.

And I want to note that I am presenting also on behalf of Ninez. So, she will also help with some of the questions.

Next slide.

So, in terms of the question that we want to pose, we want to ask whether LEP, when it is added to the conventional risk adjusters, does it provide a better risk prediction tool? And in this conceptual model, we show several pathways in which LEP can affect the outcomes. So, you know, it can affect it through some of the underlying health, which would be captured in some of the comorbidities, the diagnosis data.

But it also really affects the process of care, the appropriate care that people get. And so, with that, it can limit access to care, affect patient/provider communication, which, then, can affect the outcomes.

But one thing I really want to

note is the enabling services that are provided. These are non-clinical services, like language interpretation, that really address some of these barriers. And so, when we are considering the data, we really want to consider enabling services down the line.

Next slide.

So, in terms of whether LEP data does exist, at least for the Community Health Centers, we are funded by HRA, and we report annually to UDS, the Uniform Data System. And among the variables that we report, one of the variables is patients best served in a language other than English. So, at the Community Health Center we have this data at the patient level and, thus, at the Community Health Center level.

Next slide.

So, what we did, we were modestly funded by the California Endowment to really do an exploratory study, more of a proof-in-concept study, and it is still in progress.

But what we did is we wanted to see whether we can take our patient data and use LEP, as well as poverty, but we focused more on LEP in this presentation, use our LEP status information and put it in the risk adjustment.

So, we actually got only a subset of our patients due to limited access to some of the data with the health plans. We got about 50 percent of our entire population, and this mostly focused on the Medicaid managed care group, the Healthy Families group.

But it is about almost 17,000 members, and we looked at the years of 2011 to 2012. I want to note that 89 percent of the ones included in there are of LEP status. So, it is really a high proportion of LEP.

In the analysis, we actually worked with Dr. Todd Gilmer from UC-San Diego, who is one of the co-developers of Chronic Illness and Disability Payment System, the CDPS, which it is very similar to an HCC model, and I think most of you are familiar

with it.

In terms of our data, we mostly looked at demographics, enrollment, diagnosis, and pharmacy data. And the scores, the CDPS scores, what it does is it accounts for age, gender, and diagnoses. And what we did is we added on the LEP status.

Next slide, please.

So, in terms of our results, here you see, on top, we looked at LEP stratified. So, you know, on the middle column, you have the LEP. And then, on the righthand side, you have the non-LEP risk scores. And it is broken up by the four aid categories: adult, children, disabled, and elderly.

And you can see that for LEP we are seeing lower risk scores when compared to a national benchmark. That national benchmark is mostly Medicaid, based on a Medicaid dataset relative to the non-LEP.

On the bottom, we added in LEP status as a risk adjuster, so in the model,

and it shows you for each of those aid categories as well as combined. And again, LEP compared to non-LEP, in our patient population you see that the LEP has a lower risk. I do want to note that the model that included LEP, the r-squared for it was slightly higher than the model without it.

One thing I do want to note here are the weights. So, a lot of the weights here, you know, the CDPS program weighted it, the diagnoses, to age and to sex. But, because it lacked information on language, it was not weighted on language. So, that is one of the big limitations to our results.

Next slide and final slide.

So, you know, again noting the fact that the data is not weighted by the LEP status. We couldn't do it for our data because it was quite small. Also, that we compared to a national benchmark rather than to California, because California really differs. So, we want to note that.

In terms of outcomes, you know, a lot of the risk adjustment is so based on diagnoses data, and we know that with a lot of these populations, as well as other disadvantaged populations, they face a lot of barriers. So, this issue on underutilization and underdiagnosis is major, and we really want to underscore that point.

When it comes to our risk adjuster, LEP as a risk adjuster, we want to note about the selection bias. We did an internal comparison in that analysis, and we are comparing it to non-LEP within our patient population. Well, who are these non-LEP coming to a health center that mostly provides language services needs to be considered, right?

Data limitations for us, we couldn't get all of our patient data. So, it is subsetted.

And then, the issue is that we didn't have hospital and mental health data.

For the stratified, we compared to the national benchmark, which did have that. So, that needs to be accounted for as well.

The issue of stratification versus risk adjuster, I think we will continue to have that discussion. But a major point I want to make here is in our model and in our analysis, at Asian Health Services we provide language services and a whole bunch of other enabling services universally to our patients. So, if you are looking at that and you are not considering the enabling services in the model, then you really are going to downward-bias your analysis.

So, it is something that, you know, it is not just LEP, but, as we consider other social factors like homelessness, we really do need to consider what the providers, what these primary care providers are providing in terms of enabling services, because it is already addressing that pathway. And by not adjusting for that, by not

controlling for that, you may have not the best accurate results when you are looking at these things.

CO-CHAIR FISCELLA: Thank you, Thu. Very succinct.

Tia?

MEMBER SAWHNEY: Okay, next slide.

Okay. This is kind of like field notes from someone who plays with data. So, when we talk about, and it is noted in some of the reading, there is difference between health and healthcare. Within health and healthcare, there is a difference between incidence and prognosis. And this came up in the presentation regarding hospital care. Once the health event begins, that is one part of the path, but, then, who is at risk for the health event to begin? And I think we always need to be thinking along those lines because it has a big impact on the models that you build.

A classic case in an SES

adjustment. People with no diagnostic history, two young men 23 years old, one is on the streets in the south side of Chicago and one is a student of U of C. Which one has a different risk -- you know, is there a different risk profile? Yes, you'd better believe it.

And, in fact, the one who at U of C and going to student health may actually have a diagnostic history that the kid on the streets of Chicago doesn't have. But I know which one, coming from an insurance background, I would rather be insuring.

Traditional risk adjustment looks at age, sex, and diagnostic history, and pharmaceutical history, which is really a proxy for diagnostic history. So, it is really limited for those who don't have a diagnostic history. The takeaway is you really need to think of what we are adjusting for and what the risks are that we are adjusting for.

Next slide.

It is usually not just diagnostic history. It is usually one year of diagnostic history, which I think is also reflective of regular contact with the healthcare system. And that is another thing we all need to be thinking about.

So, some research that I did, and I'm not going to spend a lot of time on it because it is not necessarily as applicable in this forum because it is total healthcare cost, and it is risk-adjusted for SES in that context.

But I did look at income as a marginal variable after traditional risk adjustment in order to predict total healthcare cost. And I did find that there is a relationship after adjusted for age, sex, and diagnosis between SES and total cost.

And now, it levels off. It is the difference between what MEPS would define as poor, near poor, middle class, and high not

class but income, high income. Now their definition of high income is not all that high. It is 400 percent of the federal poverty level.

But, actually, even the middle income and the high were relatively flat to each other. The gradient really seems to be between poor, near poor, and middle.

And then, my theory, unproven, is that there isn't as much of a gradient between middle and high, and that's because, whereas health clearly continues to improve as you go up the spectrum, so does the sophistication of the demands that people make on the healthcare system. So, health improves, but costs go up because -- well, we all go to the doctor; we are pretty demanding customers, right?

(Laughter.)

Okay. There's a lot of problem in trying to look at SES because, historically, it just hasn't been systematically captured and connected to the healthcare experience and

the cost, to the healthcare data and the health cost data. And every dataset is imperfect, including the one I used.

Next slide.

Another point I really want to make -- and it may cause some of your eyes to glaze over, and I'll make it very fast -- is that, when looking at models, focus on r-squared in the context of SES adjustment just doesn't work, especially when we are talking individual -- not necessarily hospital data, which we saw earlier, which has that nice bell-shaped curve. Individual-level data, it doesn't matter whether it is cost or whether it is prescriptions or number of doctor visits, or whatever. It is statistically ugly data. It has its huge density at zero. It has extreme outliers. It is skewed heteroscedastic, which I love that word; it is just so fun to say.

(Laughter.)

And until you can take a

population -- and the fact is it is a spread-out mess. That is a scientific term, a spread-out mess. And until you can start to differentiate one mess from the other, then the r-squareds don't come out, even if there are cost differences.

And I'm like, wait, how can there be a 20-percent cost difference between populations? And I was working with real data, and I was finding 20-percent cost differences between two populations, but there was no difference in r-squared. I mean, it was out in the nth decimal place.

And so, I started modeling it. I started with a population and, then, I modeled the different ways I could drive, artificially drive, a 20-percent difference. And I realized the fundamental problem was, until you can start creating two mountains instead of marginal changes to the first mountain, you just don't get the r-squared. So, we can't focus on r-squareds.

Go on.

The other word is a word from -- the last one is a word from the actuaries. This goes back, and I don't know that they were the first ones, either, but this is for 1996. Practical considerations for risk adjustment variables, I mean, you have to have the data for most patients. It has to be reliable. It can't be susceptible to gaming, and it has to be stable over time.

And that's just like so important, and that is one of the problems with homelessness because, yes, the man living under the bridge is clearly homeless, but there are a whole lot of gradients beyond that. And it is not necessarily stable over time.

Next. Done. Okay, cool.

(Laughter.)

MEMBER GARRETT: So, I am going to talk just briefly about some work that one of our physicians, Scott Davies, has been doing

at the Hennepin County Medical Center. And it really speaks to, if we decide that we are going to recommend that there should be some kind of adjustment for SES and sociodemographics. And this kind of gets into a bit of the questions we are going to have to answer with the "how," and part of that "how" is going to be, well, what's the definition of sociodemographics and SES.

And so, Dr. Davies has been challenging us to take a look at this -- and if you could do the next slide? -- and start to think about tobacco use. And is tobacco use really an outcome variable or is it actually more of a sociodemographic variable?

And so, just to tee this up, we use a measure in Minnesota called the D5. There are five components of this diabetes measure. It was NQF endorsed in 2010. And so, it is probably used in other places as well.

And one of those five components

is self-reported that you are tobacco-free. So, that is an outcome variable. And that makes a lot intuitive sense because, as we all know, which a huge impact smoking status has on our health.

And there certainly are things that we can do, as the healthcare system, to help people quit. And so, we are hoping that we are giving an incentive to use as providers to actually intervene, get people into the right cessation programs, and try and address tobacco use.

And so, go to the next slide.

So, the premise of having this in the D5 as one of the measures is, if you look on the left, if you go from an excellent clinic to a worse clinic, then your overall D5 score is going to go from high to low. And so, the idea is that those five components are going to measure provider performance.

Now, just as an illustration, if you look at the tobacco-free rates by four

actual real-live clinics within our system, the tobacco-free rate also kind of follows that curve. There is a lot of variation in that rate. And because it is an all-or-nothing, you have to hit each of those five measures in order to get a 1 in the numerator. If you have a clinic -- like we actually have a real clinic where the smoking rate is 70 percent, and we think that might be one of the highest rates within a clinic in the country. And so, our theoretical maximum for being able to achieve on that measure is very low.

And so, then, the question is, well, how amenable to change are smoking rates. And Dr. Davies, who is a pulmonologist, has done a lot of research on this. The very best, most expensive interventions maybe at a population you can see 3-percent decline a year. So, how amenable is that measure really to clinician intervention, is one of the questions that he is asking us to think about.

And then, another thing about smoking is that the most successful interventions have really been kind of at the community and public health level. And so, how do you factor that in, when you're trying to incent performance is another thing we have been thinking about.

Next slide, please.

And so, kind of as I have been saying, a lot of the future improvement will come from environmental efforts. And how much can you really do with individuals? And once you get into people who are smokers and have been for their whole lives, and have a lot of others who are going on, how much opportunity is there really to change?

Next slide, please.

So, you are all very familiar with this, but there is high correlation between smoking rates and other dissociative demographics. So, these are a couple of results from some surveys within Hennepin

County about associations with race and ethnicity and education. And there's just lots and lots of correlation here. So, again, it kind of raises the question, is this an outcome variable or more of a sociodemographic variable that we want to consider controlling for?

Next slide, please.

And we have done some multivariant analysis to understand what impacts are diabetes scores, and some of those results are here. We see things that improve the score are age, diagnosis of CAD, primary language other than English. Things that make it worse: younger age, race/ethnicity, some factors there, substance abuse and psychiatric illness, which are really huge in our population. So, it just really kind of has gotten us thinking about this question about variable versus control.

And then, the last slide is a bit of a different view of this. But we have been

collecting data on one of our Medicaid expansion populations with a tool we're calling the Life Cell Overview Survey. And these are the different types of factors that we are collecting in that. Tobacco use we are collecting in that as well because we are very much thinking about it as a key thing that we need to be working on.

And I know Steve had put an article in our packet about this Life Circumstances Index that he has proposed, which I think is a really interesting idea. It is kind of related to this idea of the lifestyle overview, that we are trying to understand all these different factors that impact health.

And one of the things that is at the very top of our list in terms of prevalence is social support. So, we have a couple of questions, including: how many people can you count on in times of need? Do you have a spouse or a partner? Are there any

adults, including spouse or partner, with whom you have regular talks?

And so, I also just want to throw out I think that that social support is something also that we believe is very highly correlated to ability to change health. And I think we should consider that as well, as we are thinking about definitions of sociodemographics.

Thank you.

CO-CHAIR FISCELLA: Norbert?

MEMBER GOLDFIELD: What I have been impressed by the conversation is that there are sort of two kinds of conversations that are going on. No. 1 is whether or not to incorporate SES data elements, and some presentations seem to present that we shouldn't. And then, others talk about different data elements that we should consider.

As already indicated in my initial remarks, I think low-income populations are so

discriminated against in this country, it is not a question of whether, but how.

And just as a way to kind of contrast our approach to the world, which is I have been only with the research group that developed the DRGs for 30 years, but Rich has been working with it since the beginning with Bob Fetter and John Thompson. And arguably, it is certainly the methodology that has had the greatest impact on healthcare policy, both in the United States and beyond.

I would just say that the way we look at the world is try to have these kinds of conversations and look at new data elements that we should be collecting. And so, typically, that might be I-9 or I-10 or additional data elements, such as homelessness and make as rigorous a definition as possible, and then, test it out in a large state. So, we are much more interested as much a possible in working with states to get at that.

And so, the point that I want to

say, building off this slide here, is, again -- I have said it already, and everybody seems to say it, but, then, we just say, well, poor people can be discriminated against and that's okay. You know, that these classification systems are going to be used for payment, and we have just got to hammer that home over and over again.

I also say that it is really important to try to use clinical data in extreme detail in ways that DRGs have pioneered and continue to pioneer. And I will give you an example of that in just a second.

I want to highlight, as a consequence, I tend to be very practical, you know, which is to say I like to specify the healthcare encounter question. Obviously, readmissions are different from complications.

In a positive way, Susannah and I look at the work of readmissions very differently. That is to say, we would look at readmissions the way the hospital thinks about

it, that is to say, at the point of discharge. And we would look at all the conditions at the point of discharge. And at least initially, unless we can specify which ones are complications, which is a separate one, we should include them.

So, we want to have as detailed and rich model as possible, understanding that we also want to look at it the way the hospital looks at, which is to say the hospital looks at readmissions as the point of discharge.

And then, of course, I have already said the issue of -- I think it is important, and that is certainly the reason that I'm here -- is the whole issue of a national/state strategy that, hopefully, can come out from NQF, in particular, with respect to homelessness.

The next slide.

The only thing that I want to say on this slide is that it is very important to

identify those individuals, as I will show on those slides, who are often with a lower socioeconomic status who have higher severity of illness. It sounds like an obvious thing, but I'm going to show an example why most of the models do not get into that.

Then, obviously, higher payment will minimize adverse selection. And that is constantly an issue that we face, that everybody has highlighted.

I want to just point out with respect to what kind of items to include, DRGs are a categorical risk model. As I was joking with Pam Owens, that people have access to the APR DRG Manual which is used in some of the AHRQ QI Indicators and this 5,000 pages of detailed model that people can look at. And I think, from an item point of view, it is important to try to identify those clinical variables that have the least gamability.

Lastly, on timing, partly because Karen asked me to comment on that, it really

should not impact the classification from our perspective. The results will be different, obviously, for readmission at two versus four weeks, but the classification should be the same.

Next slide.

So, I want to give an example here. So, a patient with cerebral palsy needs to be stratified by severity of illness. Of course, most models don't even have cerebral palsy identified. So, putting that aside, we have different categories not only for that, but categories for patients when we are looking at a year's period of time in terms of that dependent variable; those patients who are in foster care. And that is an interesting question, as to whether or not to use a use variable such as foster care as an SES variable. That is, yes, just to raise the issue. And I met with foster care providers for over two years on that very issue.

But, without this detail, the

approach to risk categorization, it is inevitable, and we know it -- we talk about it, but, then, we ignore it -- that poor people will just be ignored. And that managed organizations will assiduously create any kind of risk incentive to avoid these patients.

Next slide.

So, here's a bottom line. And so, Nancy brought up -- I think it was Nancy who brought up -- the issue of the patient with diabetes and schizophrenia. So, you could replace this issue of diabetes and schizophrenia, of CHF. And so, what you have here on this slide is patients on the top slide who only have diabetes as a chronic illness. And these are four levels of severity. So, that is their only major chronic illness. And on the bottom, rather, row is patients who have diabetes, COPD, and CHF, and those are levels of severity. That could be replaced by schizophrenia, also, instead of COPD.

So, I think it is that kind of detail that, for example, New York State and Texas have gotten into with respect to looking at paying for better outcomes. Because, in fact, if you want to recognize certain aspects of socioeconomic status, at a minimum, we know that diabetes who are schizophrenics, you know, those are very different patients. And we really need to look and stratify by severity.

Now, again, we know that that will not be 100 percent of the variation, but I am already accepting and wanting to test out, for example, homelessness.

So, on the next slide -- and that will be my last slide, actually, because we are really not supposed to talk about payment -- there are three different types of clinical data that can be incorporated into risk adjustment, data that is available today. So, actually, bottomless index is actually not a terrible piece of information and foster care

is actually also available.

We are the developer of the I-10 procedure classification system. That will actually make some significant impact on the types of information that we can have available starting next year.

There's data that is available for some individuals, but not reliably collected. And I have raised that homelessness, and I am not going to get into it.

Data that is not generally available, but should be available in the next three to five years that I am hoping that the panel will get into that. So, for example, patient-derived health status, that is already available for certain PPSes such as the home health. Incarceration, there are some linked databases in New York State and other states that we can do.

I am particularly a fan myself -- and I'll put out my point here -- it is that I think that patient activation or

empowerment, I am a much bigger fan of that as opposed to something like English language proficiency.

My last comment is that I think we need to have humility, folks. I mean, I think it is in short supply. And so, we talk about CHF, and I obsess about CHF. And we at the same time forget that there's maybe an increasing body of literature that looks at the relationship between readmissions and mortality.

So, I think that there is a lot that we need to learn while we try to put in this issue of SES. And I am hoping that this panel will have a clear, or as clear as possible, strategy as to how we can incorporate SES, not whether.

CO-CHAIR FISCELLA: Thank you, Norbert.

We are going to take public comments first. And then, we can make a decision as to whether to do comments before

lunch or after lunch.

MS. PACE: We will see, first of all, if there is anyone here present in the room that wants to do public comment. If you would come up to the microphone?

And then, we will go to the phone.

MR. SHAW: Hi. I'm John Shaw from Next Wave in Albany, New York, and we're a health services researcher and interested in this issue for quite a while.

One of the things that helps inform the discussion is what happened last week. There was a discussion of the population health framework down here. And they suggested one thing that would be helpful in discussing these new, complicated topics is to make the implicit assumptions explicit.

And two things come to mind relative to this. One is the big controversial measures are either all after discharge for readmissions or for the cost measure 80 percent of the variation was after

discharge.

So, what we are really doing is assuming that the patient, their informal caregivers, and their local community are capable to understand what to do and have the resources to be able to do it.

We have had a few examples where someone that is homeless versus someone who is a Wall Street executive might have differences there. The current system assumes that they are both the same. The current system also assumes that the measures that we are looking at that are constrained by the data we have for Medicare is all of the data that there is.

And we don't collect much of the data after discharge on what's really going on. We're starting to. The patient activation, the social supports at home are critical factors, as well as some of the population data where the provider may be situated.

And so, what may be useful in

having a complex situation a little bit simpler to see is make sure that, if we are making assumptions, we have them listed explicitly on the table, not just assumed to be the same.

Thank you.

MS. PACE: Is there anyone else on the conference call line that would like to make a comment?

MEMBER CALLAHAN: This is Mary Beth Callahan.

MS. PACE: Oh, thank you, Mary Beth.

And, Marshall, also feel free.

MEMBER CALLAHAN: I just wanted to kind of make a connection between something Dr. Goldfield said and something Nancy Garrett said, and futuristically thinking, I think Dr. Goldfield said there could be data to be collected in two to three years. I think he was maybe referring to patient-perceived quality-of-life data in terms of empowerment

and such. And I don't want to put words into his mouth.

But, then, Nancy Garrett, I believe, related that -- and I believe also that -- social support is very important in how illness is going to play out and can mediate factors of socioeconomic determinants in ways that I don't fully understand.

And so, I just think that is an interesting concept, and I don't really know how to put it forward, but I just wanted to connect those two thoughts.

Thank you.

MS. PACE: Okay. Thank you.

And anyone else in the audience here?

(No response.)

Yes, Operator, would you open the lines? Maybe that is part of the problem.

THE OPERATOR: At this time, in order to ask a question, press *, then the number 1 on your telephone keypad. That's *,

then the number 1 on your telephone keypad.

MS. PACE: Would you turn off your microphones?

THE OPERATOR: At this time there are no questions.

CO-CHAIR FISCELLA: We can do clarifying questions now before lunch or come back and do it after lunch.

MS. PACE: We can do five minutes of clarifying questions and, then, we can go for lunch.

CO-CHAIR FISCELLA: Okay, we'll do it.

Questions?

MEMBER NUCCIO: Yes. I have a question. I'm sorry. I had a question for Nancy.

Nancy, I noticed on one of your charts there was a curvilinear relationship between some of your variables and your outcome, and there was linear in the other cases. Could you talk a little bit about

that, especially where smoking appeared to be curvilinear? That is, it went down and came back up.

MEMBER GARRETT: So, the first example was really more theoretical, showing differences across four different types of clinics. Either you could consider them good to bad or, then, I was kind of putting that next to here's how smoking status actually looks in our clinic. So, that was more theoretical. And then, the other one was a regression analysis, which was a linear look at things. So, does that help?

MEMBER PONCE: For Norbert Goldfield, please help me understand the rationale why patient activation over LEP is preferred.

MEMBER GOLDFIELD: That is a great question. And at the end of the day, there are going to be patients with limited English proficiency -- I, myself, speak Spanish with an Italian accent -- who are very activated.

In fact, I am not sure if in one of these -- oh, it was a different document.

I think it was a document that was posted in one of my blogs where I spoke about a patient who has no English language proficiency and has significant intellectual disability, but over a period of time became extremely activated and empowered, and her diabetes is very well-controlled. That doesn't mean she doesn't have other issues, you know, but I am just talking about the outcome of diabetes.

So, I guess just from my own clinical background, and just from reading the literature that has been pioneered by Judy Hibbard, Kate Lorig, and John Watson, I believe that these items can transcend, shall we say, limited English language proficiency. So, that is how I look at it.

Then, I was just going to comment on the question that was posed on the phone, if that would be okay?

So, I just want to say that, with respect to social situation, I consider that just as important as activation. In fact, a study that we are doing right now for a federal agency, in OASIS they do collect living alone, right? And so, I think a schizophrenic who is living alone as opposed to a schizophrenic who is living with his or her family, again, I'm not talking rocket science, folks. And we have a linked database that looks at that. That should be something that should be tested and moved.

So, I definitely accept, and, in fact, I am very excited about, this project that we are doing for a federal agency that looks at that specifically for the severely mentally-disabled.

MEMBER GARRETT: So, a question for you, Norbert. You talked about that the current risk-adjustment tools, clinical risk-adjustment tools, don't do a good job with severity. So, can you tell us a bit more with

the existing data that we have, with existing diagnosis code systems and procedure code systems, do we have the ability to improve that or does that really require moving on to new types of data?

MEMBER GOLDFIELD: The short version is both, which is to say, for example, what New York State uses for looking at outcomes and payment for year-long patient-based episodes has a thousand categories and has very detailed categories specifically for children.

That said, I spent literally maybe 50 hours with foster care providers, and they finally positively beat me over the head that we need a separate category for foster care. And I highlight that because it is actually interesting from an intellectual perspective.

So, I believe that you have to start somewhere. That is the whole premise of DRGs; you have to start somewhere. But there is a lot more detail that can be captured, and

I just tried to give one example.

At the same time, we should absolutely -- and I think all of us here, this is just an incredible experience from everybody that is here -- that we could really lead the way in terms of setting out an agenda as to what kind of data elements should be collected for extremely discriminated-against individuals.

MS. PACE: All right. Lunch should be ready. So, feel free to get up and take a break and grab your lunch. I think we will be able to make up some time. So, let's just plan to reconvene at 1:15. Maybe try to get back to your seats at 1:10 and we'll proceed from there.

Basically, you will have to bring your food back to your place. There are a few empty seats in the other area. The buffet is in the back.

(Whereupon, the foregoing matter went off the record at 12:41 p.m. and went

back on the record at 1:15 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:15 p.m.

MS. PACE: Hi, Kate. Welcome. Sorry I didn't see you when you first came in. And I don't know if you want to make some remarks now. If you are prepared, we can do that before we start the next panel.

MS. GOODRICH: Sure.

MS. PACE: So, this is Kate Goodrich from CMS.

MS. GOODRICH: Hi there.

So, I am Kate Goodrich. I am the Director of the Quality Measurement and Health Assessment Group at CMS. And not only does my group oversee a lot of the quality measurement, public reporting, pay-for-performance programs, we do measure development. We work with Suzanne and her team and other folks on measure development.

We also oversee the work under the HHS contract with the National Quality Forum, and really saw over the last year sort of the

opportunity to tie in, to basically leverage NQF's convening function to address some of these issues that we are discussing today.

So, as you all know, CMS has not traditionally adjusted for race or socioeconomic factors or other types of related factors. And we also don't in most of our programs have the authority to be able to address those factors through stratification of payment or anything like that.

So, as we started to implement these outcome-based measures, thinking the mortality, readmission, and even our cost measures in our pay-for-reporting and, then, ultimately, pay-for-performance or pay-for-value-type programs, like Hospital Value-Based Purchasing, like the Hospital Readmissions Reduction Program, there has been a lot more attention given, obviously, to the measures and how they're constructed, but around this particular issue that we are discussing today, and the fact that within the measure we have

not traditionally accounted for these factors.

And I think over the last few years a lot has been written, a lot has been said about how CMS and, presumably, other payers should account for these factors or how we should handle this within the measures or within our payment systems. And so, we really saw this as an opportunity to use the NQF convening function to address this head-on in really an evidence-based, data-driven kind of way, where we have evidence and data.

Because, again, I do think a lot has been said. There has been a lot of real concerns and perceived concerns. So, we felt that it made a lot of sense to just get it all out on the table, but, again, in as much as possible, data-driven, and evidence-based way. So that we can have a really smart discussion, hear all the viewpoints.

And I think what would be helpful for us at the end of the day is going to be some, essentially, principles around this

issue. I don't think we necessarily are looking for or think we can get out of two days specific direction, but just having those principles that are agreed upon by a Committee such as this, and this is a phenomenal Committee, I think would really, really help, not only us because my group, we oversee the Medicare Fee-for-Service programs. I think it would help other components within CMS. And certainly, it would help the private sector as well.

So, I am very thrilled to be here. I wish I could be here all two days, but I will be here this afternoon. So, I'm looking forward to it.

Thank you.

MS. PACE: Thank you.

CO-CHAIR NERENZ: Okay. I think our next session, essentially, continues a couple of the main themes we had before lunch. We are talking about examples of inclusion of one or more SES variables in an adjustment

model and what effect that has. We have a few examples.

As we did before, we will try to work through very quickly, five minutes at a time. I'll try to be the enforcer on that, but people have been pretty good. So, I haven't had to be too strict.

I don't think I need to say anything more by way of overview. We will do some questions after.

Atul, it's all yours.

MEMBER GROVER: You can go ahead and advance it one slide.

Actually, this paper just came out online on Monday in HSR, and we have made the language available. I am happy to share it, if you ask me any questions. I am going to look at the paper because I didn't actually write it. My colleagues at AAMC and AHA did it with Hugh and Kaynig.

Next slide, please.

You know, interestingly, even

during the whole readmissions discussion where CMS did have the ability to stratify but chose not to, it was interesting to watch the presentations from their data shop and Brennan presenting how readmissions vary across HRRs, and really sort of laying out the striking difference between the place with lowest readmissions in the country -- anybody care to guess? -- Idaho Falls, Idaho, and then, the place with the highest readmissions in the country -- care to guess? -- Chicago, and saying, "Gee, you would think people would understand Chicago by now."

And what CMS's data shop showed was essentially the biggest correlations there were that in Chicago you had 33-percent dual-eligibles. That was about 16 percent in Idaho Falls. Idaho Falls is also 95-percent White non-Hispanic. So, if you didn't know there were places like that that still existed, there are. And Chicago is, of course, a majority/minority city at 60-some percent

minority.

So, what we did was, looking at proxies for SES and really focusing on dual status with the Medicare Program, looking to see if we could find differences, as has been pointed out. And I know that those differences seem small in many of the analyses that have been presented here, but they do make a difference in terms of payment policy. And I think that is one of the things we wanted to highlight.

And what my colleagues found was that duals were certainly more likely to be readmitted to a hospital within 30 days after discharge, even after adjusting for age, sex, and comorbidities.

What was also interesting is that the share of patients discharged by a hospital that were duals also seemed to have an effect and appeared to work as a proxy beyond just looking at those individual patients. And again, I think some of that has been discussed

today, as, you know, does that reflect other conditions in the catchment area of a hospital?

I will also note that what my colleagues found was that in those areas, those hospitals that served high-percentage dual populations, they also had more admissions that were tied to ambulatory-care-sensitive conditions. So, clearly, there is something going on in the ecosystem of that neighborhood that may not be fully adjusted for if you just look at the status of that one dual patient.

And we know that hospitals with higher shares of duals are, then, disproportionately penalized under the readmissions program. And interestingly, of the hospitals with the highest quartile dual shares, over half had negative total profit margins in 2008 and 2009 compared with only 20 percent of the lowest quartile. So, again, are we at risk of entrenching disparities

because of removing resources from those that serve the most vulnerable?

Next slide, please.

And again, you know, looking at the absolute change in readmission rates is one thing, but if you look at what this means from a payment policy perspective, we know that comparing hospitals that have the lowest share of duals in their discharges and those in the highest quartile and those that have no reduction from the readmissions program, 23 percent in the lowest quartile, 10 percent in the highest quartile.

If you look at the maximum penalty of 2 to 3 percent, we project using the three existing conditions, and then adding in COPD, CABG, PTCA, and other vascular conditions, that, again, in the highest quartile 10.5 percent of the hospitals will have the highest penalty; whereas, that is only less than 6 percent in the lowest quartile.

So, the next and final slide,

please.

So, what my colleagues did was, then, within the regression, hierarchical regression, try and adjust for nothing. So, looking at comparing, say, heart failure, which is circled here, in terms of the gap between those with excessive readmission rates, in the highest quartile it was about 61 percent and 41 percent in the lowest quartile. When you adjusted for the individual dual status, you equalized that a little bit at 43 and 57. But when you adjust for both individual-level characteristics as well as the hospital characteristics of having a larger share of duals, that equalized almost completely, 49 percent and 50 percent.

So, again, I can't tell you exactly what it is we're measuring. I think we have seen some examples here. But maybe looking at a way to stratify based upon that population is the right way to go.

I'll stop there, and I think

people will have similar presentations.

CO-CHAIR NERENZ: Are there some very quick clarifying questions before we move off? Because we will cycle back. We have got a lot of discussion block coming.

MEMBER BERNHEIM: I'm just curious what the readmission measures they were using were. I mean, I know the conditions, but were they using the CMS --

MEMBER GROVER: They were using the CMS.

MEMBER BERNHEIM: Okay. Because for some of those conditions, they are not out there yet. Maybe I can just look at the paper now that it is out and try to understand.

MEMBER LIPSTEIN: That's his slide.

MEMBER GROVER: That's my slide.

MEMBER LIPSTEIN: Oh, did you want me to start my presentation. Oh, okay. Okay.

(Laughter.)

So, if there's anything that you

take away from these slides, what I guess I would hope you would write down, especially since I think Kate just mentioned empirical data, is I hope you would think Census tracts with high housing vacancy rates. That is the first thing to remember. The second thing is patient discharged to nursing homes. And the third thing is, yes or no, is the hospital located in a local taxing district for a regional safety net, yes or no?

And so, I am going to talk about each of those three variables. I picked those three because there's not a lot in the literature that was sent out to all of us, No. 1. And No. 2 is that data is available at a national level. And so, those become really kind of important indicators.

I want to start, before I just jump into the slides, because my slides go really quickly, to kind of give a framework for this. And I am going to use a non-readmission example real quickly.

In St. Louis, Barnes-Jewish Hospital is the only hospital in the city limits that still delivers babies. And we used to think that a good outcome was when we would discharge a healthy baby into the community.

We have now broadened the definition of outcome. And we now know that, in addition to discharging a healthy baby, we want to help that healthy baby get to third grade. And we want them to get to third grade fully immunized with complete eye and dental care and with a Body Mass Index appropriate to their age and height, and here's the kicker, reading on grade level.

And the reason we want to do that by third grade is we know that, if you get to third grade with those health indicators and reading on grade level, there's a much higher likelihood that you're going to graduate from high school. And if you do graduate from high school, we know that that affects life

expectancy and mortality rates long-term.

And so, if we want a good outcome for this healthy baby born at Barnes-Jewish Hospital, the definition of the outcome has to go beyond what happened at discharge. Okay? And so, one of the things I like about readmission rate is it really has focused our nation and our community on what happens to the patient after they leave the hospital. And that's a good thing.

Now fast forward. I shared with you earlier that I serve on the Board of the Patient-Centered Outcomes Research Institute. And at my very first dinner meeting I sat next to Harlan, Dr. Krumholz, from the Center for Outcomes Research and Effectiveness. He was having a glass of wine; I was probably having a Budweiser since I'm from St. Louis.

(Laughter.)

And that was when Harlan first told me the conclusion that Susannah shared with us earlier, that socioeconomic status

isn't determinative of hospital outcomes, hospital performance with regard to readmissions. Did I say that about right?

And so, I looked at him and I thought, what is he drinking?

(Laughter.)

Because at BJC we have 12 hospitals, and our two biggest disproportionate share hospitals have always had the biggest challenge on readmission rates. And so, I was trying to figure out, going back, then, and really looking and beginning to read the literature that was stimulated by my discussion with Dr. Krumholz, what do the national people who study this, the researchers in this room, what don't they know about St. Louis? Okay? Or what don't they know about BJC patients? Or what don't they know about patient-reported outcomes?

And that's when I began to separate in my mind socioeconomic status from difficult life circumstances. And that's why

-- I'm glad, Nancy, you're probably the only person, other than me, who has read that paper -- why I wrote a paper on life circumstances, and a Life Circumstances Index is probably just as important as a case mix index in determining patient outcomes.

So, setting about this journey on how to figure out what is it that really happens in St. Louis that people at the national level don't know about that makes the readmission rate challenge so much greater at Barnes-Jewish Hospital and Christian Hospital, our two big safety-net providers, compared to the other 10 BJC hospitals.

And so, here's what we learned. If you go to the next slide, what we learned is that, if you would just remove -- within BJC, so this is not kind of a national scientific study, just within BJC -- if we took the patients out of the data that were either discharged to Census tracts with high unit vacancy rates or we took out the patients

that were discharged to nursing homes, and I am making a real big presumption here. They were discharged to nursing homes either because they weren't well enough to go home or because they couldn't go home because there wasn't really a home to go to. If you those two out of the database, Barnes-Jewish and Christian had readmission rates that were about the same as everybody else within BJC.

So, I said, okay, nobody's going to look at just data that is only applicable to BJC. So, how do we begin to find out what really happened?

And so, what we did is an analysis for the whole State of Missouri. And what I like about this analysis at the State level is it is the same Medicaid program throughout the State.

So, for example, when Susannah was talking about Medicaid programs, as you know, eligibility for Medicaid is differential across the entire United States, and Medicaid

programs pay providers differently across the United States. So, Medicaid eligibility or not, it is not a great determinant of individual ability to manage outside the hospital environment.

And, David, when you brought up SSI indicators earlier, what we have also learned is that a poor person in Chapel Hill, North Carolina, is not the same as a poor person in East St. Louis, even if they have the same incomes. Because the person in Chapel Hill has access to a high-density environment where there are social support mechanisms. There are restaurants. There are grocery stores. There are drugstores. There are laundromats. There are taxicabs. There is a bus service. In East St. Louis, if you just come with me, I think you will see the difference.

And so, what we did is -- if you go to the next slide? -- what you will see is we took that data that is readily available to

hospitals and contains some factors that are not typically in administrative or claims data. And we recognize the limitations, that when you look at Census tract data, it could reflect either individual or neighborhood effects. And it doesn't capture all the social factors that we have talked about today.

But, when we ran the analysis -- next slide -- we compared the replicated CMS models alone to the models using the Census tract variables. And the one that we really haven't talked a lot about today is this high-vacancy housing unit variable. And the data came from all over the Missouri hospitals, and the Census tract variables from the analytics.

Next slide.

And you will see, for the three conditions, that what really shows is variability gets condensed.

Go to the next one.

And this is myocardial infarction.

And the next one.

And so, this data is online. But you will see that what CMS is trying to do is reduce variation in readmission rates. And so, housing unit vacancy and Census tract variables, not Medicaid, not income, not taking anything away from dual-eligibles, can do this at a statewide level, and the data is available nationally.

And so, one of the things that, when you go outside the State of Missouri -- and this is really the last important point I'll make -- St. Louis City -- I was in Baltimore when Baltimore closed Baltimore City Hospital. And then, I went to Chicago, where we had Cook County. And then, in St. Louis, St. Louis Regional Hospital closed in 1997. So, we don't have a publicly-funded safety-net environment at the local level. They do in Kansas City with the Truman Medical Center. They do in a number of geographic areas.

But especially in the rural communities of our State, there are not local taxing districts to support a safety net. And it makes a big difference in what happens to a patient when they leave the hospital.

What I guess I am hoping is, if you were to look at readmission rate or Value-Based Purchasing data, and you were to look at patients discharged to high-vacancy, high-housing-vacancy residential neighborhoods, patients discharged to nursing homes because they either can't go home or they are not well enough to go home, or patients discharged into a community that has local tax-based support for a safety net, it has a pretty big influence on outcomes.

And so, I guess I will just leave you with this final thought, because I always think it all the time. If you take the Henry Ford Medical Group and you move them to Scottsdale, Arizona, would they still pay one of the highest readmission penalties in the

United States? And if you took the Mayo Clinic folks, arguably, among the best doctors in the world, out of Scottsdale and you moved them to Detroit, would they still pay no penalty?

And so, at the provider level there is a credibility issue here when we think about rewards and lack of rewards or punishments or lack of punishments for outcomes that are not adjusted for socioeconomic status.

My biggest concern is that, by not adjusting for socioeconomic status, not only are you entrenching disparities, but you will make them permanent. Because what will happen is, it isn't the providers will avoid the poor and the uninsured; it is just that they will invest their capital in places where they can be successful. And if they know that there is going to be a higher degree of difficulty and a financial penalty associated with investing their capital in neighborhoods, Census tracts

with high-housing-vacancy rates, they just won't invest their services there. They won't grow their services there.

The readmission rate for hospitals in East St. Louis is now zero, as is the admission rate, because Kenneth Hall Hospital has closed. It isn't just because of the lack of socioeconomic adjustment, but is reflective of an environment where the federally-funded safety nets, as opposed to the locally-funded safety nets, the federally-funded safety nets are facing challenges on the Medicare front, challenges on the Medicaid front, especially in those states that are not taking the expansion. They are facing financial penalties on readmissions and Value-Based Purchasing, and they weren't financially strong to begin with.

And so, one of the things we learned about getting that newborn baby to third grade, reading on grade level, is that it involves interventions that are expensive:

a parent educator from birth through third grade, nurses for newborns. And those resources have to be provided by society because they are not available within the family or within the local school district.

And so, if we are going to take resources from Detroit and send them to Scottsdale, it does have real serious ramifications for patient outcomes.

CO-CHAIR NERENZ: Thank you.

Quick clarifying questions, any?

Yes, Nancy? And then, we will probably move on from there, but we can always come back to these. Yes?

MEMBER GARRETT: Yes, thanks, Steve.

So, I'm curious, among the SES variables you tested, did they all end up being significant in your final models?

MEMBER LIPSTEIN: Pardon?

MEMBER GARRETT: Did all of the variables, SES variables, you tested end up

being significant in your final models or did some of them fall away, because they are very highly --

MEMBER LIPSTEIN: Well, if you go back to the models that were in the -- can we go back a couple of slides? One more. Yes, right there.

So, you know, we looked at whether income alone was determinative. And it turns out that, if you just look at income -- and there's some problems with patient-reported income, as you know -- but what we started to do was we started to collect data on every single patient for all their difficult life circumstances, income, education level, whether they were above or below the federal poverty. We looked at obesity. We looked at substance abuse. We looked at disability, physical, emotional, or behavioral disability. We looked at smoking status.

And what was really kind of eye-opening for us was the housing unit vacancy

rate. Okay? That was the one that -- and you have to go to the Census tract level because, even within zip codes, the neighborhoods are very, very different.

And so, there's a video that is on the internet about St. Louis and the communities north of Delmar. But what you will find is that you needed to go to the Census tract level. The zip code or geomapping of the zip code level didn't expose this.

CO-CHAIR NERENZ: Then, to Gene. And again, we're borrowing a little time here from the next time block that is much more open discussion, just because I think we are getting some important material out here. So, if there are other questions more substantive, more thoughtful, about where this is all going, we are just about to enter that time block anyway.

So, Gene?

MEMBER NUCCIO: Hi. There's the

title slide with my colleagues listed down there. I apologize for the size of the font.

The work that I am going to be reporting on was partially funded by Kate and her group with the Home Health folks and with MedPAC folks.

Next slide.

Just a quick history. Nursing Home Compare happened in 2002, Home Health Compare in 2003, and Hospital Compare is a sort of Johnny-come-lately in 2005.

The risk adjustment in home health is based on a 12-month rolling average of the observation period. Risk adjustment, as I mentioned, is a two-part process. For us, there is a prediction model, and then, there is the application of the model's result.

For Home Health Compare, we take the observed value and we add the national predicted value and subtract the agency's predicted value from that to adjust that value down.

I am not going to speak to the issue of whether that is right or wrong, or how it can be improved. As I mentioned to Sean, if you want to see the results of that, you have to Academy Health.

The prediction model that we are going to be talking about here, basically, is based on OASIS data, the OASIS assessment instrument. I took this model from a million quality episodes of care. And it was validated against another million episodes of care. So, it is reasonably well-based.

There's about 12,000 home health agencies in the country. And obviously, this is a post-hospital thing. So, if you are wondering what happens to your patient in that 30-day window, they come to us. And then, hopefully, we can help out. There's about 5 million episodes of care in the last 12-months period.

The results that I'm going to be describing here in the next couple of slides

are a claims-based, OASIS-adjusted, acute-care hospitalization measure within 30 days of the end of care. Now that's different from -- the end of our care. Okay, the end of the home health agency care.

And I want to note that this is different than the claims-based measure that is currently being reported on the Home Health Compare, which measures a 30-day window from the start of care.

Now we obviously overlap because, if a patient comes into care and, then, leaves care within the first 30 days, they would be, obviously, in the hospital within 30 days of the end of the care.

Next slide, please. Actually, could you flip to the next slide? Okay.

The short model for this -- and this might seem long, but the larger models have more than 100 variables, and there is a reason for that. But the short model starts; the most important variable has to do with

whether or not the patient has had multiple hospitalizations, multiple prior hospitalizations, obviously, a very strong odds ratio.

Notice that the relationship for the variable care for joint replacement -- basically, if they had a hip or knee -- is negative. That is, we're very likely not to have a rehospitalization if the patient comes in with those sorts of things.

So, now if you could flip back?

And all those are P to the three zeroes and a 1. So, wildly significant.

The full models that we create for risk-adjusting the outcome, this particular outcome or the outcome similar to this for Home Health Compare for the home health, we purposely use many, many risk factors. And the reason for that is because people call up and say, "Hey, what should I put on this particular -- how should I answer this question, so that I get the most points?"

And I just tell them, I said, "You should put on there what the patient is presenting as." Okay. So, if the patient is a five on this scale, then put a five. And if they're a four, it's a four. "Don't worry about it because you can't beat the model." When you have 100-plus variables in the model, it can't be beaten.

I give you some of the statistics there that you might be interested in. So, for the full model, the C statistic is .7, which is a pretty strong model. And if you look down at the mini-model, which was the one that you just saw, you will see that the C statistic is .69.

So, when we drop from 113 variables down to 22, we capture virtually everything. So, it is just a way of preventing cheating, if you will.

If you look at what we did following this, we said, starting with this model -- and if you would focus on Table 2 --

if you, then, add in the characteristics of race, how does that change the C statistic in the model? We improve it 1/1000th. If you look at dual-eligible, what happened? Well, we improved it by 2/1000th.

If you look at length of stay -- and now, let me just point out we had a question about did you consider length of stay as a patient variable. This is a length of stay based on the agency. So, all these variables, race, dual-eligibility, and mean length of stay, are agency or organizational variables.

What we find is that agencies in Region 6, okay, which is Louisiana, Texas, Arkansas, all the way over to New Mexico and Oklahoma, those agencies in those states tend to have double the average length of stay for other agencies in the country.

So, it is not necessarily a patient perspective, but it is an organizational perspective. And so, what I am

arguing here is that, if we are going to be looking at SES kinds or sociodemographic variables, if we consider them as an organizational characteristic or a context-of-care characteristic, that's one way of dealing with this.

Another issue is that, for home health agencies -- we have 12,000 of them, and someone was asking about, well, if we stratify the results, can't you report the information? The problem is that we have about 25 percent of the home health agencies nationally that don't have scores reported for one or more outcome measures because they are too small.

And so, if you are reporting at a provider level, and you are, then, basing your analysis in terms or your incentive, pay for performance or whatever, on that, you're automatically disqualifying a quarter of the agencies because they are too small to have reliable data. As Sean said, there was too much variation.

Next slide.

Before you stratify.

Next one.

So, just some final thoughts. Adding the provider level, even demographics don't seem to add a lot of predicted power between the models. The provider levels, when you look at length of stay, actually, does the best, and that one was one I showed.

We did another analysis that I did not show that talked about dual-eligibles. And what we found was that the relationship between performance and percentage of dual-eligibles was not linear, but curvilinear. That is, agencies that had very low rates of Medicaid patients and agencies that had very high rates actually did better than agencies that had sort of the middle level of patients. So, we have to be careful about imposing linearity on all of our variables.

Parsimonious was better than others. And then, the issue of how you are

going to apply it, whether you are going to use regional or state predicted values rather than national, and whether or not you should use a difference model or a multiplicative model or some sort of ratio model or some sort of index model, all is to Steve's point. You know, how can we drive down the variability, so that when we find an extreme value, either extremely high or extremely low, we are not getting false-positives or false-negatives in those extreme positions.

I think I made it.

CO-CHAIR NERENZ: Thank you.

We had a couple of clarifying questions. Are there others, again, directly and pure clarifying?

Nancy?

MEMBER GARRETT: So, are you making a recommendation, Gene, in terms of whether the right approach to doing risk adjustment is at the provider level or at the patient level?

MEMBER NUCCIO: I'll defer that until tomorrow. Okay?

(Laughter.)

MEMBER GARRETT: I'm going to be in suspense. Okay.

(Laughter.)

MEMBER NUCCIO: I think it is something we should consider. If you are going to be representing the characteristic -- okay, are we representing the characteristic as a patient characteristic, like we're doing with their medical condition, or are we going to be looking at or should we think about representing demographics as an organizational or a care context variable?

And in that sense, then, it would be more appropriate to look at the context of care. The context, whether it is rurality -- and, clearly, Colorado has lots of rural outcome kinds of things -- or if we are going to be looking at it at the patient level.

CO-CHAIR NERENZ: Last but not

least, Alyna.

MEMBER CHIEN: Okay. I heard five minutes, and I said there's no way I'm going to put any data on any slides. So, if you feel like you need a visual, the paper I am going to talk about is in the set of papers that were put on the SharePoint. It's the Geocoded SES Factors Change of P4P Program in a Primary Care Setting.

But what I thought I would talk about is this sort of total question mark about what we think we are capturing when we are using geocoded information. And I have worked on the project with Larry in the P4P primary care setting, and then, also, I have a grant from NICHD looking at the same information geocoded almost exactly the same way and the same set of variables, but I'm saying that it represents the neighborhood health risk of a child in a spending model.

So, if you want the variables that I use -- well, first of all, let me say why we

were attracted to Census information. It goes back to something that Tia said, which is, aside from being stable over time, unless your address is changing is a lot, Census data is very readily available. It's free. It's reliable. It's very well-collected. It's very well-studied, also, in terms of how it might map onto social determinants.

So, in the United States there haws been a lot of work by a lot of people, but we relied on Nancy Krieger's work that assembled about 10 different SES variables from the Census. So, their median household income, which we think represents wealth. It is percent education level within whatever area you're talking about. And we could be talking about something as small as a block, a block group, which is a little bit bigger, maybe 200-300 people, or a Census tract, which can go up to 500 people, as opposed to a zip code, which is very heterogeneous and perhaps explains why it doesn't pick up some of the

things when we are looking at putting zip-code-level poverty levels into models.

So, it assembles all of those 10 factors. You average them out. So that, if you're average, it's a zero. And then, if you are at one of the two ends of the tails, you get plus-2 or minus-2 in terms of standard deviations, and you add it up.

So, the poorest person, the poorest Census tract ends up as a negative, like 24 or 26, and the most wealthy Census tract ends up being a positive 26. And it's nice because there is a pretty long gradient.

So, in the study that Larry and I did, we looked at IHAs, pay-for-performance program in California. So, we had pretty much 200 physician organizations, which was 10,000 practices, delivery care sites across the entire State.

And we thought that this Census tract information, if we mapped it to practices, it could represent practice-level

resources or the patients that are going there. We weren't really sure. But we thought, either way, it might matter in terms of your ability to deliver care that is high quality.

So, what we found -- and it's in the paper -- is that the folks who had more of their practices in low SE Census tracts had almost a 1.5 -- the ones who were wealthier had 1.5 greater odds of getting the P4P bonus, taking percent Medicaid into account.

So, it is an application of quality measurement in a setting that we are kind of being guided towards not talking about, but it is just another representation in a more primary care setting of how socioeconomic information might matter and widen disparities.

So, then, my next slide. So, I'm on slide -- I only have like basically two slides.

So, in the spending models for

pediatrics, it is the same geocoded information. We are playing around with Census tract versus block group. But, again, it is unclear what information we think we are picking up on.

There's lots of Census tract area information that suggests that, if you live in a lower SE tract, you are more exposed to violence. You have greater risk for infectious diseases like bacterial vaginosis, strep pneumo. There's lots of reasons to articulate that there might be, if you are living in that tract, you might actually have more health risks.

So, in this case we are attributing almost individual-level exposures based on the same Census tract information. None of these models have any multi-level components to them. They also don't take into account how the size of the area that you are using might impact things differently.

For example, block groups really

matter for exposure to violence and perhaps injury, but very wide swatches and sprawl indices suggest your risk for obesity. That's all being mooshed in here.

So, those are the two things I am working on. That is how we are thinking about what we think geocoded information might actually bring. It is very easy to work with. It is easy to get. It is difficult to game because it is hard to change somebody's address, unlike ICD-9 codes where we know that providers just will upcode to make their patients look more complex.

So, if you want to see the Census tract information that is in what we are using, it is the next slide. The next slide. Oh, it's that one.

And then, I thought I had to weigh-in -- so, can you go back one slide? -- had to weigh-in on what I thought whether to adjust or not to adjust. I just made it binary like this, even though I don't think of

this as black and white.

So, for spending, definitely, I think you need to take these factors into account or you will mismatch resources with the complexity of the patient.

For payment that is based on quality, I think it really just depends on the program design.

And risk adjustment is not our only tool at getting at disparities problems. The pay-for-performance programs are currently much more achievement-oriented, but we can design them to be improvement-oriented, so that maybe some of these factors don't get so widened-out the way they are. So, we could do a piece-rate pay for performance instead of the achievement point.

And I just don't want people to forget that we can advocate for different program designs.

That's it.

CO-CHAIR NERENZ: Thank you.

We said we are going to try to make a seamless transition into this upcoming block of time, but let me talk a little bit about how we might do that.

We still should take some time for questions to the speakers collectively, those who just went. But I wanted to just at least look around the corner to what is coming to help frame questions.

If we think of the charge in front of us, there are really two big meta-questions. The one is whether or not to take into account SES or sociodemographic variables, which effectively is a yes/no question in principle. And then, there is the how, when, whether, which, and all the details.

At least as the agenda is printed, the next hour or so in front of us is to talk just free-flowing as a group about the first of those two, the "whether".

So now, as you think of questions

to the people who have just spoken, if your question can transition us into this larger "whether" discussion, that would be good. You don't have to do that. But I would say, if you want to ask someone about a very picky methodological question that's more the "how," perhaps that could be held for later.

But, again, just a general suggestion, because now, from the perspective of the front of the table, this is when we enter the herding-cats part of the discussion.

(Laughter.)

It will be very hard to keep this all linear and flowing. So, first, a token attempt.

So, Nancy, you had your tag up.

MEMBER GARRETT: I just had a quick followup. I'm just a little confused as to whether you are using Census data to assign characteristics of where a provider is located to the care that is delivered there or patient care choices where they live, or both.

MEMBER CHIEN: So, both projects we are using Census tract. So, that is the area. But in the P4P program evaluation, we were ascribing that to the practice. In the one for the spending models for the children, we are ascribing that to the individual's neighborhood, for a lack of a better term.

MEMBER GROVER: I also have kind of a methodological question, but it is also, I think, a bit of a philosophical question. And that is sparked originally by Steve's comment about looking at the variation in Medicaid eligibility and how that differs. How do you sort of normalize for that.

But, even, then, with the Census information, if you are looking at household income, and you realize that poverty is very different if you are living in New York City or San Francisco versus living in St. Louis, how much adjustment do we do at a local area level for how much your home is worth or how much money you make?

MEMBER CHIEN: That's directed at me? Okay.

(Laughter.)

Because I am sure others can answer this question.

For the way we constructed it, it is a relative measure. So, it goes after that idea that it is that social ladder and your distance between the highest person and the lowest person. That is what matters here, not what the actual value is.

CO-CHAIR NERENZ: Okay. I should also just make sure we don't forget the folks on the phone. Questions for any of the folks who have presented in this last group?

MEMBER CALLAHAN: This is Mary Beth. I don't have anything. Thank you.

MEMBER CHIN: Thanks for the great presentations.

CO-CHAIR NERENZ: That's good. It's always tough, though. We just forget that there are folks up in the ceiling here.

We can't see them.

(Laughter.)

Okay, Susannah?

MEMBER BERNHEIM: I have a question for Atul and anyone else who wants to weigh-in on it.

I have tried to pull the paper up, but I think what you showed is that, when you risk-adjust for patient SES, there is a small difference, and when you risk-adjust for this sort of hospital SES, there is a bigger difference. And I just wanted you to talk a little bit more about your interpretation because I think there's lots of ways you could interpret that, and I wanted to hear a little bit more about what you thought.

MEMBER GROVER: I think one interpretation could be that -- and this is looking at adjustment with both individual- and hospital-level characteristics -- one way to interpret it is I think there is something going on in that community that affects

patients more broadly, and particularly as it relates to the ambulatory-sensitive hospitalizations. Is there just low quality or lots of barriers to good quality access to care in the community, which at least for the case of readmissions, then, would make sense. You are going to be admitted in the first place and readmitted after the fact.

And I think that my own philosophical bent is to suggest that, I think similar to Steve, that you look around that neighborhood and you realize there are huge differences in the populations being served locally.

Another interpretation may be that poor people are just admitted to bad hospitals, right? And I think that is sort of the other extreme of this, that people who are disadvantaged are going to low-quality hospitals. I think you have to get at that through a mix of measures and looking at processes and saying, you know, what's really

going to affect the patient and the outcome that is related to the hospital care versus what is going on outside the hospital and how you adjust for that.

MEMBER BERNHEIM: Can I ask a follow-on?

So, again on this concept of -- I mean, I am really just trying to think about this --

MEMBER GROVER: Uh-hum.

MEMBER BERNHEIM: -- because I think it is relevant to sort of how we would make recommendations about what variable you would or wouldn't use.

If these were targeted to ACOs instead of hospitals, so that there was a sense that there was more of a community-level responsibility for the readmissions or the follow-on care for these patients, would that change your thinking? I mean, is this more about sort of the hospital is not responsible for the fact that there is lots of ambulatory-

sensitive conditions because it is just not their purview? I mean, how would that change how you think about that hospital-level variable?

MEMBER GROVER: Well, I think if you are talking about an ACO that says, "Well, we're going to be responsible for all this stuff," yes, I probably would consider it differently, although I would still want to take into account the fact that there are differences in populations. And you could probably see this among Medicare Advantage plans as well. If you've got a Medicare Advantage plan that's got 60-percent duals, you might want to adjust the metrics for bonuses for that MA plan compared to one in which you've got 5- or 10-percent duals.

So, I think it would make some difference, but it is more a sense of there's lots of levers here. They're not all in the hospital. And how can you better figure out what is within the control and not within the

control of the provider whose outcomes you are trying to measure?

MEMBER PONCE: This is for Steve. I was convinced with your arguments about whether or not -- yes, you should -- but your outcome was a high bar. You know, it was your outcome of it is not just about a birth outcome, but about getting them to third grade, the immunization.

And so, I think the question, then, this high-level question for this group is it is not just adding these as risk adjusters, but thinking about what the outcomes should be. I think that was your message, and that is something you should think about.

So, posing it another way, if you had just looked at conventional birth outcomes, do you think you would have seen that compression?

MEMBER LIPSTEIN: So, you know, I think by giving the example of the high bar of

getting to third grade, it was my notion of just looking at what didn't happen to the patient after 30 days. Is it nearly as important as what did happen to the patient after 30 days? Just keeping them out of the hospital may not be a good indicator of what really has happened to them, especially given where they live. And that is the point I was trying to make.

But, if you just assess outcome at the day of discharge on a healthy baby, it doesn't necessarily mean that you are going to have a good outcome by third grade or long-term. That is the point I guess I was trying to make.

I was going to respond to Susannah's query because she was asking about kind of a methodological question about whether or not individual dual eligibility or, you know, the socioeconomic status of the individual versus the socioeconomic status of the hospital.

And what concerns me about that is that we are using, especially in the case of readmissions or Value-Based Purchasing, we are comparing people. We are comparing one person's socioeconomic status to another. And it is very variable across the country.

So, for example, if the Medicaid program pays you at 100 percent of cost in California and pays you at 40 percent of cost in Arkansas, to suggest that those two hospitals have the same resources available to care for patients just gets you to the wrong answer.

And so, that is why this whole area of Medicaid eligibility or Medicaid status as an indicator of the socioeconomic status of either the individual or the hospital just worries me a lot. It gets you to the wrong place. It doesn't get you to the right place.

And so, at least within state -- and this was Gene's idea earlier -- if you

come up with benchmarks within state, at least you are dealing with a standardized Medicaid program. And that at least makes the result a little bit closer to consistent across those providers.

CO-CHAIR NERENZ: Tia?

MEMBER SAWHNEY: Just a comment, the difficulty of what we are doing. We have talked a lot about disadvantaged populations and the very incredibly-dedicated people that are working to improve that. But we should keep in mind that there are also sharks out there and very bad providers who couldn't practice in any other hood. And their story always would be that they're not really bad; it is just their population.

A near and dear case in Chicago, a hospital that kept coming in again and again, a self-proclaimed safety-net hospital, and asked for and got all kinds of preferential treatment from the Medicaid program, et cetera, was shut down this year because they

were taking homeless people and doing unnecessary tracheotomies at $150,000 a pop, and were caught on wiretap discussing this.

So, we can't eliminate all deviation. We do want to be able to discern who the bad apples are.

CO-CHAIR NERENZ: Just let me speak in support of that point. I think that could be overlooked in some of our discussion, that there's this intellectual challenge that we can't avoid, that we don't want to label providers as bad if they are not bad, but we don't want to label providers as good if they really are bad. And so, somehow we have to try to end up at that point because the really bads do exist. So, an excellent point.

I think, Norbert, you are still on?

MEMBER GOLDFIELD: Just I definitely agree with that statement, and I have said that in my first slide, what you just said, David. I definitely agree.

Even though it is not part of the conversation, I just want to highlight, partly in comment to yours, is that some of the conversations we have with states is to try to give ACOs, for example, or whatever the month is, an leg up for the first year. So, the adjustment can be dynamic. It doesn't have to be static. So, that is not really part of the panel, but certainly, as we think obsessively about implementation, you know, that is something just to be cognizant of.

And certainly, we want to encourage providers to take on the sickest and the poorest people, but it doesn't necessarily have to be over 10 years. You know, it could be a period of time. With data, there should be an expectation of some improvement or maybe the risk adjustment has selected it out.

CO-CHAIR NERENZ: Thank you for giving me the opportunity.

What I am going to try to do is follow the sequence in which people put their

tags up, but if you have something that must immediately follow on something someone just said, that is okay and just indicate, like Steve just did, or put your hand up. I will try to recognize and follow that, because sometimes you don't want to lose track of a point.

So, yes, please.

MEMBER LIPSTEIN: Susannah and Norbert both made a point about ACOs. BJC is an ACO, and we have about almost 40,000 attributed lives.

But one of the important things to remember about ACOs is that it is a provider enrollment model, not a patient enrollment model. And one of the things we could study is how many of those providers or those primary care physicians -- it is a primary care attribution model -- how many of those primary care doctors are really in Census tracts of what I described?

So, what we have found, at least

in the ACOs that we have studied, is that a lot of these are enrollment models in more affluent than less affluent communities. And so, you are not going to get a normalized distribution of patients by looking in ACOs.

MEMBER SUGG: So, just a couple of comments. So, when I was first thinking about this last night, you know, the pathway, yes-or-no pathway, I was actually standing before the pathway thinking, is there such a thing as a quality measure? I mean, does that really exist? Can we actually come up with a measure that measures what we want?

And so, to get back to what Tia is saying, what I hear from the city and different grant-funding agencies is: we want to know what we are paying for. We want to know what we are getting for our money. And we want to make sure that we get rid of the bad apples.

But the problem I see with how the quality measures get rolled out is we get rid

of maybe some of the bad apples, but we get rid of a lot of good apples in the process. And how do we develop a quality measure that gets at what are we paying for and get rid of the people that truly are not giving quality service?

So, in my practice I have people who are schizophrenic, diabetic, and they use crack on a regular basis. I will work with those folks for five years. We may get them housed. We may get them on meds. We may get their A1c from 14 to 11, which for us is a hallelujah.

I will suffer financially under the current system. That said, I still want to get rid of the bad apples. So, I really struggle with how do we come up with something. We are never going to be pure because we can't possibly get all the socioeconomic factors that could be put in the equation. That is just not financially possible. But how do we come close enough

that we get what we want and not ding the people that maybe just need some extra training or some mentoring or better resources, and they are not truly the bad apples?

And that is where I really struggle with even the whole concept of quality. You know, it is going to be sort of a shadow of what we are really trying to measure, but how do we come close with that?

The other thing, because I think language is important, I really get concerned when I see articles that say socioeconomic status is not associated with pay for performance in this or that way, when the reality is what it is actually saying is perhaps Medicaid as a proxy is not. And we know that and we know what questions to ask. But I can tell you my state legislatures will not know those questions. They will just see the headline and maybe the title of the report and make an assumption. And I struggle with

that, of how we also communicate when we talk about these measures.

MEMBER CASALINO: The bad apples question keeps coming up. I think there are three reasons why a hospital or a medical group that takes care of a socially-disadvantaged population, let's say, might have bad scores, bad quality scores, right?

One is the patients are just really hard to take care of, right?

Two is they may not have transportation to get their mammograms, to take a process measure, and so on. We have had lots of examples today.

So, the patient bag, that's No. 1. The second is they may be bad apples, right? They may just not care. They may not be good at improving, whatever.

And the third is they might be good apples in terms of their intentions and even their skills, but they might lack resources, as Atul keeps pointing out, right?

So, either one or all three of those, or two out of the three of those, could be in effect at anytime, right? So, there is going to be no perfect solution to this.

But I think that, if you have a method that pays both absolute score and improvement over time, which some programs do have now, and that pays based on both your kind of performance in comparison to everybody else, leaving open for question at the moment who "everybody else" is, and for organizations in your category, okay, then I think the bad apples are always going to get paid really poorly under that formula and the good apples that just are having a tough time because they don't have the resources and it is a hard patient population, there will still be some shift of resources to the really highest-performing hospitals who may be in better situations. But at least there won't be this total shift of money from East St. Louis to Scottsdale, and the good apples will have some

chance.

So, I think some kind of a compromise is probably the best we are going to be able to do. Without that compromise, we put the good apples and the bad apples in the same barrel and leave them forever, basically.

MEMBER BHAREL: Larry said everything I was about to say much better than I would say it.

You know, I think when we went back, I have had up most of the time Kevin's initial picture, and we talk about to adjust or not. And when we say, "Yes, adjust," and then, it says, "Well, are we, then, developing a two-tier system?", I mean the truth of the matter is in our current situation we already are in a two-tier system or three or four. And depending on where you look, it is either within a city with two different hospitals, one for certain types of patients, another for another, or it is within a hospital system or it is within an ACO where patients are already

getting differentiated care.

So, you know, I think Tia's point about how do you make sure these abuses don't happen, we have to do that even in the current system. I think that is always an issue.

And one of the ways to do that is to look at things like pay for improvement, which has already been said. But, for example, if you look at my own program where we've tried extensive integration across the community spectrum, five years ago our pap smear rate for women, for multiple reasons, was 19 percent, 1, 9.

But, through multiple different incentives, including HRSA publicly making available reports, our pap smear rate now is 50 percent. If you were to just do pay for performance, we still wouldn't make it. But we are hugely celebrating that as an achievement. So, I think there are ways to make this work if we think creatively.

MEMBER GARRETT: I just wanted to

follow up on a point that Nancy made. So, we have heard a number of presentations today, some of which showed a small effect of SES adjustment, some of which showed a large effect. And I am not sure that that is really relevant for a decision about whether to risk-adjust for these factors or not.

To me, it is interesting to see that there are differences in how we are applying these methods, but I think we need to decide whether in principle we should control for these social determinants. And then, if so, it is the "how" and what's the best way to do it. And whether or not they end up affecting the outcomes and the results is not really relevant, I think, to that question about the principle.

MEMBER CASALINO: Before you control for other things, like clinical factors, if SES doesn't affect outcomes, then, in principle, why would it be relevant? I mean, if I was really convinced that it made

no sense, it added nothing, to put in SES variables, then why would it make sense for me to say, "But I think SES variables are important ethically; therefore, I want them in."? I don't see that.

MEMBER GARRETT: Yes. So, I guess what I am saying is, if we believe that there is enough evidence that social determinants do affect outcomes, then we would decide, yes, we should risk-adjust. And then, the "how" gets into, okay, well, what's the best way to measure that. In some situations we are going to see a bigger effect than in others because of the measurement and because of the processes underlying it.

So, if we believe there is enough evidence that there is an association, then I think we could make that first decision. Does that make sense?

CO-CHAIR NERENZ: I've got Sean, Susannah, Pam, and Kevin so far.

MEMBER O'BRIEN: I wanted to talk

about the inclusion of hospital-level percent summary measure of an SES factor, percent Medicaid or percent other types of examples of measures that have been adjusted for. When you include those types of variables in the model, I think it is important to talk about, well, what's the motivation for adjusting for them. And it really changes the interpretation of what question you are asking and answering.

So, all the models express performance of a unit of like a hospital relative to some group. And so, your question is, relative to whom? And if you don't adjust for hospital-level covariants or hospital-level summaries, by default, it is relative to everyone. And you can leave aside the question of who is everyone. Is it nationally? Is it some subset of interest, as Dr. Casalino indicated?

But when you adjust for hospital-level summary statistics like that, you are

implicitly saying the relevant comparison is relative to other hospitals that share that same summary measure. And that may or may not be the relevant comparison for some particular purpose of interest.

And I just think, when this is all wrapped in the questions about complicated modeling questions, sometimes it is hard to think through these issues. If there are no real case-mix confounding issues going on, there still may be reasons for wanting to look at homogeneous competition pools, as Dr. Casalino's paper called them, when comparing groups of hospitals, you know, for incentivizing performance and avoiding unintended consequences.

So, if we are going to make any kind of recommendations about the inclusion or not inclusion of adjustment for case-mix variables, I just think we have to really make those two distinct issues. One is adjusting for patient-level and one is the hospital-

level summaries of case mix. When you adjust for the latter, you are changing the question that you are asking and answering with the model.

MEMBER BERNHEIM: Sorry, I was reacting to a different speaker. So, two quick things.

One is just this idea of sort of there is a lot of concern about the unintended consequences. And I just want to raise again sort of the intended consequences and reflect a tiny bit 9on my own clinical practice and experience at Yale, where it is fairly large underserved population, and the hospital has done some astounding things recently to improve readmissions among the low SES patients, including setting up respite beds, so that they are never discharging homeless people to nowhere.

And that took an enormous investment of resources at a hospital that is losing a lot of money on the Readmission

Reduction Program right now. And I don't know if that would have happened, but it is certainly providing better care to these patients. It is certainly ensuring better outcomes.

And so, again, just in the mix, and it is not straightforward on either side, but I do feel like I am seeing -- and I am actually asking myself different questions in my clinical practice. I am finding that I am thinking about this SES issue so much that sometimes things that I would have just said, like "I can't do this today," I am finding myself in a setting that I have always worked in and thought I was the most committed provider, like going the extra mile to ensure different care for patients.

So, I just want us to keep in mind I don't know what would happen if you put this in there, but you are going to set a different standard, and there are both intended and unintended consequences with the current

measure. So, I worry about both sides of that, and I don't think it is a simple answer.

And there was something else, but I have forgotten it. So, I will wait. If it is important, it will come back.

MEMBER OWENS: So, I just wanted to build on something Nancy said, and then, Larry, you followed up on. And I think the question is not, "Do you make a blanket statement, yes, you adjust or account for socioeconomic status or, no, you don't?" Rather, for me, my answer would be you need to consider it.

My concern is, if you come up with a blanket statement that says, "Yes, you should risk-adjust," that will become policy. Every measure will have SES, and we have gone down a trajectory that really was not intended.

But I do think it is an important consideration, and perhaps if you go down the path of NQF endorsement kinds of things, you

would need to specify why it wasn't or what you did to consider it. Maybe that is, for me, more of the question than the blanket statement of, "Yes, do this" or "No, don't do that." I don't see it as a possibility.

CO-CHAIR NERENZ: Okay. Just a quick response. And again, let me just express my own previously-unstated assumptions, but I think this is a really important point.

When I have thought about this first big meta-question, which I think we phrased as, "Should we do this?", I actually am doing a quick mental translation that essentially makes it "Can CMS do it?" or "Can measure developers do it?", as opposed to the current position that says, "No, you cannot."

I wouldn't take it quite as far as I heard you express it, to say you must or this must now be a characteristic of every NQF-endorsed measure. I had never sort of understood a yes answer to our first question

to imply that. But, again, that is just purely in my own head, my own assumption.

But I thank you for bringing that up, and I would be interested if --

MS. PACE: I mean, I think the question really is, "Should it be considered?" And I think, if we work through -- it is kind of, have we had enough discussion? And we have a few more questions about whether it should be considered. Then, the "how" gets into, you know, if it is not related to a particular outcome, as someone already brought up, then there is no point in adding it to a model.

But, you know, some of where we want to get to with the recommendations, as we talked about earlier, and Kate just mentioned, is principles. You know, when people submit a measure to us, would there be particular analyses we would want to see that justified including or not including?

So, I think you're on the right

track. It is definitely not going to be a black-and-white and applies to every situation.

CO-CHAIR NERENZ: So, maybe just to paraphrase, it is currently the case -- again, I am trying to imagine myself in the position of a measure developer -- even if there were empirical data or analysis with which to include an SES factor, one would probably not bring a measure forward with that, just because the current NQF policy says no.

Where I think we may conceivably shift to is to say a measure developer can include that. But, if in a particular domain it didn't make sense, no empirical relationship, no good available data, you wouldn't have to. Is that a fair restatement?

MS. PACE: Yes. I think that is what we really want to try to work through these questions of whether we should consider it. Then, how we look at selecting risk

factors. And then, how you think about putting them in a model.

CO-CHAIR FISCELLA: I wanted to respond to Larry's query of why would you consider it if it doesn't seem to make a difference or maybe not much. And a couple of thoughts come to mind.

One is that it may not make a difference nationally, but it may make a difference in a particular state, in a particular locality. It may not make a difference right now, but our healthcare system is changing enormously rapidly, and resources are going to be distributed in very different ways and deployed in very different ways. And I think it is very hard to predict what is going to happen. So, it might make a difference today; it might make a difference tomorrow.

We have conceptual reasons why SES and resource deployment make a difference. And at worst, if one were to do that, one may

actually develop the data on which to make better decisions in the future. You would begin to see, well, what conditions, at least currently, where it is making a difference versus those where it is not.

And right now, we don't really have good data on that. We have selected examples that people have published. But, by and large, we really don't know the answer to that question: for which measures, for which conditions and which regions does it matter more or less?

CO-CHAIR NERENZ: So far, I just run right down the table. I have got Atul, then Steve, then Larry.

MEMBER GROVER: Getting back to this question of, is it a can you use this as part of risk adjustment, I think part of it depends upon whether you are talking about a process measure or an outcome measure. On the outcome measures, I would say, yes, you would probably want to default on the side of

measuring it.

But we were having this conversation the other day with colleagues around, well, should that affect at all, SES affect the ability of a nurse to administer antibiotics in a timely way? Well, probably not, right? Like I could go down that path of saying, well, you know, she's really rushed and she's got five patients, but probably not, right? So, you would probably think on a lot of the process stuff maybe you wouldn't include this.

So, I think, yes, getting to "Can you," but then, I think it becomes very difficult, and I don't know how this works at NQF particularly well. But if you have got measure developers, and you may want to say, "Well, don't use it within the risk adjustment at the individual level," or maybe you do, but how do we actually, then, decide when it comes to policy? Because, again, some of these things are very small shifts in, you know, 260

hospitals moving a decile. Well, maybe all of those, then, just went from no penalty or a small penalty back to gaining money.

And I think, somewhere in there, there needs to be room to give guidance that NQF can say to CMS, if they adopt a measure, you know, "You should adjust this for socioeconomic status." And I just don't know where the flexibility is or the mandate is in that process right now.

MEMBER LIPSTEIN: I guess I was going to caution us against our experiential lens. And the reason I was going to caution us is because many of us work in or around academic medical centers because we are academics or researchers.

Academic medical centers and teaching hospitals, in particular, have resources that not all hospitals have. And their ability to care for a sociodemographic segment at scale, not one patient in one examining room and one discipline at a time,

but at scale, is different than it is in a rural hospital or it is in a Community Health Center that is independent of an academic medical center.

So, I just caution those who do see patients and do research, if they do it in a teaching hospital environment, those teaching hospitals -- Boston Children's Hospital is an example -- have resources that other communities just don't have. So, that would be one point I would make.

The second is that we have to acknowledge that NQF-endorsed Quality Indicators are determining millions and millions of dollars of federal government funds flow. And we are going to see this change dramatically in the next three years, as Medicaid dollars flow disproportionately to 25 states compared to 25 other states.

And so, I don't think we can ignore that in our policy recommendation: that whatever we decide has implications for

federal government, federal taxpayer dollar funds flow. And if don't risk-adjust, the money may go one way, and if we do risk-adjust, the money may go a different way. That's irrefutable almost.

So, this decision/recommendation we are making has lots of consequences. And I have expressed my concern about this: if you don't risk-adjust, money will flow disproportionately to providers who don't disproportionately serve the kinds of Census tracts that I talked about earlier.

CO-CHAIR NERENZ: Larry first. Then, Tia.

MEMBER CASALINO: Again, just maybe at the risk of putting my foot in my mouth, just some practical context, I think, for the decision about whether to include risk adjustment in an incentive program of whatever kind.

And very specifically to the question of whether NQF, which I take it has

been saying, "We don't want to hear about SES as an adjuster for proposed quality measures," to, I mean, NQF could say, "You must show us in some detail or with some data that you considered whether SES should be a risk-adjuster or not, and why you decided yes or no." That is separate from the question of, if it should be, what you do with it, right, in terms of the payment form it takes?

So, if that is an important question, I think practical context would say you have to look at the kind of real-life world of the people who set up incentive programs. Now the real-life world of those people is, whether they work for a commercial health plan or whoever, their job is to set up an incentive program. And their job is to do it fairly fast, right? I mean, that's what their boss wants, whatever context they are in.

And I'm not attributing bad faith to people at all; quite the contrary. But any

things that complicate that are not welcome, right?

So, Alyna headed a study that we did in Chicago, and not very many years ago. Alyna, you can probably -- correct me if I get the details wrong, but the gist of it was she went out and interviewed a lot of people who were responsible for running, instituting pay-for-performance programs at major health plans. And of these people who were doing it, and there were a fairly large number of them, almost none -- this was about, what, seven years ago? -- almost none had even thought about the questions we are discussing today. They were actually surprised by them, right?

Now I don't think that would still be true today. But, once they heard about them, they still didn't really want to think about it because it was in their way.

So, this is why it is so important, I think what NQF does. This would change fast if NQF says, "No, you have to at

least show us how you thought about this, ideally, with some data, and why you are not proposing that as a risk-adjuster." That would kind of force people who, otherwise, you know, would rather not have to deal with this, because it is politically controversial and, as we can see today, it is a thorny methodological issue as well. So, there is a lot at stake here really in what NQF decides.

MEMBER CHIN: I can't see the queue, if I'm in the queue, or else maybe I'm in now.

CO-CHAIR NERENZ: Okay. Let me just say, yes, because we can't see your intent, Marshall and Mary Beth, you just have to jump in like you just did. And then, we will let you do that.

MEMBER CHIN: Okay. So, I have been thinking throughout about how can the issues that have come up be practically framed with an NQF mandate in a sense, but still be relevant.

If I had to summarize, I mean, I heard from Helen Burstin and Chris Cassel this morning that they were very open to us going beyond sort of the traditional expectations at NQF. They specifically said that the world has changed. And so, they seem to be very open to hearing new perspectives that may go beyond traditional recommendations.

Similarly, Kate's request in the group seemed to be, well, it would be great if we would provide a list of key principles to use, and she was pretty open with that also. They will use them or not use them as they see fit.

I think the other part of the context is that I think overall, a point made at last week's NQF MAP meeting, NQF has had a hard time addressing the disparities issue coherently.

In the past, maybe it is like two or three of us in this group that were on this Health Disparities/Cultural Concept Committee.

But that was, again, a very forced Committee where the goal was to try to come up with like 40 disparity measures, which was kind of a losing battle from the beginning.

You know, this Committee is like the next one that is close to the disparities issue, and it is a key issue, this risk adjustment. But there hasn't really been sort of an overall coherent look of, if we are going to think about equity as one of the pillars of quality, how are we going to come up with it in a coherent, sort of consistent, holistic way? And there has been plenty of holes.

So, I think this Committee really is a great opportunity. And I think we have been given sort of the green light and the open door here to really sort of, as people like Larry said, and Steve also, really sort of discuss the issues which are relevant, which are: it is going to be public reporting. It is going to be incentive

payment. It is going to be, as we are talking about, what would you might think about adjusting on an individual level for SES versus the ecological hospital/regional basis?

I think we should go for it. This is a great opportunity, and we shouldn't be constrained, I think.

CO-CHAIR NERENZ: Thank you.

MEMBER SAWHNEY: I am going to introduce the constraints. The fact is you can't risk-adjust anything without data. And if you are going to risk-adjust at the system level, you need to have the data at the system level. And, by and large, we don't have it today.

That is the reason why we have relied heavily over the decade on claims data. Because, even though we all hate claims data, in many respects it is standardized.

Patient address or insured is available either from the clinics or from the insurance companies. So, address is

something, and so geocoding has good possibilities. The constraint there, though, is that address under HIPAA is identity. And it is hard to pass addresses around either in address format or geocoded format because it is telling precisely where insureds are.

So, it is like who runs it through the geocoders to get it to an aggregation level where it is no longer identified. Or how do we change practice standards, which is a very difficult thing to contemplate, to gather data that we're not consistently gathering today?

So, we are not adjusting for SES. I mean, at the conceptual level, I think we all want to adjust at SES, but we need to be specific as to what SES variable and whether it is doable.

CO-CHAIR NERENZ: Thank you.

Just a very quick process observation, but we will continue here, and, actually, for some reason, it all now has

moved to this side of the table.

(Laughter.)

No, that is okay. Once we get done with that sequence, I am going to call a temporary pause. I'm looking at the clock and thinking of our overall agenda, and we will just try to take stock a little bit of where we are.

So, before anybody else puts a nametag up, let's just take care of those that are up. And then, I want to do a little take-stock pause. And then, we will certainly get right back into our discussion.

Actually, I have Atul. Let's see, have I not crossed-off Atul? Okay. I have Tia. It's up to me. Okay. I have got Norbert, Alyna, Pam, Rachel, in that order. Okay. I got Pam, right?

MEMBER GOLDFIELD: So, I think some of the issues that Tia obsesses about I obsess about also in terms of implementation. Because, for me, at the end of the day it is

implementation.

But I think there is a good solution and there are solutions. In fact, I was just curious and I emailed one of the states. And they have decided to collect homelessness data.

Don't forget, when DRGs were implemented in 1982, it was preceded by six years of pilot tests at states. And I would say that, especially with something as controversial, you know, I guess -- it is not controversial to me -- but I think what is interesting is not the "whether," but I think what is interesting to me is the "how" and the "what".

And so, I think different states, CMS could be encouraging different states to try to collect information in different ways. For sure, at least in this one very large state, they will be collecting homelessness data in the next few months. And so, we will have data.

And here is the bottom-line point: we will not only have data; six months after that we can adjust payment rates. We can adjust looking at outcomes within a year. No, it's not immediately, but it is pretty quick. So, I am totally onto the program, which is why I have tried to encourage states. Because I think CMS, for issues that are different, not so much in terms of your area, but since, for example, the readmission penalty is a zero-sum game for sure, the AHA is not about to start dispossessing suburban hospitals, and neither is the Federation, and try to give more money to safety-net institutions.

I want to just take exception to Larry's comment with respect to the private sector, what they are thinking about. Until very recently, most of them couldn't care less about Medicaid. You know, it was just a part of the book of business. Now they are very interested in the Medicaid book of business, you know, for obvious financial reasons. And

I get that.

But what is happening in several states is that state Medicaid programs are mandating that private insurers look at data in a particular way. So, stronger states as opposed to weaker states are mandating that the benefits managers, or whoever is calculating the rates, start to think about these type of issues, whether it be incarceration, homelessness, and so forth.

So, bottom line: I think there is a two-track road that we can take here. Obviously, you all know that I believe that it is not "whether," but "how". But that could be done in an experimental type of way, encouraging states to do, and you could also format some suggested definitions. You know, for example, homelessness, do you recommend the HUD definition? Obviously, we are not going to decide that by tomorrow.

But I have been steeped into that literature for quite some time, as I discuss

with different states. So, I think it is imminently possible. And we could have socioeconomic disparity risk adjustment certainly within a year and a half.

MEMBER CHIEN: Just listening to the discussion, it makes it sound like we are going to flip a switch and all the quality measures are going to be risk-adjusted. And then, we are going to flip a different switch, and we'll flip it down, and then, they're not going to be risk-adjusted.

So, I think the problem is that what is happening with socioeconomic status is that it feels invisible. And when it is invisible, you can't tell what people are doing with it.

So, I think it is the option or the requirement to see it both ways which tells us what might be happening at the patient level or the provider level or the state level. And so, it is actually do it, but do it both ways.

MEMBER OWENS: Yes, I was just going to reflect on measure development from wearing a measure developer hat and the submission to NQF, and whether it comes across in terms of -- well, there are two parts to your submission to NQF. You put together a well-specified indicator and you show it and very well could present it both ways on SES. That could be a bucket, you know, as part of the consideration with SES.

But there is another part out of the NQF, and that is, then it is implemented, right? And not all measures are implemented in the same thing. A lot of the AHRQ -- a lot of NQF measures are implemented by CMS, but, frankly, for AHRQ QI measures, they are implemented in all different kinds of ways. And I can't project all of those ways.

Some are used for quality improvement. Well, within a hospital, it would make no sense to risk-adjust on socioeconomic status. So, if you make that

part of NQF endorsement mandated, that might be a bit problematic. On the other hand, it may actually increase the validity and reliability of the measure. Without testing, you wouldn't know.

So, I guess I am trying to separate out the measure specification and going through endorsement as it is for reliable and valid data. And then, maybe building on what you said, Helen, which is, what if NQF moves towards "For what purpose?" and what you need to do for what purpose? And say a little bit more about that implementation because now you have got some analysis to support it.

Still, it is a consideration, but I guess I would like personally to stay away from sort of a mandate to do it, when you can't know all scenarios, and to separate those two aspects of this NQF document.

MEMBER WERNER: So, I just want to comment on something which has come up a

couple of times, which is about whether it is feasible to do this because of data availability concerns, and just say that I don't think that the lack of current available data on SES should hamstring us against recommending that it be there. Because, often, it is a strong recommendation from an organization like the NQF that really tips the scale and makes people start to collect this data.

You know, 20 years ago, we didn't have information about the kind of quality that is being delivered at hospitals that we have today. And the reason that we do is because CMS said that we needed it.

And so, I think that that data can become available. And it is important to keep in mind that these things change, and that it is not that hard to start collecting new data in many cases.

I do have sort of a followup. I don't know if it is a procedural question. It

may be too detailed for this conversation right now.

But, when I hear you talk about sort of the vision for recommending adjustment for socioeconomic status, there is a lot of different ways to measure that, what socioeconomic status is. And I am not sure that the science is currently available to know how to best do it. And I am just wondering if this is something where we just kick the can down the road and say we recommend adjusting for socioeconomic status, but we are not sure how to define that, or if that is something that is going to have further followup.

CO-CHAIR NERENZ: Okay. Let me try to respond. Again, I have no special standing here than that I happen to be in this quadrant of the room.

(Laughter.)

And again, the NQF folks can answer.

I will try to respond just to the very end of that, but, then, see if I can try to pull some threads together and take stock of where we are.

I don't think is charge is specified in terms of literally SES as sort of a single measurable concept. I think our charge is phrased as "sociodemographic variables," which I think actually makes the task somewhat easier because we are not asked to decide whether a single great measure of SES exists or whether it is available or even conceptually what it means.

Again, my own personal understanding has been that we are talking about domain of variables, meaning variables plural. Some may make sense for one measure. A different set may make sense for a different measure.

So, I don't sort of feel a problem or concern or barrier just on the label "SES". In fact, to me, SES is part of the larger

domain of sociodemographics.

MEMBER WERNER: Do you mean to separate SES from sociodemographics? I sort of meant sociodemographics in a larger way. And it sounds like what you're saying is that the charge is not to specify what sociodemographic variables are.

CO-CHAIR NERENZ: Again, just kick me under the table if I am wrong here. It seems like that clearly enters the discussion under the "how". And we have already had some things about, well, percent Medicaid is good or not good or Census tract versus zip code is good or not good. And this has entered the discussion already. And clearly, it seems to me, under the larger "how" umbrella, we are absolutely right into it. Is it a good variable or a bad variable? Is it a good variable for this kind of measure or that kind of measure? So, I don't think we're done with that issue.

MEMBER WERNER: My question is,

are we going to define it by the end?

MS. PACE: If we could, that would be great.

(Laughter.)

But, you know, I suspect that we are not going to get to that level of granularity to say this is the best way to capture all of the various things that we are talking about.

I know there is a lot of interest in talking, but when we move on to the next topic, we want to talk about selecting risk factors. What are the general principles? Are there some unique things? And, you know, as David said, you have already talked about some of the problems with some of these risk factors.

You know, people have their special interests. Whether that should translate into a recommendation -- you know, that's why we were trying to look at, do these variables correlate with other variables that

are maybe easier to measure? You know, someone also mentioned the problem, well, sometimes when you use proxy measures, you get the wrong answer.

So, we do want to try to work through those, and we will see. I can't say. You know, ideally, that would be great. I'm not sure that, as you've all mentioned, the difficulties with data will lead us there, but maybe, then, there's some short-term and longer-term recommendations that we need to think about for some of the more problematic areas.

CO-CHAIR FISCELLA: Just a quick comment. I mean, at least from my perspective, I would see this as a first shot at coming down, hopefully, with some very important principles that sets us in a general direction. I think the more specificity we can have, the better, but what is realistic within a two-day I think is really the question.

But I wouldn't see this, and I hope NQF wouldn't, as a one-time initiative. I think if we do go down this path of figuring out at least some of the time accounting for sociodemographic and the social disadvantaged, that this would be an ongoing effort, where -- I don't know -- three years, five years down the road, we look again at the data. We look at new data that has been collected and maybe figure out better ways to do it.

DR. BURSTIN: Just to pile on a bit, I especially like your term, Rachel, of not wanting to kick the can down the road, and we don't want to, just to make that very clear. We want to be able to come out of this with as much clarity as we can for both the measure development community and those who implement measures. If there is additional work to do, we will try to seek additional funding to keep this work going. It is really, obviously, critically important.

And I could start list out a list

of the other sort of major measurement science kind of issues that trip us up. I think there is a really important role here for us to get groups like this together to do this work.

And this, just from my perspective, has been an amazing discussion that I have not heard at a lot of other tables before.

So, we will try to do whatever we can in the context of this project. And I think the developers are also looking to us to give clarity. So, the more we could say those variables might work, those are problematic for the following reasons, that I think is a good starting point.

CO-CHAIR NERENZ: Okay. Let me tiptoe gently out on a limb and try to see if I can capture the sense of the room, and you will, then, tell me if I have or haven't, on this first big question, which I would sort of phrase as: in principle, should NQF consider or allow the inclusion of sociodemographic

variables in adjustment for performance measures?

My sense of the sense of the room, it is yes, but with several important qualifications. And I am just going to pick people at random to look at it.

In terms of a couple of Pam's comments, in the sense that I'm thinking of it, the "yes" does not have the meaning of "shall" or "must" as applied to any one particular measure. It has the sense of "may" or "can".

And in saying that, I am thinking of kind of a visual metaphor where my sense is, up until now, that particular door has been closed or at least that it has been closed. I think now what we are saying is that door should be open.

Now it doesn't imply that we have, then, blessed everything that might walk through that door. That's the "how" and that is the detail. But at least we are saying

something quite significant if that door should be open in principle.

We are also, in saying in principle yes, not saying that there is a standard SES variable that would be automatic included in relation to any measure. We are not saying that all performance measures would naturally or inevitably have this form of adjustment. It may make sense someplace; it may not make sense. But I think what I am hearing is that, where it does make sense, that we would like that to be done in principle.

Also, in terms of Tia's thing, there clearly are questions of feasibility and data availability that would be part of the "how" question. So, a "yes" here does not force movement into directions where it is not good because the data don't exist. We don't end up saying that such-and-such performance measure should be adjusted by a variable for which data do not exist. Okay.

So, I am seeing a few nods, but also, Larry, I know you had your hand up.

I am just trying to see if we can take at least a preliminary position that this first big question would be answered yes, with all of those caveats, and then, swing into what I think will be even the bigger discussion of the "how's," the "when's".

Okay, Larry?

MS. PACE: I am just wondering, because we have had a lot of discussions about the potential benefits of this, could we very actively ask about the potential negatives, so that we have those out on the table?

CO-CHAIR NERENZ: And I guess I should ask, as a process, are we being transcribed?

MS. PACE: Yes.

CO-CHAIR NERENZ: Okay. Because, clearly, I mean, I'm thinking again of the bad apples example. I mean, that has been out there.

MS. PACE: Okay.

CO-CHAIR NERENZ: And I think, again, a couple of the presentations pointed out situations in which an approach to adjustment was taken, and it didn't seem to matter. Okay. I mean, I guess that would be at least a type of negative where you might attempt to do it and find it made no difference, and, therefore, had wasted some time.

And actually, then, if you went deep down that path and committed to a whole lot of data collection that didn't currently exist, and in the end result it made no difference, that would be a negative. But I think that's on record already.

MS. PACE: Okay.

CO-CHAIR NERENZ: Burden of data collection, that would be another one. Fair enough.

Tia?

MEMBER SAWHNEY: We have talked

about it in terms of it benefitting providers who are working in challenging circumstances. But, to the extent it is a zero-sum game, it is going to make other providers look worse, and they are not going to take kindly to that and they are not going to take kindly if there is money attached to it.

I mean, I have worked in state government. I understand the political ramifications of taking money away from people.

CO-CHAIR NERENZ: Fair point.

Yes?

MEMBER CASALINO: Yes, it seems to me that your questions -- the discussion the last 45 minutes or so has made me think that the question you are trying to ask isn't actually a yes-or-no question. I think there are three possibilities for NQF, right?

One would be leave it as it is, right? We don't want to hear about SES in proposed quality measures.

The other would be, as you put it, open the door and say, "You can bring this up" or "We welcome it if you bring it up." And also, maybe go on to say, "And if you do want to consider this, yes, here are things that you ought to think about."

So, I think those were the two possibilities you were raising. But I think there is a third, which would be, "No, if you want to propose a measure, you have to show us that you considered whether SES should be brought into it or not, ideally with some data. And if you decide not, that's fine, but we would like to see the justification one way or another."

So, I would say there are three possibilities rather than two. And I suspect we may have maybe not that much division about a yes-or-no question. I don't know; we might have more across the spread of the three possibilities. I don't know.

CO-CHAIR NERENZ: Just could I add

a friendly amendment, that the last two of the three I would consider variations of "yes," at least in terms of the question posed to us. Okay. Thank you.

I'm kind of losing track. I think Gene was up next or Steve.

MEMBER CALLAHAN: Following on that idea, I'm concerned that, if we say we are now in a permissive mode, and that NQF will be looking to us to provide evidence in some form that we have considered the use of sociodemographics in our models, that Karen and others are going to have to come up with some criteria for judging how well we did that.

I mean, you know, if you look at reliability, there are certain techniques in determining reliability, and the same with validity. In the same way, that not only will we have to think about how we define and what is the domain of sociodemographics -- you know, what's in; what's out? -- but we will

also have to define sort of a scientific methodology of how well we tested to provide evidence that we should or should not include these variables in there.

So, I just caution us -- I mean, I'm not above looking at that issue, but we have complicated the road, you know, the pathway.

MEMBER CALLAHAN: This is Mary Beth.

CO-CHAIR NERENZ: Go ahead.

MEMBER CALLAHAN: May I speak?

CO-CHAIR NERENZ: Yes, go ahead.

MEMBER CALLAHAN: I'm in favor of including sociodemographic and socioeconomic factors, but in terms of just throwing out a con, I think that in medical settings, as guidelines come out, as I was involved in the KDOQI guidelines, I think sometimes as recommendations or suggestions come out, the medical community, even though they may just be suggestions and "can use," the medical

community feels like they are being pushed upon.

Now I think these are good things, but I'm just saying, you know, you asked for the possible negatives, and I think that could be seen as one.

CO-CHAIR NERENZ: Good point. Thank you.

MEMBER GOLDFIELD: Just a quick thing. I think another variation on the friendly amendments is that, again, states, as I have already highlighted, states are going to down the road with or without NQF. So, that is a given. Okay? It is already beginning to happen.

The question is, I think, whether or not NQF can provide -- and I think they can provide -- really additional expertise and suggestions on how the data collection can occur. So, I think that is a part of the process.

And with respect to my OASIS

colleage, I mean, I have to believe that the group here can provide advice to the NQF group as to what are the criteria. Maybe I am being eternally optimistic, but I don't think it is impossible.

MEMBER LIPSTEIN: So, I tried to convert your construct into a non-methodological researcher vocabulary, which is --

CO-CHAIR NERENZ: I didn't think it was that bad.

MEMBER LIPSTEIN: -- as it relates to sociodemographic adjustment of performance measures, including outcomes, you could leave the door closed, the way it is now. You could open the door or you could open the door and walk through it.

MEMBER CASALINO: And make people walk through it.

MEMBER LIPSTEIN: And that was in English, so I understood that.

(Laughter.)

So, you could open the door and make people walk through it.

And the reason I want to push for a third option is because I don't think we have three to five years. When you said that, Kevin, you scared me a little bit, because I do think that our regional safety nets are in a volatile place right now.

If this was the only conversation taking place in America about how to reward, punish, or pay, then we could say, "Yes, we've got three to five years to think about this." But, in the context of the Budget Control Act of 2011 and the Affordable Care Act and the American Taxpayer Relief Act of 2013, these hospitals, these providers are going to struggle if we don't move quickly.

And so, it may be that, if we open the door -- is it Lawrence or Larry? -- and Larry said we make people walk through it, we can say, "If you walk through this door, here's the upside and here's the downside."

And the downside is that we may have a bad apples that get rewarded inappropriately or a few good actors that get punished inappropriately. But, on balance, okay, we think the implications of walking through that door are better than the alternative.

And so, I would push for that third alternative.

CO-CHAIR NERENZ: Let me clarify for understanding. How is the third alternative different from Larry's one that said the door is open and you must at least say that you thought about it? I don't know how to keep translating it.

And maybe there is a flavor of this that says, if a measure can be adjusted through some acceptable technical means with available data, it should be. Is that okay?

MEMBER GROVER: Because of the time limits, also -- I'm probably the only registered lobbyist at the table. But we are

talking about the SGR and we are talking about moving towards another quality metric that would collapse the Value Modifier and PQRS and all these other things.

And again, we are going to be up in the very, very short time, hopefully, if they can get it done, to have to deal with this policy issue in the very near future. And it would be sure nice to have something out there for both Congress and CMS.

CO-CHAIR FISCELLA: Yes, Steve, I wasn't suggesting that we take three to five years to do this at all. All I'm saying is that we shouldn't let the enemy be the perfect of the good, and the process will continue to evolve and improve over time. I just want to be clear about that.

MEMBER BERNHEIM: I just want to go back to some of our earlier conversations because most of the reaction has to do with how payment penalties are applied to hospitals, and that does not have to be dealt

with through risk adjustment of measures. And, in fact, I suspect that risk adjustment of measures wouldn't actually have the impact that people wanted based on what we have done.

I mean, one of the slides I didn't show, but it is very similar to what is in that paper, 3 percent of low SES hospitals would go from being penalized to not penalized, 3 percent.

And I know we are going to fight about what the risk adjuster is and whether Medicaid status works, but the first job of these measures is to look at quality. And the problem is, if you ask people to bring data, they are going to bring data that looks like what I brought, and they are going to say there's a small difference. And we're going to say, "I don't know if it is because those high outlier hospitals have patients they just can't accomplish low readmission rates with or whether it is because they are not doing what they need to." And I worry about hanging up

the entire endorsement process if we spend all of the NQF endorsement time in the future arguing about what the right risk-adjustment measure is and whether I have tested my measure right.

And then, we end up with something where I don't know what the criteria is. I don't know what this Committee would have us do with the readmission measures, given that the differences between the low and high SES hospitals aren't that great and risk adjustment doesn't make a big difference.

So, if we open the door, I think we have to have really clear criteria and some priority based on sort of, you know, what kinds of outcomes is it more important to do this in? Do you prefer clinical factors over SES? And if they seem to be adequate, would you prefer a measure that did or didn't?

And I think it becomes complicated. But, again, I wouldn't let the payment be this thing that makes us all so

anxious that we drive forward. Because we can tell Kate right now, "We don't like the way you're doing the payment penalty," but we think we don't know enough with the quality measures to necessarily drive that fast.

I mean, that's my caution. I think you could cause more good than harm.

MEMBER LIPSTEIN: Susannah, I have to disagree. In other words, the data that you brought using the surrogates for socioeconomic status that you utilized showed that there wouldn't be a difference in payment methodology. But other papers are coming forward and other research was presented that showed something different, using other surrogates for socioeconomic adjusters. And at least we have to be open to other folks' point of view beyond the people that CMS pays to do this. I mean, don't we have to be open to other people?

MEMBER BERNHEIM: Of course. But I will say Atul's paper on the patient level

shows very little difference. The OASIS shows almost no difference. I mean, there is a real mix of data here, right?

And this is my concern: if you tell developers you have to have investigated every single possible socioeconomic variable, and if you see a difference, you have to risk-adjust for it, it is not clear to me that you are improving quality measurements and it is not clear to me that you are going to get any quality measures through again.

I'm not saying you shouldn't do this. I'm saying we shouldn't push so hard to do this that we lose track and set NQF and CMS in a position of sort of not being able to move forward with quality measures.

These measures have brought down Medicare patients' readmission rates, which hasn't happened in 10 years, right? I mean, there's a lot of people whose lives are better. There is a risk/benefit to this, and we just have to be conscious of that.

And if it is about the payment policy, I mean, it is a separate --

CO-CHAIR NERENZ: Steve, microphone.

Okay. Let me just say that this is why we all came here. This is good.

(Laughter.)

Let me, first of all, speak in support of the beginning part of your thing about that you don't strictly have to do risk adjustment of the measure in order to have some effect on how you apply the penalty.

I would point out, and I understand perhaps it wasn't clear when I went through the MedPAC recommendation, that is, in fact, what that recommendation was about. If you remember, nowhere in those slides was there an adjusted and an unadjusted measure. The measure at the hospital level did not change.

It is about how you group hospitals in order to apply the penalty. And

that is why there were some questions about, well, gee, how are your things different from Susannah? Well, they're not really because we weren't seeking to do the same thing.

So, I agree, absolutely, but I don't think that somehow splits us or takes us in different directions. At least my sense, then, is that, as it relates to any one performance measure, it may be technically possible and powerful and good to do adjustment. And I think in our open-door analogy, we are saying that that door should be open, where in the past it has not.

But we recognize absolutely that some desirable policy outcomes may be achieved by changing the way the penalties are applied without literally adjusting the measure. But that second point doesn't trump or somehow argue against the first one.

Are we all okay with that?

MEMBER BERNHEIM: And I wasn't arguing that we shouldn't consider the door.

It was a cautionary tale about how we approach it.

CO-CHAIR NERENZ: Understood.

MEMBER ADAMS: Okay. I actually had a different, completely different point. So, I don't know if you guys wanted to respond to her first.

MEMBER OWENS: So, to speak to the point about walking through the door and keeping separate the measure technical specifications defining the measure and its implementation, something stronger could be a stronger recommendation of what is required regarding implementation and what that might look like.

If you are to apply this to this program, you must consider X, Y, and Z. And that is what the measure developer is recommending or has done some testing around. I mean, it is separating out those two parts of what is a reliable and valid measure and, then, what's its application.

You could make the recommendation NQF. "NQF, you have to move towards for what purpose and what is the measure developer's recommendation of how this is a valid and reliable measure for implementation."

Do you see what I'm saying. And make that a more active -- in other words, walk through that door.

But, Steve, to your point, your focus is really on the implementation of the measure, how does that go into payment programs, rather than -- what I hear you saying, but correct me if I'm wrong -- rather than, in that particular measure, no matter how it is applied, doing something about risk-adjusting on SES. Yours is about once it gets to the payment program, once it is sent the other way.

I don't know. I'm just trying to disaggregate it and to come to some compromise.

MEMBER GROVER: I think it

depends. And certainly, we had this conversation when it was being implemented on the readmission side.

If you do it in the risk adjustment itself, are you disadvantaging particular individuals who are at risk in some way versus, if you stratify, which is what we eventually came up with as a proposal, in terms of how you implement the payment side of this, can you do it in a way that helps not already disadvantaged providers, but that doesn't sort of treat a whole class of citizens more poorly?

MEMBER LIPSTEIN: I think, Pam, what I was responding to was the idea that the end goal -- and this is what I brought up -- the idea that the end goal is to bring down Medicare's payments for readmission rates. So, that is the end goal, and that is the definition of quality.

Then, I think what I want to do is illuminate the fact that, while we bring down

the readmission rate for Medicare, there are other costs. So, in other words, all the money that is being spent to reduce those readmission rates is being shifted to somebody else because Medicare isn't paying for that. It could be shifted to commercial payers.

But in those hospitals that can't do cost-shifting because they don't have a commercial payment base, they don't have anybody to shift it to. So, they're paying the penalties. They don't have a commercial payment base, and they are beginning to really suffer.

And so, what I want to say is, when we walk through that door, we need to illuminate the fact that the end outcome may not be an improvement in quality. It may be just a big cost shift. And that is what we have got to come to grips with.

MEMBER CASALINO: I'm sorry about that. I'm from New Jersey; I'm used to just speaking louder and faster, like certain

politicians there and other New Jerseyites, or at least the latter part.

(Laughter.)

Originally from New Jersey.

I think, you know, Pam and Susannah raise a really good point. A lot of the energy in the room comes around really how measures are used for payment or for public reporting, three of us and other people, many other people in the room, I mean three of us most recently, that is where a lot of the concern comes from.

And that doesn't, as Susannah pointed out and Pam, really have anything intrinsically to do with whether the measure itself has to be risk-adjusted statistically, as opposed to what gets done with the measure once it is there.

So, I think that, for me, and I think for some of us, that is a distinction that, while it is clear, in a way I think it is easier to keep forgetting it, right?

So, it does, in terms of these three possibilities, leave the door closed, open it and let people walk through, or open it and require people to walk through -- what are we talking about? Susannah and Pam took it that we are talking about the measure itself; should it be statistically risk-adjusted so that it will automatically get used for payment or pay for performance, right?

There might be different feelings about requiring people to present why they do or do not want to risk-adjust the measure if NQF was given some thoughts about what might be appropriate uses of this measure in pay for performance and public reporting.

CO-CHAIR NERENZ: Let us turn to Alyce.

MEMBER ADAMS: Okay. So, this discussion has been fantastic and really interesting.

One of the things I am still

struggling with a little bit is this issue of the construct of race and ethnicity as being similar or separate from socioeconomic status. And here's my issue, my dilemma:

So, in terms of quality metrics, we are moving towards a world where we consider equity to be a component of quality. And in our goal to improve equity, we actually want to compare across racial and ethnic subgroups in an effort to identify differences that should not be there.

And as part of that construct, we want to control for things like clinical differences, but not for SES, because of its tight correlation with race/ethnicity, right? So, that is sort of the one piece. It is stratification, but not for risk adjustment, but, rather, for direct comparison purposes.

We, then, on the risk-adjustment side are talking about SES, race, ethnicity, gender, and age as potential risk adjusters. And that's where I get a little bit stuck on

race/ethnicity. And the reason is there is, it seems to me, a qualitative difference between saying the resources of a particular neighborhood are such that it disadvantages the providers or it puts additional barriers up for the patients. It is something else to put a variable that says "X" percent Black or "X" percent Hispanic should be adjusted for. Do you see what I mean?

Because, to me, that doesn't necessarily speak to resources per se, but it is, rather, wrapped up in all these other SES measures. And so, I am really struggling with that particular component of it.

Gender and age are similar, except that most of our quality, many of our quality measures anyway actually already adjust for gender and age based on the nature of the quality metrics themselves.

And so, part of me is wondering to what extent, if we believe that race and ethnicity are not primarily measures of

biological differences, but primarily measures of socioeconomic differences and things like that and social/political history, what have you, are we doing well enough with our clinical adjustment to say we can get rid of those variables; we don't need them in there?

And so, I don't know what others of us are feeling, but I am really struggling with that, of whether or not those variables should really be a part of, you know, sociodemographic measures that we actually use for risk adjustment specifically.

CO-CHAIR NERENZ: If I could just observe, a perfectly good point. I think, at least to my ear, that is in the "how". Which variables, which relate, which are markers for what others?

So, rather than get deeply into that in terms of response, I guess I think this can surface many times over the next day and a half or day and a quarter, or whatever we have.

Okay. Norbert?

MEMBER GOLDFIELD: So, two comments. At the end of the day, the reason I'm here is because it is very clear to me that certain types of SES variables must be included in the risk adjustment.

On the payment side, the response that people talked about the payment side, there are 50 different ways to screw the patient or the provider. That's just how it works. And that's okay. You know, we recognize that.

So, I'm not interested in all the different machinations which Bob Fetter, and so forth, were among the world's experts in how to do that.

What I'm interested in, and what I often think about, is what is a poster-child example? A poster-child example of SES is homelessness. Okay? So, if we can get, I mean, from my perspective, if we could have homelessness built into risk adjustment, and

I'm more than willing, in deference to Susannah and others, to do a data collection, which is what we are going to do in several states, to do a data collection and understand.

But, from my perspective, we really not, even though at the end of the day it's all payment -- I mean, I get that -- but part of payment is having a risk adjustment that takes into account all the different populations. And if I were to take a poster-child example, if we can't agree on homelessness, there's nothing to talk about, frankly.

MEMBER SUGG: So, I know we focus a lot on the readmission data because it is on everybody's plate right now, but I just want to make sure we come back to the other disparity that is going to happen if we don't do socioeconomic adjustment. And that's that, when physicians get their reports and it shows that they are poor quality because they take

care of certain groups of patients, they will, voting by their feet, leave that and just increase the disparity.

And I just want to make sure that that piece doesn't get lost with our readmission piece being so prevalent. I think it is easy to do because we have more data about the readmission, and we don't really have hard data, but I think everybody around the table can pretty much imagine what will happen as far as people being willing to serve in those underserved areas.

MEMBER BERNHEIM: Could you just say what measures we are referring to? Because I think it is really important that we don't -- because the whole world isn't about readmissions because I do think it is different for different measures. So, what kinds of measures are you referring to in that setting where you think it is going to be a problem?

MEMBER SUGG: So, I am talking

about the public reporting, like physicians' individual quality scores that will be available on the web. And if you do these quality scores and not risk-adjust for the socioeconomic parameters of the patients they serve, then it will make individual physicians less likely to take on risky patients.

CO-CHAIR NERENZ: Some of these are HEDIS-like measures: mammography rate, medication. But some of them are maybe outcome, and it depends on the context. It depends on the payer. A1c control, for example, is an --

CO-CHAIR FISCELLA: So, I have been thinking about the issue of, you know, relatively-modest effects. I agree that, at least for the outcomes we have looked at, that the effects aren't enormous, although I think, as some of us have said, that if you are doing multiple reporting, over time these effects can be cumulative as well as set up feedback loops.

But in some ways I think the fact that the effects are moderate is perhaps a good thing in terms of moving forward because it means, one, that the really egregious providers, adjusting for SES is not going to adjust away the fact that they are doing a horrible job.

And it means perhaps that initially for things like rehospitalizations that the shifts are not going to be seismic, for the reasons Norbert has already alluded to. Politically, it would become very, very difficult if there was a huge initial shift with this. So, it actually may not be such a bad thing from my perspective.

DR. BURSTIN: Just a comment, and perhaps some of this is I think we need to sort of do a little bit of discussion at your break and perhaps bring a formal proposal forward that I think we can work with.

I mean, at least the way I see it, we are in a transition phase. We currently

have one stop, yes/no, endorsement, which I think makes this problematic, and we recognize that.

I do think, though, for the sake of this discussion, I think we have clearly heard that we should allow developers for certain kinds of measures, using certain kinds of variables, all TBD to follow our discussions, to consider you putting sociodemographic variables in their risk-adjustment model if justified.

I think what we are clearly getting into the discussion of, the forcing it is I think more about the implementation of the payment, and it may be that we can actually add a second principle there. And Kate may be helpful here in helping us think this through. For certain kinds of uses, there may be more of a suggestion to demonstrate that you should do it or not do it as opposed to leaving it fully optional.

So, I think there are ways for us

to create a pathway here that I get the sense people would agree to. But perhaps during the break we could try to write that up.

MS. PACE: But I think that, you know, we are going to have tomorrow to come up with specific recommendations. So, what we were hoping to do today is really air all of the issues, the pros and cons, the things that we need to consider. And obviously, we have multiple ways that we can try to move forward.

I also want to challenge us to separate the policy response from the performance measurement. If we really have a reliable and valid indicator of quality, why can't that be used for public reporting versus pay for performance versus accreditation? What is it that is going to change that reliable and valid indicator of quality?

So, to me, the question is -- and I think our statisticians can help us with this -- if we put something into a statistical risk model, and we start looking at the

results of that, I think we are basically asking the question, if the average provider cared for my case mix, this is what we would expect the outcome to be.

And so, the question is whether looking at the average provider -- right now, we are looking at clinical and health status factors. And so, the question is whether looking at those patient-level SES indicators, whatever they might be, need to go into that mix when we are asking the question, if the average provider took care of my mix of patients, what would the outcome be?

And I think part of that question is what kind of unmeasured clinical severity is accounted for by those factors. And, you know, one of the things that we have talked about -- and I will leave you with this question. We can go to break and we will come back and talk more methodological.

But if we did include these in risk adjustment, and we have these questions

up here about the responsibility for addressing these disadvantaged populations, and many of you have already presented the higher cost of having interpreters for people that English is not their primary language or higher cost of taking care of patients that don't have a home to go to when they leave your hospital.

But there are some strategies that can address those things. Is that part of the quality question? Who is responsible for that?

But, if we adjusted for those things, we would still, I think -- and this I my question -- so, if we had those things in the risk adjustment and we're asking that question about, if the average provider took care of my mix of patients, would we still see the difference in applying those good strategies for taking care of these more difficult patients?

So, I don't know if we can go to

break with some of those questions, and then, come back and really get into some of the key things about how we would select risk factors and how we would adjust. But I just wanted to kind of lay those out there for you.

Why don't we go ahead and come back at quarter until 4:00, 3:45?

Thank you.

(Whereupon, the foregoing matter went off the record at 3:27 p.m. and went back on the record at 3:46 p.m.)

MS. PACE: Okay. We are going to reconvene.

(Pause.)

Okay, everyone, we are going to reconvene.

We are really going to move into the "how" questions. We on our agenda had this kind of parsed out to first talk about the factors and, then, the methods.

And what I am going to do is just present a couple of slides, just in usual kind

of considerations, just kind of to frame this. And then, we will take off from there.

And I'm going to go ahead and do a few slides on adjustment models, as we have been talking about. And then, we will have the open discussion, because I am afraid we will end up kind of crossing those two topics anyway. So, we might as well get this out of the way.

So, I just wanted to put this out here. These certainly are things that we can add to or subtract from, but these tend to be some of the usual considerations for selecting risk factors.

One is that there is a clinical or conceptual relationship with the outcome of interest. Usually, we look for an empirical association with the outcome of interest. As part of that empirical analysis, often looking for contribution of unique variation versus redundant. If, you know, two variables are basically highly correlated and accounting for

the same thing, you may not need both of them.

From a risk model standpoint in the context of quality performance measurement, we want things that are not related to the quality of care because that is what we are trying to do, is isolate differences that we want to attribute to differences in quality.

So, we have already talked about we really want to focus on things that are present at the start of care, not things that happen days into the care that is started.

Accurate data that can be reliably captured, and I think that is something that we are going to need to come back to because that certainly is a consideration. And data limitations often are a practical constraint.

And often, you know, we want to see improvement in the risk model metrics. Does it improve discrimination? Does it improve calibration? We have talked about improving the moderate effect on the overall

r-squared or C statistic, but one question we might want to look at is, does it improve calibration when you look at different levels of caring for patients with these different factors?

Okay. Next slide.

So, we will come back to this in terms of your thoughts about how you select risk factors and how that applies to sociodemographic factors, whether there are some of those that don't apply or new ones.

So, let's move on, and I am going to just mention a few things about the methods. And basically, what many of us are used to is comparison of observed-to-expected outcomes for the accountable entity, often indirect standardization, which is, then, extended to multivariable statistical models.

Go on.

And as we already said, the models for risk adjustment are used to isolate the effect of the quality of care. And I think,

as Susannah and others have mentioned, they purposely do not include all the variables related to the care provided. You know, if your goal is strictly to predict the outcome, you would include lots more, but, purposely, we are not including those variables related to treatment and the care processes versus, as I said, an explanatory model.

And so, one thing that we need to keep in mind when we are looking at model metrics -- for example, r-squared or a C statistic -- they are not necessarily going to achieve the same values that you might have when you are doing a total explanatory model that you are including those treatment kinds of variables.

Okay. Next slide.

The other thing that we want to talk about, and certainly we can have some discussion, I think people seem to be in agreement that stratification is a way of doing risk adjustment versus just these

statistical models that we often think of. And stratification could be done by constructing risk categories based on SES or other sociodemographic factors.

And this could be done, as we have talked about, within a provider organization or accountable entity or something that we are kind of terming "organizational stratification," like David presented that is what the MedPAC recommendation is. And certainly, then, there could be combinations because, actually, the MedPAC recommendation, they are using the risk-adjusted, the clinically risk-adjusted performance rate and, then, stratifying organizations. And you all may have some other suggestions that we should be considering.

Next slide.

So, this is just to illustrate what we're talking about, stratification within an accountable entity, the hospital, the physician, where you actually would look

at the patients served by that organization and perhaps dividing them into quintiles by median income that was determined by the Census tract of where that patient lived. And then, each provider would have five performance rates for each quintile of the variable.

I think, as was pointed out, when you have low numbers to begin with, and then, you have a provider dividing those cases into five, we get into other issues with reliability.

And then, the next slide.

And this is just another representation of what the MedPAC recommendation is about, what we're calling organizational stratification, where you just use the performance rate using, in that case, already risk-adjusted for the clinical variables, but, then, using stratification to identify peer groups in terms of looking at comparing performance within peer groups.

So, I think that is my last slide.

So, I think, with that, I'm going to turn it back over to Kevin, and we want to have some discussion. We can decide whether you want to try to start with factors or leave it open to both factors and approaches to risk adjustment.

Kevin?

CO-CHAIR FISCELLA: My suggestion is to continue with what we have and go to risk factors. I know Alyce raised this question as sort of a prelude to getting into this.

So, it gets into what are the sociodemographic risk factors that should be considered. What are the criteria we should use both at an individual level and at an organizational level, since those are both open questions on the table, if we are going to stratify under some type of stratification, at what level it would be done?

So, why don't we just open it up

to thoughts about criteria for those sociodemographic factors that warrant consideration?

MEMBER SUGG: So, I wanted to follow up with what Alyce started talking about with race because that is something that I have been thinking a lot about recently.

And some of this actually comes from my front desk staff who tell me that, when they are doing the intake for registration and they start to talk about race, they often get perplexed looks from the person standing in front of them. And then, when they get to ethnicity, they get even more perplexed looks about, well, first of all, why are you asking? And secondly, what does that really mean and how do I identify myself?

And I especially have issues around ethnicity because at Harborview we have the International Medicine Clinic. And so, we have a large Southeast Asian and Somali population that we take care of. But when I

see somebody who came over as a refugee 20 years ago, ethnicity in some ways makes sense to me. When I see their granddaughter in clinic, not so much because they're Americanized, and although ethnically they relate they're Cambodian, I don't know that that particularly changes anything I'm going to do medically for that person; whereas, it definitely would with their grandmother.

So, I have spent a lot of time kind of figuring out why do we still ask this. I mean, part of it is disparities. We want to make sure that groups are not being discriminated against, and I think that is an important variable. But, beyond that, I actually am not sure how I use that clinically anymore.

MEMBER LIPSTEIN: Yes. No, if we're having any conversation about what risk factors -- and, Kevin or David, keep me honest there -- when I raised the issue earlier of Census tract housing vacancy rate, you would

call that a risk factor?

Okay. So, the article that we presented, which the manuscript is with the peer-reviewed Journal now, and that will become available during our work here. So, the Census tract variables that I presented in that paper, including especially focusing on the one that didn't get a lot of discussion, was high housing vacancy rates, and that would apply to -- so, for example, the discussion we were having just before the break of individual doctors, and individual doctors who serve those Census tracts characterized by difficult life circumstances, that would apply in that kind of a setting as well as the institutional setting.

I did have one other question I wanted to ask. And the question is, if measure developers -- I think that is what they are called -- if measure developers were required to look at the impact of socioeconomic variables on an outcome, and

they could conclude either they do have an impact or they don't have an impact, okay, one of the two, that would be good public information to have. If they include that it doesn't have an impact, then there would be no consequence of adjusting for them. So, why not go ahead and do it? Because that would at least quiet the people who believe it really did have an impact and didn't. And then, if it really did have an outcome, then you would want to adjust for it.

So, it seems to provide rationale that the provider community, if they really do believe that these variables affect or impact outcome, it would take another obstacle out of the way to progress in improving outcomes if we actually adjusted for things that don't affect outcomes because no one could argue the result.

MEMBER SAWHNEY: A true story. I did work, strategy work, for the Cook County Health and Hospital System, but I was using

the State discharge database. And I noted the Cook County Health and Hospital System, less, though, than any other hospital in the State, could not identify the race and ethnicity of its patients.

And I'm like, what's so wrong with these incompetent people at the front desk that they can't fill the form out correctly and get it into the database? Until I spent an evening in the ER of the Cook County Health and Hospital System, and I just literally watched people come and go. It is an experience everyone should have.

And I sat there, tried to look at people and figure out what their race and ethnicity was. How exactly do you classify the woman who is in full burka and speaking Arabic? What category do you put her in? But, clearly, she has a social disadvantage in the healthcare system. And I wouldn't put her under White, but she doesn't fit in the other boxes, either. So, if the front desk put her

under "other," I understand why they would use "other".

CO-CHAIR NERENZ: Just as a quick response to that, with all due respect to the complexities and agreeing with the points you make, three or four years ago, there was an IOM committee addressing the issue of the standard collection of race/ethnicity data, including a couple of language variables. A template exists, at least in the form of the report of that group.

And I can say, upon having implemented that in essentially faithful detail at Henry Ford, at least in our environment -- emphasized, in our environment -- it works pretty well.

So, granted, it is challenging and there are clearly places and people for whom it is very complicated, but at least that template does exist, and recommended as a standard template.

MEMBER SAWHNEY: A counter-example

is my children are of mixed race and ethnic and minority religion and wear the uniform of the religion. And they look kind of strange by our society's standards, but they are not socially-disadvantaged. So, it is a tough one.

Now, that said, there are some unique challenges that I don't know how to figure out. Granted, this is an incidence problem. It is an incidence, not an outcome problem.

But the threat of violence, if you are a young man and you're Black on the street of Chicago is very different than if you are White.

The birth outcomes, which those are outcome measures, we don't know what is going on and we sure would like to fix it, but birth outcomes are very different by race.

So, I would prefer to stay away from race, when possible, but I am not sure that is always possible. It does have an

impact in our society.

MEMBER CALLAHAN: Hi. This is Mary Beth Callahan again.

So, just kind of thinking outside the box possibly, if we took away the idea of race and ethnicity, but looked at the barriers that those things created for people, and I know that that would have some barriers itself. And I don't know if it would be a person filling this in themselves or someone at the administration desk filling this in. I don't know how we would get this data.

But if we are looking at language barriers or homelessness or low literacy or no health insurance or unemployment, so we are looking at the barriers that some of those things create. That is just an outside-the-box way of looking at another situation.

MEMBER BHAREL: I have two comments that are semi-related. First, this wasn't my main comment, but on the race/ethnicity one, I would like us to just

keep those on -- from my point of view, I think they must remain on the table. Most race and ethnicity in well-trained centers -- we actually went through a very extensive training in our Health Center -- it is about asking the patient, not looking at them. So, just to take that off the table. And it can be collected very well if well-trained individuals do it.

And we are here to talk about socioeconomic disparity, but, given the history of our country, race and ethnicity remains an issue at every level, and particularly at low socioeconomic status. So, I would like to advocate for that to stay on.

But I am here to advocate for homelessness as a factor. And you have heard my arguments about it before, but just a couple of things that have come in the readings and discussion, just to highlight.

So, homelessness, again, can be kind of this composite that takes into account

all of these different issues that we are talking about, and would be a great place to start experimenting with some of these issues that we are talking about.

In terms of data collection, I think it is really parallel to the race and ethnicity. If you want to do it correctly, it can be done correctly, and there is actually at the federal level, HUD requires, as an HMIS system that is used federally, which is the Homeless Management Information System, where every state has tracking of when individuals stay at a shelter.

So, taking away the issues of being able to cross information between two different state groups, an organization like Medicaid could see what patients, and their panel, have slept at shelters. So, there are ways to get at the data.

And Norbert is gone now, but New York is doing it at a clinic level and requiring it when there is Medicaid funding

involved. Massachusetts is looking at it as part of their payment reform, requiring individuals to report on homeless status. So, there are ways to get at the information.

The other thing, some of the questions came up around, you know, these should be consistent things that are traits. Very much agreed. Most homeless individuals, thankfully, are transiently homeless, 80 percent of them. So, it is a state at a given period of time. That said, to become even transiently homeless is in itself a marker of chronic stress and other issues, but it can be used if you looked at it, say, on a yearly basis.

Thank you.

CO-CHAIR FISCELLA: Just to clarify, Monica, in terms of race/ethnicity, how are you advocating that it be used? You said "collection," but once it is collected, what are you suggesting?

MEMBER BHAREL: So, my point was

really about the issue around collection to not be the barrier. So, I think if we are looking at socioeconomic determinants of health, the traditional ones have not -- you know, it is considered in disparity, but not necessarily in sociodemographics. And I would advocate for including in something that might be measured and, then, adjusted for.

CO-CHAIR NERENZ: Okay. Actually, in my sequence -- and a quick apology -- I said Alyce, and I heard someone else say "Alyse," and I wasn't listening carefully. What is correct?

MEMBER ADAMS: It's Alyce. That's okay. And its spelling, it throws everybody off. It's all right.

CO-CHAIR NERENZ: All right. We will get it correct from now on. Okay.

MEMBER ADAMS: Okay.

CO-CHAIR NERENZ: You are next.

MEMBER ADAMS: So, just briefly, I just wanted to reiterate, it is true, the gold

standard now is to ask people what they are, not for the provider to make that assessment. And that's helpful.

But I think it is also important -- and I don't want to get too far down the rabbit hole of race/ethnicity -- but this issue of sort of would it matter. So, I am sort of going down this list and trying to check off each of these boxes, as it were some other type of factor.

One of the things I run into is that, with race/ethnicity, truly it depends on the subgroup. So, sometimes what you will find, for example, is that when you control for other socioeconomic factors, sometimes race falls out, but not always.

And so, trying to figure out race/ethnicity as a single factor is challenging in the context of risk adjustment because, yes, sometimes it is almost saying the same thing as SES; sometimes it is not.

And so, I really struggle. I am

not sort of advocating for throwing it out completely, but I do think that it requires a little bit, a lot of thoughtfulness about exactly what do we think we are capturing when we talk about race/ethnicity. And my guess is it is not truly biology, nor is it purely SES. So, we just need to figure out what that is that we are trying to capture with that and whether or not we're getting it through other means.

MEMBER PONCE: Thank you.

So, I think for race/ethnicity I have the other concern, which is that, if it is included in a model, it might be picking up. So, racial/ethnic might look like they are doing better, but it might be an access problem. So, they're not getting access to care at the individual level.

So, I go back and forth whether it should be at the individual level, but at what is termed organizational or neighborhood or system level, I do think it is important to

get at the compositional racial/ethnic, you know, clinical profile.

And the ACA does have a mandate on collection of race/ethnicity. I was with David on that IOM panel. So, an Arab would be White. It is self-reported. That is No. 1. That is the gold standard. It shouldn't be assessed by the front desk.

Going through this checklist, I want to have a friendly amendment on, and building off of what Rachel said earlier on accurate data that can be reliably captured, data limitations. If that is going to be one of the criteria, then, you know, some of the variables that are really important might be thrown out.

MEMBER SUGG: I wanted to make sure people understood. The front desk was not choosing the race. They were asking the patient. And the confusion was on the patient's end of how they saw themselves.

And I'm not saying that we should

throw race out. I just think maybe it is worth thinking about how we are going to use that data and where it is important and where it isn't. But it is never about somebody checking it off and making an assumption, but it is about sometimes from the patient level they don't quite know how to answer the questions.

MEMBER PONCE: I agree. Thanks for the clarification.

MEMBER BARGER: As a researcher for race/ethnicity, it's interesting how your subjects or your patients have difficulty answering the question. And I find the younger ones actually want to check more than one race because that's one of the choices. And then, as researchers, we don't really know what to do with more than one race. And so, we sort of leave them out when we do it.

I think that the data should be collected because, especially for perinatal things, there is huge racial differences. I

mean, the difference is Black women die at three or four times the rate of White women in this country, and it is the largest disparity, although my husband tells me, no, it's male and female mortality rate that's larger.

So, I think we should collect it. I'm not so sure that we should adjust for it. I think it allows us to report about those differences.

I go along with what someone said about, at this day and age, a lot of things are geocoded. In California they geocode this chart where you live with your discharge data. Birth certificate data is geocoded. And the Census does do a good job, and there are lots of variables in the Census that Steve has pointed out, and I think Pam pointed out, that you can use to allow for sociodemographics that get at access, poverty of resources, those kinds of things, that might be a little bit more stable. So, that is what I would say.

CO-CHAIR NERENZ: Okay. We've got Dionne, Gene, and Larry so far.

MEMBER JIMENEZ: Should I press the right button.

So, the point I wanted to make was, I mean, I think it is really important to have some sort of income indicator in there, but I know we want to be cognizant that it may be difficult to figure out which one to use, based on what the quality measure is going to actually be used for.

So, my example would be for measures that are used for the Value-Based Purchasing Program. I mean, that is supposed to be only applied to sort of the Medicare population. But, then, when the measure developers are trying to do something, what if that same quality measure might be used on the commercial side or on the Medicaid side?

So, when we try to develop this criteria, we have to keep that in mind. And maybe there should be sort of -- I don't know;

maybe you have to require multiple variables that need to be tested on the side.

In terms of the race thing, I totally agree from a personal standpoint it is always very hard for me to like check that box of White because I don't feel that that is my race.

And also, you see the demographics of the country changing. But I think it is critically important that the information still be collected. Like I totally agree with what Mary is saying and Alyce and others. It should be collected, but not necessarily adjusted for, as you see the demographics changing, as well as over time people are becoming more affluent and less socially-disadvantaged, you know, for some parts of the race. So, I think you are leaving out a big group of low-income White people as well, too.

MEMBER NUCCIO: I would just like to chime-in on the idea that I'm really not sure what race measures or what it represents.

One could argue that what we are trying to measure and we claim to measure SE, economic status or something, but I would argue that it more likely is more cultural perspective on how healthcare gets delivered and should be received.

And so, in that sense, trying to use the value of race in a prediction model, I don't know what that represents in that equation. Is it representing the economic status of the individual? Is it representing decisions about whether or not you should or should not receive vaccinations and when you seek healthcare?

So, I am really not terribly excited about including it in an equation because I don't know what it is that I ultimately predict when it is in there.

CO-CHAIR NERENZ: I am wondering if I could just call this friendly amendment, but I am just checking the framing issue for our discussion. When we talk about any one of

the social demographic variables, so let's just now use race because we have been talking about it, presumably, when we speak favorably about it, we're not crossing the line to say that it should or must be included in all measures in all possible models. Again, I think we are talking about "can" or "might". And again, I see a few nods around the table. We ought to check that.

I also, I think, have the auxiliary assumption that the actual interpretation or meaning of it in a model, meaning how it influences the dependent variable, can vary measure-by-measure. And it may be a proxy for something like income or education somewhere, but it may have a different influence someplace else.

I am imagining that a measure developer in explaining an adjustment model, bringing it forward to NQF, could include either a verbal or a diagrammatic conceptual model of how the included variables are

presumed to influence the outcome, whether they are direct or indirect effects, whether they are variables that are presumed to essentially mean what on their face they say, as opposed to being the best-available proxy for something else that cannot be measured.

So, again, I am throwing out a few assumptions here, but I think what is in my head when we have this discussion about variables is mainly that they can be useful, but we are not seeking to decide whether they must be included. Is that fair? Okay.

MS. PACE: But I do think, if there is agreement on something that should not be included, that is certainly fair game as well.

CO-CHAIR NERENZ: Yes, and I would be okay with that. And then, presumably, there would be a rationale why that is strictly a bad idea and why it may be a bad idea across the board, if we have that.

Okay. Larry, you were next?

MEMBER CASALINO: Two points. One is, you know, all day this has been a very hospital-centric discussion and a very hospital-centric/readmission-centric discussion. And it is easy to understand how that happened, but it is a mistake, I think, right?

So, if we talk, for example, about how data can be -- or we talk about race/ethnicity. Where does that data come from? CMS has data on that and can say whether it is good or not, but, right, it's there at CMS?

There's Census data on that, right, which you could use not for individuals, but you could use for -- you could see where the Census tracts are that the hospitals confirm or medical groups, and do some kind of organizational-level assessment there.

But are we talking about physician offices collecting this information? And how

accurately will they do it? And I think any of us who have spent time in small practices know that it is ridiculous to think that small practices are going to collect that information accurately. It is just never going to happen, either if it were mandated, I don't think.

So, specifically to the race discussion, but more generally to our discussion I hope the rest of today and tomorrow, I think let's try to think of other cases, because we are going to have public reporting for individual physicians. We are going to have public reporting for medical groups. We are going to have payment for individual physicians and medical groups, pay for performance.

And there is going to be lots of yowling, for example, about, you know, "It is really hard for me to get my pap smear rates up to 50 percent. You cannot compare me to a physician in Mill Valley, California, where

the women want to get mammograms every week, and they have nannies and they have BMWs to drive and get there." So, I think we need to consider that case, too.

The other point is just we can spend the two days discussion race/ethnicity. It is an important subject. But I'm looking at, actually, the 3:30 questions on the agenda, and they are quite different kind of questions, as I read them, than what variables should we use. I won't read them, but I will just point out they are different, and I think they are interesting.

CO-CHAIR NERENZ: Agree. Thank you.

Susannah? That's too bad; I don't have anybody else with a nametag on.

(Laughter.)

CO-CHAIR FISCELLA: Yes, I wanted to comment a little bit on the perinatal issue because I think it does highlight the challenge. The relative Black/White rates of

low birth weights haven't changed in 50 years.

As you get into very early gestation births that are extremely premature, the rates get up to close to fourfold, and these are not accounted for by traditional sociodemographic variables. And, of course, this has big implications for cost, for hospitals, for readmissions to the NICUs, which are very, very high-cost areas.

So, the question, then, becomes, well, what do you do? In this particular case, and it is probably fairly unique in that we don't understand the pathways; a lot of people think it is due cumulative lifelong disadvantage and perhaps even early prenatal factors in the mother themselves, but we really don't know.

But what we do know is that there is a huge difference, and it will matter in terms of the infant's readmission and NICU stays. And so, when you have a variable like that, what should we do? Should it be

included? Should it not be included? Should it be stratified? Or should we forget race?

MEMBER SAWHNEY: I would summarize what I had said earlier, in that race would be not my priority if other variables work, but there are situations where there's no substitute for race.

MEMBER GARRETT: So, there is a point that we have been focusing a lot on the readmissions measures which raised a question for me, which is, do we consider patient satisfaction to be a type of outcome measure that we might be including in this or not? The question is whether we would consider patient satisfaction to be a type of dimension that we would in the outcome measure definition.

MS. PACE: Yes, we consider experience with care a patient-reported outcome. And so, yes, it would be in the discussion, and that is part of our questions to you all. When we talk about these things,

is there a difference by type of outcome? So, I don't know if you want to say that it should be the same or there are different considerations, but we should consider that, yes.

CO-CHAIR NERENZ: Since I don't see a nametag up at this instant, perhaps -- no, that was an omen.

(Laughter.)

Just in terms of responses to the question that Karen put forward, are there any variables in this set that we are talking about that people feel should not be included as a matter of principle? Karen talks about there were no answers to that. I may take that to say that we don't think there are. But are there?

MEMBER CASALINO: Let's look at Kevin's question just as a particular case. And I would like to hear what would people do.

So, this is a subject area I'm not familiar with. So, if you have higher

perinatal or you have higher premature birth rates in certain ethnic groups, say, right, even after adjusting for other factors, right? And let's say you want to look at individual obstetricians' C-section rates, and you think high rates are bad, right? And this is a performance measure or a proposed performance measure.

So, you bring in various clinical factors. And I would just ask the group, would other factors be brought in and how? What would they be and how would you bring them in?

And individual physician, an obstetrician, you're looking at C-section rates. Now, obviously, if you have a high African-American percentage of patients in your population, you're going to have higher C-section rates, everything else being equal, because you have more premature births, right?

MEMBER BARGER: African-American women do have higher C-section rates, but it

is not necessarily because they have more premature births.

MEMBER CASALINO: All right.

MEMBER BARGER: Because you don't do C-sections for premature births necessarily. I mean, it is not a standard of care.

So, it is a good question, and it is one we have struggled with. In California, we have something called the Quality Maternity Quality Care Collaborative. And I'm on the Data Committee, and we're actually given physicians now almost real-time data on their statistics, one of them being their C-section rates.

And so, then, the question is, you know, the obstetricians come back and say, "Well, you know, I have this high-risk group. I take care of obese women." You know, they have X, Y, and Z.

And so, we are now being able to sort of fairly quickly adjust. We are on the

verge of giving them the adjusted rates, right?

And it is sort of like some of the data here. So, being on the Data Committee, I have been looking at the adjusted rates. And really, once you adjust, it still doesn't make a whole lot of difference, but it will certainly make quiet their objections to being compared. And so, I think, from that standpoint, I am for adjusting, just because I think, then, there is no way that they can sort of say, "It doesn't really apply to me."

MEMBER CASALINO: How are you adjusting for that?

MEMBER BARGER: Oh, we are adjusting for age, ethnicity, BMI, comorbidities such as preclampsia, which is a reason. So, we are adjusting for preexisting conditions.

MEMBER CASALINO: Income?

MEMBER BARGER: We don't have income. So far, we are not doing income or

SES. Maybe we did. I wish I had them. I don't have my computer here. I just have my iPad, so I don't have all the things, but a fair number of things.

MEMBER CASALINO: And you are adjusting for individual physicians?

MEMBER BARGER: Yes.

MEMBER CASALINO: So, there is no stratification. So, this makes differences invisible --

MEMBER BARGER: Uh-hum.

MEMBER CASALINO: -- and are of different patient groups, right?

MEMBER BARGER: Uh-hum.

MEMBER CASALINO: They just get a single number adjusted, right?

MEMBER BARGER: Uh-hum, uh-hum. They just get an adjusted rate.

So, anyway, but it is also done on an organization level. So, I mean, it is a pretty very cool thing. So, each hospital who is part of the group can look at within their

kind of hospital, within their region, within the state. So, it provides like the benchmarks, the average for the state or within their teaching hospital group, or whatever kind of group they want to compare themselves to. So, it's pretty cool.

And then, where you are getting dinged, it actually gives you the -- if you are the hospital, you can find the patient that you are getting dinged for. And then, you can go back and look at the records and, then, go to the provider and say, "Why did you do this C-section?" or "Why did you do this induction?", or whatever. And if it is coding issue, they can fix the coding issue.

So, it is now making sure that the data is really, really clean, because, then, the physician would say, "Oh, the data was bad." Well, then, here's the data you gave us. Either fix your coding or is it correct," right? So, I mean, it's cool.

So, I think that it is fine to do

it, but I want some data that we can use. Some of the problem is this sort of dual race, you know, where do you put those people? And sort of what another person said is, I think what Gene said is, what are we measuring? In the perinatal, we are measuring some long-term kind of thing because we know, even if you control for acquisition of education, acquired wealth, all of those things, the perinatal things still are there. Among high-income women who are very wealthy, who are Black versus White, there's still a huge difference for pre-term birth. So, I think that it is a proxy measure for something beyond access, beyond where you live, access to grocery stores and resources.

CO-CHAIR NERENZ: Sean? Then, Steve. Then, Susannah.

MEMBER O'BRIEN: I was just thinking it would really help clarify my own thinking if somebody could give an example of a measure or a scenario where you would really

want to avoid adjusting for a sociodemographic factor. But I think of lots of scenarios where you could very easily defend not adjusting variables, if you had issues of data availability and cost of collecting the data or data quality or a situation where just adjusting really made no impact. So, I can envision lots of those scenarios.

But it used to be that you didn't need to really worry about the reason for not adjusting because the NQF policy said the door was closed, and that was your reason and you kind of said, "Well, we don't want to mask differences." And you didn't have to really think too hard about whether that held water or not.

But if we are opening the door at this point, then, presumably, that is no longer a good enough reason. And I know there are good examples. I think it would be helpful if someone could kind of lay out a way of thinking about it and an example of here's

a scenario where you really wouldn't want to adjust for any demographic variables.

CO-CHAIR NERENZ: Go ahead.

MEMBER COHEN: Yes, it is not so much that we don't want to -- I'll talk about the miSCRIPT program, which is purely a quality improvement program. And we might get slightly better models, but it is not worth it.

So, essentially, in miSCRIPT we look at 30-day surgical outcomes. So, it is an acute situation, and all of the risk factors are probably all represented in 30 or so clinical variables, even though we do include race and ethnicity. But it is purely quality improvement. Money is not involved, which makes a big difference, you know, I would suppose.

And the data that we are using we give to hospitals so it is blind to everyone else. And it is sufficient for them to do their drilldown or to do comparison to other

hospitals to see where they fall on many different outcomes, so they can allocate their resources well, even though there might be some unknown bias with not including these things that might take a little effort.

The difference, we expect this to be small, and it is not really worth additional complexity, you know, to go through that process.

And also, a very important part of the program is to identify best hospitals for purposes of case studies and leadership and that. And that really won't change very much if you do it. So, it is not a matter of it doesn't help, but it is really not worth the resources to do it in that sort of pure quality improvement context.

Does that answer it?

MEMBER O'BRIEN: Yes, I think that is a scenario where you can very well defend not doing it. But I guess I would kind of reframe my question. Imagine a scenario where

the data were available. They were high quality. It made a difference in terms of how performance was assessed. In one of those situations, what is an example where you would be doing the wrong thing by adjusting?

MEMBER LIPSTEIN: Okay. So, Sean, what I was going to say was I would never ask the question, "Can you think of an example where we should not risk-adjust or we should risk-adjust?" It's always risk-adjust and not risk-adjust. You always want to do both.

And the reason, you want to look at unadjusted data and you want to look at adjusted data, and you want to be able to look at them side-by-side because that illuminates for you, or, hopefully, it will illuminate for you whether something makes a difference or it doesn't make a difference. So, you know what you should be working on.

And so, one of the reasons why I always nervous about risk-adjusting for race is because the interpretation of that is

sometimes that either the provider -- there's a discriminatory situation as opposed to a disparities situation, which is you are using race as a surrogate for something else. And because people of color, at least in the cities where I live in, are not randomly distributed across your communities -- in every city I have lived in there is an Apartheidian element along racial and ethnic lines.

So, when you look into small -- somebody talked about small building blocks of geography -- what you find it either there are access disparities, service disparities, prenatal disparities, to get at Larry's outcome issue around birth weights. But you want to do the "and", not the "or".

And it becomes very illuminating. And we can talk about birth outcomes or we can talk about readmissions. You want to illuminate what is happening.

MEMBER SAWHNEY: I would also say

that sometimes you cannot separate cause and effect or what happens before someone comes in from what happens once they get in. So, there are certainly scenarios where there could be racial differences in outcomes. And yet, that may be because there are, in fact, differences. When patients present a problem, the disadvantaged and the advantaged were put on different treatment paths.

And if you just adjust the outcomes according to national or state index of racial difference outcomes, then you have masked that. You have masked that problem.

So, was it the race? Was it a disadvantaged person walked through the door or was it how it was carried through the system, through the treatment? And that can happen. I am hoping it doesn't happen often.

The other thing is, it was alluded to earlier, but let me point out there, you know, let me put the White rednecks of the world -- and I don't meant that too

pejoratively because I come from a White rural low SES area, and those areas can be very challenged, too, the areas east of East St. Louis in southern Illinois.

CO-CHAIR NERENZ: Okay. I have at the moment Susannah, Helen, Nancy, Larry, Kevin, in the order of plackets going up. Does anybody want to go out of order because speaking immediately to a preceding point?

Okay. Yes, go ahead. Go ahead. Good.

CO-CHAIR FISCELLA: Yes, I'll be quick. One possible example might be on the experience of care, like the CAHPS measures, where a person feels not respected. I believe, at least from the literature I have seen, that African-Americans are more likely to have that experience in hospital settings. And the question becomes, is that something you would want to adjust for or not, for example?

DR. BURSTIN: And just another

thought. I mean, just to the issues that came up earlier about whether it is within someone's control, whether it is logical to actually adjust for something, one could make the argument, for example, CLABSI, central line bloodstream infections within a hospital, completely within the control of the providers, would not be necessarily an outcome perhaps that this group would think should be adjusted in that way.

So, I don't think we want to have a blanket statement about all outcomes, either. I think it should have a logic model here for why would you would adjust or not adjust. And I think that was brought up this morning in a lot of the discussions.

CO-CHAIR NERENZ: Susannah?

MEMBER BERNHEIM: Just to add on that, I mean, the sort of obvious thing is, if the patient of low socioeconomic status, those hospitals are doing worse, if you knew that it was because they were providing lower-quality

care, you wouldn't want to risk-adjust for it, right? The SES, is it really a marker of poor-quality hospitals? And that is the sort of classic scenario in which you wouldn't want to be adjusting. And it is also what we are trying to tease apart, as sort of how much is a quality issue versus a patient-level, inherent factor.

MEMBER BARGER: I think part of what Larry was trying to do, to try to use the race and premature birth example to say sort of, can we start to say something about criteria you would want to use? And I am going to throw a couple of things out there for people to react to.

So, I would say, if we thought race, in that case if we thought that it was really just a marker for more underlying disease, I would preferentially risk-adjust for the underlying disease rather than the race. If we thought it was a marker for quality of care, I would not adjust for it.

If we think it sort of meets these criteria, which those criteria aren't up anymore -- we are looking at a different slide number. Can you go back to Karen's where we think that there is a relationship with the outcome that both conceptually makes sense, which I think is really important, as well as empirically and unique. So, it is not actually a proxy for something that we can better measure. And we don't think it is because they are getting poorer-quality care.

And I think that the example you gave is one of the few where you might think about race because it is so well-studied and so hard to understand. There are very few other situations where I would argue that race was the better variable. So, in general, if I was making a list of things you would consider adjusting for, I would put race lower down. But I think in the unique circumstance where a lot of research has gone into it, and it seems to be biologic, it doesn't seem to be

mediated by care quality at the time or even during the prenatal time, and there is some conceptual model that somebody has that makes sense, then you are starting to get to sort of criteria for a measure where you would use race. And I think you could start to do something similar with other variables.

CO-CHAIR NERENZ: Nancy, you're next.

MEMBER GARRETT: So, I just wanted to respond to -- a couple of people have mentioned, I think Alyna and Steve, the importance of doing this both ways. And I agree with that. Analytically, I think that is really important. I think it is very challenging, practically.

I mean, the current process it that a measure is endorsed from NQF, and then, it is released into the world to be used in many different ways. And you would almost have to have two measures, one adjusted and one not adjusted.

And so, to give an example, the diabetes measure that I talked about that was NQF endorsed being used in Minnesota, then adopted by pay-for-performance programs, by payers, being used for public reporting. But, at the same time several years ago in Minnesota, there was a lot of noise around this issue about, well, it's not fair; we need to look at risk adjustment for SES factors.

And so, Minnesota Community Measurement created a measure that is risk-adjusted for SES. And what we used was a very rough proxy. It is payer. So, there are three payer groups, Medicare, Medicaid, or commercial. And so, the rate is risk-adjusted by payer status, which I think has a lot of weaknesses. There's lots of variations within the Medicaid population, for example. But it is a step towards having a risk-adjusted measure.

That measure is on page 150 of a 200-page report. Most people don't know it is

there. If you go to the website and use the public reporting portal, you see the unadjusted rate, and that is what is used for all the pay-for-performance programs, et cetera.

So, I think while doing it both ways really makes sense analytically, I think we have to make a recommendation of whether risk adjustment should be part of that endorsement process for the measure that is released.

MEMBER LIPSTEIN: The example I gave about test scores earlier from third-, fifth-, and eighth-grade reading scores is why I kind of take a different point of view than that. Because if you adjust a student's third-grade reading score so it looks like he is reading on the third grade, but he is really reading at kindergarten level, but you have adjusted it because he comes from a single-parent household as opposed to a two-parent household, or something like that. You

don't want to mask the fact that the student isn't reading at grade level because you want to be able to get that child the resources they need.

And so, that's why the unadjusted score is really important. What I don't want to do is not adjust that reading score, so that all we do is take resources away from the child that's not reading on third-grade level. And that's the challenge I think we have by only reporting one way or the other.

MEMBER CHIN: This is Marshall with a question maybe for NQF staff. The issue of at what stage to bring in risk adjustment in the NQF process, whether it is upfront where measure developers who are looking for approval of a measure are asked, for example, to show how the measure performs in different strata, in different races, ethnic groups, for example, as well as providing sort of an appropriate risk-adjustment tool for given purposes. I mean,

that is one option.

The other is, if it is not part of the approval process, but it risk adjustment is an issue for the user. So, CMS, for example, or if a state like New York is doing cardiac report cards. Can you tell us a little bit about, from your perspective, at what stage the risk adjustment comes in from NQF's perspective?

MS. PACE: This is Karen Pace.

And typically, we want that as part of the measure that the Steering Committee and, ultimately, the membership and the CSAC and Board endorse the risk-adjustment model as part of that because there's a couple of reasons.

One is we are endorsing a national standard. And so, you know, if we just endorsed the base measure, and then, say the implementer adjusts it in a way that works for them, then we have kind of moved away from a national standard. And people really think

that how it is risk-adjusted has direct applications for its validity as an indicator of quality.

So, we to this point haven't considered that part of the measure that is examined in terms of NQF endorsement.

Does that answer your question?

MEMBER CHIN: I think so. I guess the second part is, then, well, it came back to the beginning about like different purposes. So, I guess like, then, you know, would NQF then say, "Well, we're endorsing this measure and this particular risk-adjustment formula for purpose A or purpose B."? What is the thinking there?

MS. PACE: Well, again, as we said at the beginning, currently, NQF endorses measures that are considered suitable for accountability applications. And I guess the shorthand way of describing that is, if you have a reliable and valid indicator of quality, the thinking is that, you know,

reporting it, public reporting requires a valid indicator of quality. Pay for performance requires a valid, reliable and valid indicator of quality.

So, I guess we haven't -- and certainly, it is open to discussion if you can come up with specific rationale why you would have, you know, a different risk-adjustment model because you were going to use it in payment versus using it in public reporting, keeping in mind that what happens to it in policy is not just about the computed performance measure; it is, then, about how it is looked at in terms of putting policy around it.

But, you know, that is certainly open for discussion. We, to date, haven't come up with a strong rationale of why a measure that would be considered reliable and valid for public reporting would not be a reliable and valid measure to be used in payment, pay-for-performance program. But,

definitely interested in hearing discussion about that.

MEMBER CHIN: Thank you.

CO-CHAIR NERENZ: Okay. I don't want to close off -- at this moment, what I have, I have Larry, Nancy, Alyna, and Pam. And, Susannah, is your tab up? Okay, I thought so.

Let me just suggest, after those four, let's just do a quick pause and a time and agenda check. But let's go through those four who have indicated they --

MEMBER CASALINO: Yes, I mean, Sean's original question, and Susannah's response, and then, Marshall's question and Karen's response, made me think some more about this question of NQF opening the door and saying, "You can step through this if you want," as opposed to NQF opening the door and saying, "You have to step through it. And `have to' means you have to tell us why you are or are not proposing some SES thing as a

risk adjuster."

And, you know, I have to say I find it very hard to think about. I think I have, and I think maybe we all have -- it is hard to think about SES versus clinical factors. I mean, I am asking this not to make an argument, but as a sincere question.

If there is some outcome -- let me not specify an outcome -- but we know that there is some clinical variable that makes that outcome more likely, then there's no question that NQF is going to require adjustment for that variable. NQF is not going to say, "Well, you may adjust for this if you want." You know, you have to do it, right?

And again, I don't mean this as a rhetorical question. It is a sincere question. So, if it is also shown that coming from a low-income Census tract, say, a very low-income Census tract, is also associated, and independently, with this outcome, with

this poor outcome, is that different or is that not different from the clinical variable being associated with the outcome?

If it is different, then why wouldn't we require the data be adjusted for. And if it is not different -- if it is different, how is it different and why is it different, and how do we deal with that?

So, low-income Census tract, big effect; clinical variable, big effect. Are they different? Are they not different? What are the implications?

MS. PACE: And I think that is exactly what we are trying to say. If these are the considerations for identifying clinical variables, is there any reason that these same things don't apply equally to the sociodemographic ones? And what would be the rationale for saying they shouldn't be included?

I mean, to date, the rationale has been around this idea that including them, for

example, in a statistical risk model obscures differences. And because we are concerned about disparities and wanting to identify and reduce them, that the thinking was that that was adding to the problem versus the solution.

But, you know, we have had a lot of studies and discussion, and I think really questioning that premise and assumption, and that is why you all are here, to help us think --

MEMBER CASALINO: And, Karen, when you say that, it is a very rational answer, which I have always accepted. But, then, it makes me think we don't want to obscure the differences in care for people from low-income tracts, low-income Census tracts, for example. But how is that different from saying we don't want to obscure for diabetics, when we risk-adjust for diabetics?

And that's where I realize, wait a second, I'm just not thinking clearly about this.

MS. PACE: Right. No, I think that is exactly why this is being called into question.

CO-CHAIR NERENZ: It is a good question.

MS. PACE: Do we really have a strong rationale for saying that those are different? And, you know, I think that's what we are trying to work through. But I think part of it is a logical question. Is there really a difference when we think of it, just kind of going down this list? And you talk about a sociodemographic factor. You could check off these things for all the sociodemographic factors we were talking about.

And so, the question is, what is the uniqueness that we should consider either saying, "No, they shouldn't be in." or, "Yes, they should always be in."?

CO-CHAIR NERENZ: Does anyone have an immediate followup response to these last

couple of points?

(No response.)

I guess I will just have to say for myself, I have been asking the same question Larry just posed and saying I personally don't think there should be a difference, that setting a low bar for low income is not fundamentally different from setting a low bar for diabetes. Now I think politically we might say there are some differences, but, technically, the effects, how the concepts play, I would ask the same question.

MEMBER GROVER: This is in response to that. And that's I'm trying to think about, as we talk in our clinical health systems, one of the things that we hope to do by collecting data, better data about race/ethnicity, is to see how we are doing as individual providers, as health systems, at treating minorities compared to our outcomes on non-minorities.

And I am trying to get my head wrapped around how all that shifts, then, as we report on metrics. If we risk-adjust for race, for SES, those are the indicators we look at, are we going to sit there and look at our numbers and our outcomes on Black patients and White patients, and either say, "Well, now they're risk-adjusted and they look the same" or "Yes, I know that, from a quality standpoint, all Black patients do worse, so I don't need to worry about them." I'm just kind of trying to wrap my head around this.

MEMBER CASALINO: What you are saying, it sounds to me like Black versus White or poor versus rich. It has a valence for us that diabetic versus non-diabetic doesn't. And I think that is part of the reason it is hard to think about, you know.

CO-CHAIR NERENZ: Susannah?

MEMBER SAWHNEY: I agree it is a social -- I'm sorry.

If you are lower SES, you come in

and there are things against you when you walk through the door, and those should be adjusted for. But it is also possible -- possible; it is not in every environment -- that you didn't get the same experience once you got in, too.

MEMBER CASALINO: In this same space, right?

MEMBER SAWHNEY: In this medical system.

So, communication issues, you know, we are talking about a population that is stigmatized and maybe don't speak that well, don't speak English that well or have cognitive impairments, or maybe just bad personal hygiene. You know, are the providers even spending the same amount of time talking to those patients to educate them as they would with me? And then, to add to it, of course, I will then ask them a lot more questions and they will end up spending more time.

But the point is, I mean, is the

system doing -- okay, they're coming in with disadvantages, but is the system doing at least a level amount of effort, if not more, in the face of that disadvantage?

MEMBER BERNHEIM: Right. I mean, I think it is all about the causal pathways, right? I mean, so it is true that I could do a particularly poor job with my diabetics. And so, you could argue, if I have worse outcomes for my diabetics, we would want to not risk-adjust for that.

But, generally speaking, if I am caring for a patient -- I'm trying to use readmissions as an example -- if I am caring for patient who is multi-morbid and they are in my clinic, and I am measuring -- give me a clinic outcome -- my A1c's. That is not a great example. But, then, you have got more of this issue of process measure.

So, mortality, right? Let's use mortality because that is a definite outcome and it is simpler.

(Laughter.)

My diabetic -- and age, let's use age right? Age, they are at higher risk, and there are some small ways in which the care that they walk into and the care that I provide in that setting can better or less meet their needs that affect it. But, with SES, you blow this whole thing wide open. It is much more complicated, and that is why it is different. There is no question they come in sicker, but we can account for that pretty well.

And there is no question in the literature that in lots of settings they are going to poorer-quality providers and receiving poorer-quality care. And I don't want to lose track of that.

Now there is also probably other stuff that is going on, and I may be hurting providers if I don't. So, I'm not saying it is simple, but, to me, it is obvious why diabetes is different than SES.

The diabetes causal pathway to worse outcomes is more biologic and less easily influenced by the quality of care. The causal pathway with SES is totally intertwined with disease severity and quality of care and other factors. And so, it is much harder to figure out what to do within a risk-adjustment model.

If I have accounted for the higher diabetes rates, I have no issue, right? Actually, I know people that don't like this argument, but I actually think we account for a fair amount of SES in these models, but it is just by doing the clinical risk adjustment. I mean, if you put SES in alone, and then you throw all the clinical stuff, I will tell you we have looked at how strong a risk factor it is. It is less of a risk factor than a lot of the clinical diseases and more than some. It is sort of in the middle, once you account -- we have done it for mortality as well. We have done it for I think kidney complications.

I don't know if my team is on the phone. We have looked at other outcomes, too.

I am a little readmission-centric, I admit.

(Laughter.)

I mean, I have a whole different theory about process measures. So, mammography rates, I think about them differently, and we can go down that path, but I'm not going to do that right now. I'm really thinking about outcome measures.

But I do think different outcome measures are different, right? I mean, I think you have to think about the outcome. I think the way it plays out when you have got a patient you have got in the hospital, and you are really having a lot of control of what happens is very different than if you are looking at population base and looking at outpatient. I think it depends on the measure.

MEMBER CALLAHAN: This is Mary

Beth.

I just want to go back for a second to Thu's presentation and whoever that person was that was just talking, which may have been Thu, for all I know. I think in Thu's presentation, she talked about enabling services, which we kind of refer to as ancillary services sometimes, but support services for an individual.

And I would guess that in Thu's situation -- I really don't know -- but there are probably more funds or grants that might be available to you in that situation than a normal primary care physician in another situation. And I don't know; I might be wrong.

But what allows you to provide those enabling services that doesn't allow the normal primary care physician to? And how is that going to be able to pull in the strength of the patient and activate the individual self-management from whatever people come in

your door? Whereas, the primary care physician, because they don't necessarily have those services, won't be able to do. And I just think that is an important factor to think about as well.

MEMBER QUACH: So, this is Thu.

While there are some additional funds to pay for some of the enabling services, it is definitely not enough. For example, at our Health Center we provide 11 Asian languages services and 11 Asian languages. None of the funds that we get can really account for that.

We have just added Burmese and Karen on for some of the emerging immigrant population, not because we reach a threshold number, but because it is the right thing to do as we work towards health equity.

So, while there are some additional funds, you know, it is definitely far from enough. And we aren't getting paid on the enabling services piece.

MEMBER CALLAHAN: You are or you are not?

MEMBER SUGG: I don't know about the rest of you; my head is kind of spinning right now.

(Laughter.)

So, I can't take credit for this analogy. It was from the Medical Director at Harvard who brought this up. So, when you think of things like hand-washing in my clinic, are we going to socioeconomically adjust for that? No. We are not. I mean, there are certain things we are not going to do, because that is really a process thing that is not patient-centered at all.

However, I am held accountable for my pneumovaxes. Okay? Do we socioeconomically account for that? I would say maybe because we still have to look at culturally what is acceptable, and we still have to look at health literacy, which is part of socioeconomic, and how to adjust for that

becomes a little more problematic.

And the other piece of this that is kind of one of those feel-good things that we talk a lot about, patient-centered care. So, if I have talked through my rationale of why you should get your pneumovax and gave you the pros and cons, and you have been on the web and looked at all the stuff that's on the web and say no, I'm still dinged for you no decision because I didn't get my pneumovax rate up.

And so, I kind of feel like the powers that be that make these decisions have to say either we are going to have patient-centered care where the patient can say no, and I don't get dinged for it, or we don't.

And so, those are the other things when I am looking at what we have to put in these variables when we are doing quality measures, is the patient has to be in there in some way. And some of these things really I think we have to do socioeconomic adjustments

for.

The other thing, at some point, I would like to get back to income because I feel like that is an indicator that I have not heard anything that I feel really confident that will really help in my particular situation with my patients. If I look at even Census tract data, where my clinic is located is right across the street from the shelter, which has about 300 people, and right next door to condos that go for about $2.5 million.

So, what would my Census tract data look like and how would that be taken into account? And I know that Seattle is a little different because we have all this sort of Microsoft money that kind of mucks things up, but I think there are other urban places that suffer that same thing. You know, how do you adjust for income without just, frankly, having to ask the patient what their patient is?

I tried to Google our Census

tract. I tried to see zip code and I tried to see Census tract because I was curious what is the income they have in our area.

CO-CHAIR NERENZ: Okay. I have got Alyna and Monica, and then, I know Ninez wants to jump in. We must, with some desperation, do an agenda check shortly.

(Laughter.)

So, let's go Alyna, Monica.

Ninez, are you right on point with something here? Go ahead with that. Then, Alyna, okay.

MEMBER PONCE: So, one thing we haven't considered is looking at stratification measures, like income and equality and residential segregation. So, that is something we could throw in the mix.

MEMBER CHIEN: That was at least a quarter of what I was going to say.

(Laughter.)

But I wanted to go back to NQF's goals because I think in the beginning we were

saying that, if you want a one-size-fits-all and the only solution you want an answer to is do we risk-adjust or not, I think it is going to be like this.

The answer is it depends on what you're using it for. So, I think the focus on finding that answer in the actual variables that you want to put in the model is not the right place to start. You want to decide what people are using it for, and then, you can decide if you want to adjust and make it not transparent, adjust and make it transparent and do it two ways, or do stratification.

So, then, I would like to ask two things. One is it sounded like, when we started talking, that we were talking about risk adjustment and it was kind of a catchall phrase for doing it one way and the other way, and looking at the difference and stratifying. And the way the conversation has evolved, it is sounding very much like it is an on/off switch again.

And then, the other thing is I think I need to know more about NQF's process that it puts people through and how you specify the measures, to see where might be the easy place to insert discussions about what risk-adjustment model to use, what purpose you think the measure is going to be used for, and specifying, "Oh, if you're going to do it this way, and you really want to do it for quality measurement, then we suggest stratify because it does matter if there's a difference." Or you're doing it for spending and you're trying to -- I don't know -- do some capitation. Then, you would want to go --

MS. PACE: I think we will talk with you about some of that offline because, you know, the NQF process, it may too much to get into right here, given our time of the agenda.

But I think, as David has said, it is not just a yes/no, black/white. Part of

what we are going to be doing tomorrow is recommendations about, if so, how; what factors; when; what circumstances; what outcomes; what use, et cetera? So, those are all exactly the questions that we need to work through and make recommendations about. So, we really don't intend it to be a yes/no response.

CO-CHAIR NERENZ: Okay. A quick time observation. We have just passed five o'clock; 5:30 is our at least agenda-scheduled adjournment time. And my inclination and myself is to think of that as a hard stop. There are only so many times the synapses can fire.

(Laughter.)

And there is a dinner reservation. You know, there are reasons to take that seriously. And I have no doubt that, for those people gathering for dinner, these conversations are going to keep running.

We need to check, though, how to

use this last half-hour. The agenda shows a couple of things. There is a public comment period that may actually not take its allotted time.

We were going to at least see if together we could tee-up some possible recommendations or at least the framework for recommendations. And actually, there is a chunk that we have essentially not done that was, essentially, the methods discussion. What about regression-based models versus this stratification, that stratification? I'm dreading the direct-versus-indirect standardization discussion.

But what I am really dreading is even putting a toe in that water after 5:00 in the afternoon because it strikes me as a very important and detailed discussion on its own, and I just don't know that in the time and brain resources available we can do that.

So, a couple of thoughts. One is that in this last block of discussion I don't

think I have heard what I would call just fundamental disagreements or just conflicts that must be resolved before we can move farther. Clearly, there are some somewhat different perspectives, but at least to my ear, we are talking about some cautions, some reminders.

You know, we have had a different sense of how race and ethnicity play in, but, again, our charge is not to say global yes/no on race/ethnicity. I think we have used it as an example of the pros and cons of different things. At least that is how I have been hearing it.

So, as I think about time between now and 5:30, I don't have in my notes here, you know, these are some just burning-hot conflict issues that somehow we have to sort out. I'm sorry if I missed them, but I haven't heard.

As I look at the slide in front of us, this is actually a set of principles,

basically. I don't know that I have heard anybody say that this is wrong or otherwise bad. So, I think there is perhaps in front of us already at least some foundation in writing for moving to a set of recommendations.

So, that said, I don't know that we have a crucial set of things that absolutely must be done in the next 20 to 25 minutes, but I know we must do public comment. And then, at least we need to say something about what are we going to do tomorrow, given what we have done today.

(Laughter.)

Okay. So, Karen, am I --

MS. PACE: No, I think that is fine.

CO-CHAIR NERENZ: So, what do you want us to do?

MS. PACE: Well, why don't we open the lines for public comment and see if we have people that want to add some thought to the conversation?

And then, I think your question about are there any conflicts, we really do want those raised. I agree, I haven't heard any that are like head-on conflicts that we are concerned about.

And then, we can talk about tomorrow.

But, Operator, would you open the lines and see if anyone has any comments?

And I will ask people in the audience. I think maybe the easiest thing is to come up to this microphone here.

THE OPERATOR: At this time, if you have a question or a comment, please press *, then the number 1 on your telephone keypad.

(Pause.)

And there are no comments at this time.

MS. PACE: Okay. So, we will start with -- go ahead and sit down.

(Laughter.)

We had one that came in on the

webinar chat. So, I am going to let Suzanne mention that.

MS. THEBERGE: Sure. This comment came in earlier this afternoon from David Keller.

"I also would say that it would be hard to argue to practitioners and communities that SES doesn't make a difference. Not risk-adjusting will make it hard to sell in the community."

MS. PACE: Okay. All right.

Yes? And please tell us your name and who you are with.

MR. DEMEHIN: Thank you and good afternoon, everyone.

My name is Akin Demehin. I'm a Senior Associate Director with the American Hospital Association.

And first, I just want to add my commendation to this Committee for really bravely tackling what is an incredibly-complex issue. I definitely feel like I have learned

a tremendous amount from the discussion, and that everyone, regardless of your viewpoint on the issue, has brought so much perspective and such thoughtful perspective.

That being said, I am very glad it is you at the table rather than me.

(Laughter.)

I wanted to reflect a little bit on one of the discussion points that the Committee had earlier. And really, the central question was, if NQF should allow for the inclusion of sociodemographic variables in measures, how strong a recommendation should it be? Should it be "We'll allow for it, but you don't have to," or should it be, "We expect you to assess for sociodemographic variables as part of the endorsement process and demonstrate whether an adjustment is needed or not," and then, to apply that adjustment if it is warranted?

From the perspective of the AHA, we would really favor a fairly-strong

recommendation from this Committee, at least based on the conversation we have heard so far. The notion of really expecting that outcome measures, when they come to NQF for endorsement, have been assessed for the impact of sociodemographic variables on the performance results, do we expect that every measure will necessarily require a socioeconomic adjustment? No. And I think there are several very good examples that many of you have articulated today that demonstrate that.

But we think that including a fairly-strong recommendation in this area could really be a great opportunity to strengthen the value of NQF endorsement in a couple of ways.

We think -- and I think a couple of folks alluded to this earlier -- that the way an outcome measure portrays performance based on SES has a direct bearing on its validity as an outcome measure, and we think

it needs to be understood before it can be considered a national standard.

And the other reason is we think that it really acknowledges the reality, and several of you also alluded to this, that NQF outcome measures become publicly reported. They become tied to payment, and they have the ability to move substantial dollars around in the healthcare system. And our members are incredibly concerned that, if those dollars are allocated based on performance measurement, that they be done so in a fair way.

And then, as a final comment, I absolutely agree, particularly with the measure developers in the room, that there need to be some boundaries, some very clear and consistent criteria for what is expected when measures are submitted into the NQF endorsement process, what kinds of analyses, what kinds of factors. We absolutely agree that we shouldn't create something that is

overly subject to interpretation, overly burdensome, et cetera.

So, looking forward to a continued robust discussion tomorrow, and thank you very much.

MS. CHAMBERS: Hi. I'm Jayne R. Chambers. I'm a Senior Vice President for Quality at the Federation of American Hospitals.

And I, too, want to thank you for your robust discussion today. It has been quite educational and really wonderful to see people bring so much variety to the table and to have such a civil discussion about a topic that we have been talking about at length for a number of years. So, thank you very much for that.

I should probably just say "ditto" to everything that Akin just said, but the Federation members have long thought that measures, when we're looking at them for outcome purposes and for accountability

purposes in that context, should be adjusted for sociodemographic information and data. And the question that you tackle tomorrow, which is how to do that, is at the core of all of that.

But I appreciate very much that, from the discussion and what we have heard today, that at least having the discussion about how to do that and opening the door to doing that has been very important, and we would encourage you to continue down that road.

And I also agree that the developers need to have as much clarity as possible when they are bringing forth their measures in how they should, what they should be presenting, what should be tested, and how they should be looking at it. So, I appreciate that as well.

Thank you.

MR. SHAW: John Shaw from Next Wave in Albany.

And I also want to give kudos to the whole group. I came down for another round of fireworks and find that I did not miss them at all.

(Laughter.)

One of the things that may make the discussions tomorrow easier when we try to say, do you risk-adjust it or not, it depends. What does it depend on?

And I am not an MD, but I do know that, when I am speaking to MDs, they want to know the mechanism or the causal pathway of what is really impacting on this. And if we step back and look at things from the whole system and model the whole system, as a country, we are spending more money and we're getting worse outcomes.

In recent years, people have been modeling where and why and, basically, focused on the dual-eligible population, the folks with multiple chronic conditions, and people with behavioral or substance abuse disorders.

In that population, guess what? We have poorer outcomes and higher costs.

So, that is where a lot of the anxiety is on the part of the providers. That is where a lot of the priority attention probably should be if we are trying to move the cost curve and really implement all of the Triple Aim.

With that in mind, keep in mind that, if the mechanisms are different, and here what is driving that population is not what most of the healthcare people are providing; it is what is happening after you provide that. It is what are the mechanisms for engaging the patient, engaging their informal caregivers, engaging the community, and paying for however much of that, and not pretending that it is all for free.

In long-term care supports and services, where a lot of the dual-eligible impact is, we are trying to push everything into home and community services in an

informal caregiving environment where, with an aging population, we have got fewer and fewer people, period, able to provide that or willing to provide that or able to provide that.

So, can we really start looking at in the sociodemographic measures those items that really get at what makes it effective once the person leaves the hospital, leaves the clinic, and so on? What has worked? And we have heard a number of examples of what happens if the local taxing district provides resources to do that. If grants provide that, fine. If those are not available, then maybe we should invest some of the healthcare dollars outside of the building and into the community.

And IRS has apparently really pushed that and gotten the ball rolling quite a bit.

Thank you.

MR. SIGNER: Good afternoon.

I'm Bill Signer. I'm here on behalf of Health First, which is an MA plan in New York. It is one that focuses on low-income folks. We have about 109,000 members. About 55 percent of them are dual-eligibles. That is where we market. Everybody else is below 200 percent of the poverty line.

I have listened today, and other than the tools comments, I think everybody has focused on hospitals. MA plans are being affected by quality measures. Quality measures do have an impact on payment.

And I think that a lot of the plans that are focused on low-income folks are very concerned that, especially if you are in an urban area, not that we should get more money, more money should be directed to us, but we should be able to get as much as the fee-for-service system is. And the STAR Bonus Program is designed to help with that.

So, our concern is, and what we are looking at is that plans that have 50

percent or more dual-eligibles, and we have 80-percent low-income subsidy folks, are seven times less likely -- less likely -- to score four stars, which is what you need to score.

Now our providers are pretty good, but the problem is that the clinics that we go to are overcrowded. The demand for services exceeds the supply. And our folks are upset, frankly, when they don't get seen. Or we have language barriers. All the criteria that you have talked about, we are seeing.

And our improvement scores are good, but because there isn't a socioeconomic status adjustment and some recognition for plans that are focused on this large population of dual-eligibles, we are going to lose funds. And what we are concerned is that our members who get extra benefits, like they can't pay for over-the-counter drugs; we pay for that for them. There was discussion of transportation. We pay for that for them, and many other services we provide to them. If we

can't give that to them, what they are going to end up doing is falling out of the system and going to the fee-for-service and ending up costing more to the system. So, we are concerned that there need to be some adjustments here, so it recognizes plans like ours and what we do.

The one question I had to the group was there was a discussion about collecting both raw data and adjusted data, and which we should do. We view the collection of raw data as being very, very important because it does give us guideposts. It helps us understand where we need to improve and how we should improve.

What I am not quite sure because I am not a statistician is why you can't collect the raw data and, then, adjust it afterwards. So, it would one collection, but, then, you would adjust it. So, it would seem to me you would get the best of both worlds. You would know whether we are doing well or not, and you

would also adjust it.

And also, I would wonder from looking at some of the charts we had here, if you do the adjustments and, then, stratify, which is the other thing I think is very, very important, comparing like plans in our case to like plans for like hospitals to like hospitals, then you will find out who are your good providers and who aren't. Because you will see within that category who is above and who is below the line. That seems to answer the question of getting rid of the bad, not rewarding the bad actors.

Thank you.

MS. PACE: And, Operator, would you check one more time if there are any comments on the phone?

THE OPERATOR: If you have a comment, please press *1.

(Pause.)

MS. PACE: Okay.

THE OPERATOR: And there are no

comments at this time.

MS. PACE: There is or isn't?

THE OPERATOR: There is not any comments at this time.

MS. PACE: Okay. Thank you.

MEMBER LIPSTEIN: Would you read again the comment from the webinar?

MS. THEBERGE: Sure. Just give me one moment to pull that up.

"I also would say that it would be hard to argue to practitioners and communities that SES doesn't make a difference. Not risk-adjusting will make it hard to sell in the community."

MEMBER LIPSTEIN: The reason I thought that was important is there is a second bullet that says that "A usual consideration for selecting a risk factor is an empirical association with the outcome of interest."

Sometimes if there is not an empirical association, risk-adjusting will

help facilitate buy-in of the provider community. It is a point that came up earlier, and I thought that that's what the webinar commenter was speaking to.

Because I think for all of us who have done this Six Sigma stuff, we know that the effectiveness of the solution equals the quality of the solution plus the acceptance of the solution. And risk-adjustment, even if there isn't an empirical association, may facilitate acceptance.

CO-CHAIR NERENZ: A couple of quick things. I know, Dionne, you have had your card up, Susannah, and we are really closing in now on 5:30.

MEMBER JIMENEZ: What I was trying to say is, because I know Kate is only here this afternoon, I wanted to kind of tag along to Alyna's point about it would be really helpful to have more information from CMS to know sort of what is the process that happens. Because it seems like a lot of our issues and

concerns are really around the implementation, and I know NQF has a set role around defining criteria for measure selection. But it would be also helpful to get information from CMS about what happens afterwards. Because we know there are adjustments that happen when you are actually designing like the Value-Based Purchasing Program, for example.

And one example, I could point to that is, when you are looking at the patient experience-of-care domain, looking at the HCAHPS Survey, you know, they combine, for example, cleanliness and quietness of the environment.

And so, it just would be helpful to know sort of like more about the rationale and that processing, and how it can interplay with this.

CO-CHAIR NERENZ: I am wondering if maybe at this point --

MS. GOODRICH: I don't understand what your question is.

MEMBER JIMENEZ: To provide more information sort of about the process of how the adjustments that are made on the implementation side happen, you know, outside of sort of just selecting NQF-endorsed measures, like what happens at the CMS level. So, it doesn't have to be now, but it could be at a later time.

MS. GOODRICH: It might be helpful. We had talked about doing this, but we weren't able to make the logistics work. I actually have some of the people at CMS who actually handle the payment policy side, which, unfortunately, is not my shop, talk a little bit more about exactly that. I do think that would be helpful information. I know that we have provided to NQF some sort of fact sheets and that sort of thing about that kind of thing.

I mean, essentially, just sort of in a nutshell, we do work closely with the payment folks in helping to define the policy.

So, it is really isn't just about the measures. It is sort of at the same time that we are deciding what the measures are, we are also trying to decide what the supporting methodologies should be, and we work in partnership with our colleagues at CMS who do that work as well.

But, obviously, that is extremely high-level. There is a lot more detailed work that goes into that, usually doing quite a bit of analysis using our data of the different scoring methodologies and how that would be impacted sort of across the spectrum. So, there is quite a bit of data analytics that goes into those decisions.

CO-CHAIR NERENZ: Okay. I am wondering, we may need to turn to Karen and Helen a bit, and just tell us what do you want us to think about overnight that might be clarified and facilitated by a glass of wine or two.

(Laughter.)

MS. PACE: Okay. Well, the last question that we didn't get to was, not that we have answered any of these questions, but we have certainly been airing the issues.

(Laughter.)

But the next question is, you know, if we are going to do it, and if we have identified the right factors, what is the approach we should take? Should it be a statistical risk model? Should it be leaving the clinical things in the statistical risk model and, then, stratifying, stratifying within a provider or stratifying as in the example of MedPAC's recommendation of stratifying by some socioeconomic factor to identify like peer groups for purposes of comparison, whether it is comparison for pay for performance or comparison for how you are doing against your peer group.

So, that is what our next set of question was about, and we really are going to have to, I guess, maybe start off with that in

the morning, to at least have some discussion of the issues around that.

What we were hoping to do tomorrow, then, is to start kind of working through these and have a strawman set of recommendations that you all would list, and then, break into smaller groups to really kind of discuss some of those recommendations in more detail.

So, if that sounds okay to start off that way, we will ask you to dream about that tonight. And also, if you can come up with the answer tonight in your restful sleep, then we would love to hear that as well.

CO-CHAIR NERENZ: Okay. So, you can send an email at 2:00 a.m. if you just can't sleep thinking about it.

MS. PACE: Right, right.

(Laughter.)

But I want to thank everyone for the great discussion. It has been very stimulating, lots of issues raised, and we

knew that there would be. We know there aren't any easy answers, but, hopefully, tomorrow we can start finding a path to something that will make sense.

(Whereupon, at 5:27 p.m., the meeting was adjourned.)

G

H

M

N

P

X

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Sociodemographic Factors
Expert Panel Meeting

Before: NQF

Date: 01-15-14

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.

Neal R Gross

Court Reporter