NATIONAL QUALITY FORUM
+ + + + +
HEALTHCARE DISPARITIES AND CULTURAL COMPETENCY CONSENSUS STANDARDS
+ + + + +
MEETING OF THE STEERING COMMITTEE
+ + + + +

MONDAY, JULY 11, 2011

+ + + + +

The Steering Committee met at the Liaison Hotel, 415 New Jersey Avenue, NW, Washington, D.C., at 8:30 a.m., Dennis Andrulis and Denice Cora-Bramble, Co-Chairs, presiding.

PRESENT:

DENNIS ANDRULIS, PhD, Co-Chair
DENICE CORA-BRAMBLE, MD, MBA, Co-Chair

EVELYN CALVILLO, DNSc, RN, California State University (via telephone)
MARSHALL CHIN, MD, MPH, FACP, University of Chicago
LUTHER CLARK, MD, Merck & Co., Inc.
LOURDES CUELLAR, MS, RPh, FASHP, TIRR-Memorial Hermann

COLETTE EDWARDS, MD, MBA, CIGNA HealthCare
LEONARD EPSTEIN, MSW, Health Resources and Services Administration
DAWN FITZGERALD, MBA, Qsource (via telephone)
ROMANA HASNAIN-WYNIA, PhD, Northwestern University Feinberg School of Medicine
EDWARD HAVRANEK, MD, Denver Health Medical Center
ELIZABETH JACOBS, MD, MAPP, University of Wisconsin

FRANCIS LU, MD, University of California, Davis
MARY MARYLAND, PhD, MSN, BC, APN, Chicago State University
WILLIAM McDADE, MD, PhD, University of Chicago
ERNEST MOY, MD, MPH, Agency for Healthcare Research and Quality
MARCELLA NUNEZ-SMITH,, MD, MHS, Yale New Haven Health System
SEAN O'BRIEN, PhD, Duke University Medical Center
NORMAN OTSUKA, MSc, MD, FRCSC, FAAP, FACS, New York University Hospital for Joint Diseases
GRACE TING, MHA, CHIE, WellPoint
DONNA WASHINGTON, MD, MPH, VA Greater Los Angeles Healthcare System
ELLEN WU, MPH, California Pan-Ethnic Health Network
MARA YOUDELMAN, JD, LLM, National Health Law Program

NQF STAFF PRESENT:
HELEN BURSTIN
HEIDI BOSSLEY

KRISTIN CHANDLER
NICOLE McELVEEN
ELISA MUNTHALI
ROBYN NISHMIMI

ALSO PRESENT:

JOEL WEISMANN, PhD, Disparities Solutions Center, Massachusetts General Hospital
JOSEPH R. BETANCOURT, MD, MPH, Disparities Solutions Center

I-N-D-E-X

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

(8:54 a.m.)

MS. MCELVEEN: Good morning. Good morning. So now that we're all positioned a little more comfortably we can get started. And I apologize for the movement and the late start of the meeting.

My name's Nicole McElveen. I'm a senior project manager with the National Quality Forum and also heading up this project with a few other staff members from the National Quality Forum.

Good morning and welcome to the Healthcare Disparities and Cultural Competency Consensus Standards project. I want to make sure that everyone is here for this meeting because we do have also another NQF meeting going on across the hall for our Measure Applications Partnership.

We are going to jump right into the agenda and get started. A few logistics, as you all can see we have breakfast and

drinks in the back of the room. As you speak, please be sure to use the mic. The meeting is being recorded and the only way that we can record is for the mic to be on while you're talking. If you do have questions throughout the meeting, our normal protocol is for you to place your name card upright, allowing the co-chairs and other folks to be aware that you do have a question, if you do, during the meeting.

And I think that's it for logistics. I'm going to -- oh, let me just tell you who else -- we're here with other staff. Behind us on the table we have Elisa Munthali, who you all may have been communicating with, and also Kristin Chandler is the project analyst. And I will let the co-chairs introduce themselves.

CO-CHAIR ANDRULIS: Good morning all. It's good to be back at an NQF meeting once again. For me, this is the second time around. I also think somebody thought it was

a good idea to pair Dennis and Denice together. This is a first for me so you'll have to be a little bit precise if you're addressing us by our first names.

I just wanted to mention one or two things from the conference call that to keep in mind as we go through the couple of days here. This meeting builds on a significant body of work that's been done by NQF before. There are a number of reports that you have appendixed to the works that we are looking at today.

So there is an opportunity to kind of be added on, added on and integrated to existing work as well as the hats off to the folks who wrote the paper for this meeting, Joel and Joe and others who wrote that comprehensive and very thoughtful piece, give us a lot of fertile ground to build on. So I want to thank them for that great work.

And so as we go through between the work that's been done before by NQF and

others and as well as the paper, but keep that in mind as we have an opportunity here to not only build on what they have done and take off on variations on a theme, but set directions.

Agreement, consensus will be part of this process, and something I remember from that phone call is whether we're asking the right questions being a good part of what we need to consider here, as well as the importance and feasibility of the points that we'll be pursuing.

So with that I hand it over to my colleague and say welcome.

CO-CHAIR CORA-BRAMBLE: So good morning to all. Dennis and I are going to do a bit of a team tag here. This was probably the most difficult day for me to accommodate in my schedule but I did not want to say no to the, sort of the honor to be invited to be a co-chair.

I'm actually in two different meetings and I'm on call, starting at 9:00

a.m. so you will have to -- I will have to apologize because I will be stepping in and out. Dennis is going to do the heavy lifting today and I'll have a little bit more leeway tomorrow. So with that, I'm going to take my 9:00 a.m. conference call. It takes five minutes, and I'll be right back.

CO-CHAIR ANDRULIS: I won't screw it up.

MS. MCELVEEN: Great. A few other things I wanted to mention, I just want to make sure that everyone received an electronic copy of the materials that were emailed out to the group. We have also printed a copy of those same materials and placed them at your stations.

There was a small memo that was accidentally left out. We have printed and included a copy of that. It's about a five-page document that was also placed at your station. This memo is really a more detailed copy of the agenda.

The agenda had a few questions teed up for each section that we'll be discussing. The memo had a little more detail to it. So as we go through the sections of the paper and the different discussions, you may want to just reference the memo as needed for that additional detail.

And so we're going to then move into our disclosure of interests, and -- or maybe introductions and then?

MS. BOSSLEY: Often, we've done them at the same time.

MS. MCELVEEN: Yes.

MS. BOSSLEY: So I think it flows together. So let me just give you a little background on what we're about to ask you to do. As you may remember, when you submitted your materials you filled out a disclosure of interest form. We're asking you to orally disclose anything that may be relevant to the work of this committee today. So we don't need a full list of who you're a member of,

any of that, but any grants, any speaking, anything that may be relevant to what will be discussed, we're asking you to just state that to everyone.

The other piece that I would just remind everyone is you are here as individuals, not representing the organization you work for or who nominated you, something we like to remind everyone. But let me stop and see -- I'm seeing maybe some questions. Does this make sense, what we're about to ask you to do? Okay.

So just, I would do introductions and then just go around and say if you have anything to disclose. If you don't, feel free to say, "I don't have anything to disclose." Romana?

DR. HASNAIN-WYNIA: Does that include what we've already listed or if we've already listed it then can we just --

MS. BOSSLEY: We're asking you to repeat.

DR. HASNAIN-WYNIA: Oh, you are.

MS. BOSSLEY: Yes.

DR. HASNAIN-WYNIA: Okay.

MS. BOSSLEY: So anything you have, just briefly give a high-level what it is and nothing more than that, just who it's from and what the focus is, let's say, if it's a grant. That's it. So not to put the chair on the spot, but would you like to be the person to start?

CO-CHAIR ANDRULIS: Yes. I'm Dennis Andrulis. I have nothing to disclose. And I am Dennis Andrulis.

MS. BOSSLEY: You are Dennis Andrulis.

CO-CHAIR ANDRULIS: I did. I said -- it's easily forgotten. Dennis Andrulis and I have nothing to disclose.

MS. TING: Okay. Grace Ting. I work for WellPoint. I'm not representing them today. And I am currently receiving a grant and I'm the lead PI for a Robert Wood Johnson

Foundation grant.

DR. CLARK: I'm Luther Clark. I'm the Global Director for Scientific Affairs at Merck Pharmaceuticals, located in, well, a number of different places, Whitehouse Station, Rahway and West Point, New York. I'm a cardiologist by training. I've been at Merck for approximately four years.

Prior to that, I was the chief of cardiology at Downstate Medical Center in Brooklyn and Director of the Brooklyn Health Disparities Center. Other than being an employee at Merck, I have nothing else to disclose.

DR. JACOBS: Hi, I'm Liz Jacobs. I'm at the University of Wisconsin, Madison School of Medicine and Public Health. And I'm not sure this is a conflict, but I'll mention it anyways. I'm on Aetna's Racial Ethnic Disparities Advisory Task Force, and I receive an honorarium for being part of that.

DR. HASNAIN-WYNIA: Good morning.

I'm Romana Hasnain-Wynia, and I'm with Northwestern University in Chicago. And I don't believe I have any conflict of interest or anything to disclose. I do serve on an AHRQ Technical Advisory Panel on creating the evidence base for disparities and quality, the quality gap. And I do have a consulting with the Henry Ford Health System. So I'll leave it at that.

DR. NUNEZ-SMITH: Good morning. I'm Marcella Nunez-Smith. I'm faculty at the Yale School of Medicine and affiliated with Yale New Haven Hospital. I am also PI on an NIH grant to develop a measure of health care discrimination. Other than that I have nothing to disclose.

DR. OTSUKA: Good morning. I'm Norman Otsuka. I'm a pediatric orthopedic surgeon from the NYU Hospital for Joint Diseases. I'm on the Diversity Advisory Board of the American Academy of Orthopedic Surgeons. I've taught some courses on

communication and diversity. I have no other disclosures. Thank you.

MS. YOUDELMAN: Good morning. I'm Mara Youdelman. I guess I should disclose I'm a lawyer -- not a conflict but other issues involved in that. I'm at the National Health Law Program. Again, I don't think this is a conflict but just in case, I'm the chair of the Certification Commission for Health Care Interpreters, which has developed competency standards for assessing interpreters in health care settings.

DR. WASHINGTON: Good morning. I'm Donna Washington. I'm an internist at the Greater Los Angeles Healthcare System and UCLA. And I receive funding from the Department of Veterans Affairs to conduct health disparities research on determinants of health care disparities within the VA Health Care system. I'm also on the steering committee for the VA Principal Deputy Undersecretary for Health Group that's

addressing health disparities.

DR. HAVRANEK: I'm Ed Havranek. I'm a cardiologist at the City Hospital in Denver and on the faculty at the University of Colorado Medical School. I have -- I'm PI on one grant and a co-investigator on another grant that looks at the effect of bias and discrimination on health care outcomes.

DR. O'BRIEN: Good morning. I'm Sean O'Brien. I'm a biostatistician at Duke University Medical Center. I've been involved with a couple previous NQF projects including a measure testing task force and a couple of hospital outcomes projects.

The group that I'm with at Duke works with several medical professional societies with their databases and so frequently I'm involved with submissions to NQF and so in past activities I've mentioned my involvement with the Society of Thoracic Surgeons database as a conflict.

MS. WU: Good morning. I'm Ellen

Wu. I'm with the California Pan-Ethnic Health Network, and we receive a grant from Kaiser Permanente. I think that's probably the big thing to disclose.

MS. CUELLAR: Good morning, I'm Lourdes Cuellar from TIRR-Memorial Herman -- thank you. Last year I did receive an honorarium. I often speak on cultural barriers to clinical research and last year I did receive an honorarium.

DR. CHIN: I'm Marshall Chin. I'm a general internist in health services research at the University of Chicago. On the conflict of interest statement, it was pretty lenient in terms of I get low bar, so I'll just go over the low bar.

Research grants from NIH, AHRQ, Merck Company Foundation, Hartford Foundation, Commonwealth Fund, American Diabetes Association, Committees of the Midwest for Clinicians Network, Institute of Medicine, CMS, Robert Wood Johnson Foundation, ABIM,

AEEMC, Asian Association -- Pacific Islander Health Organization, National Association of Health Centers, VA, Indian Health Service and American College of Cardiology, NCQA and a number of general memberships that I don't think are problematic.

DR. MARYLAND: Good morning. I'm Mary Maryland. I'm faculty member at Chicago State University and a nurse practitioner with a private company. I have no disclosures.

DR. EDWARDS: I'm Colette Edwards. I'm a gastroenterologist. Until very recently I was with Cigna Healthcare as their national medical director for health disparities, and I have nothing to disclose.

DR. LU: I'm Francis Lu. I'll disclose I'm a psychiatrist from UC Davis, and I have nothing to disclose.

MR. EPSTEIN: I am Len Epstein. I'm senior advisor for Clinical Quality and Culture at HRSA, Health Resources and Services Administration, and I don't have anything to

disclose.

DR. CORA-BRAMBLE: I'm Denice Cora-Bramble. I am a professor of pediatrics at George Washington and I'm a senior vice president at Children's National Medical Center. In terms of issues to disclose I was -- I've been a chair of an advisory board for Pfizer. I am a fellow of the Academy of Pediatrics and also a consultant. I am a member of the DC Physician Advisory Subcommittee for Unison/United Health Care, treasurer and board member of the Academic Pediatric Association.

MS. BOSSLEY: Okay. Is there any questions that you have for your colleagues? Anything you wanted to discuss? Typically it's no. I would say the reasons that you listed are exactly why we've asked you to be here, so this is a good thing, but we do like to run through this in case there's any concerns. I'm going to leave and go to the next meeting so please don't be offended as I

walk out. Thank you.

DR. BURSTIN: I'll just add my welcome. I'm Helen Burstin. I don't usually sound like this. I do live with small viral vectors known as children. I'm the Senior Vice President for Performance Measures at NQF. Great to see so many familiar faces, and thank you all for being here.

We really view this as being an incredibly important piece of what NQF can do, and I think we really want to build on the work that Dennis and others have done on our previous Disparities Committee to really think through what we should do prospectively as a measure comes to us, to think through how we handle it, should it be something we should consider, always being stratified as an example as you talked about in the report.

So we really look forward to your guidance. We really do see this as something we will take from this committee and just build into all of our committees going

forward. So, thank you.

MS. MCELVEEN: I am going to do a brief introduction and recap. Some of this information you did receive during the orientation call, and for those of you who weren't able to call in, it will be new for you. You don't have a copy of the slides, but they are projected right behind you.

For folks who have laptops, not too many in the room, but we do have Internet access if you do have a laptop. We have a few cards that have the user name and password and we'll pass those around.

DR. BURSTIN: User name is laley, L-A-L-E-Y, and the password is 00800, and it works.

MS. MCELVEEN: And we have introduced ourselves. This is the main project staff.

MS. NISHIMI: Hi, I'm Robyn Nishimi. I was the founding Chief Operating Officer for NQF and now I'm a consultant

helping NQF with this project. Pleased to have you on board.

MS. MCELVEEN: So again, the main purpose of the project, we are looking to establish recommendations for approaching measurement of health care disparities across settings and populations, and to also then endorse a set of consensus standards for measuring and publicly reporting on disparities and cultural competency.

And this particular project, we do have it broken up into two specific phases. The first phase, which we're working on today, is focusing on reviewing a very comprehensive commission paper which will help us begin to establish that approach for measuring health care disparities.

And the second phase consists of our traditional consensus development process in which we will review and look at a set of consensus standards to hopefully provide recommendations for endorsement.

Our timeline across the two phases, our meeting obviously is today, July 11th and 12th. Following this meeting and the recommendations we receive from the committee, the commissioned paper will go out for what we call public and member comment. That's a time that we allow our members and the public to provide specific comments on the content of the paper. And we're looking to finalize that paper and to finalize the recommendations of the committee a little bit later this year in September.

Concurrently happening will be a call for measures. We're hoping to launch the call for measures as soon as possible. So August is the start date for that call for measures to be launched, allowing measure developers to submit measures for consideration.

Following that call for measures, which will end in October, the NQF staff does a preliminary screening of those consensus

standards before they're presented to the committee. So we're looking at a second in-person meeting where the committee will review those consensus standards and provide recommendations for endorsement in December, early January.

So we don't have an exact date yet, but as we progress throughout the project in the coming month or so, we will have a better sense of a more exact date because I know that's a busy time with the holidays.

Finally, following that second in-person meeting, we do continue on with our traditional process of consensus where we do have the standards posted for comment, as well as a voting period. And looking to wrap up an endorsement in June of 2012.

So quickly, I know you all are probably familiar with this, but just wanted to quickly recap some of our historical work that NQF has done on disparities to date. Began about 10 years ago in 2001 when we had

a small workshop with some key experts who really helped NQF begin to think about health care quality for minority patients.

And out of that workshop were ten specific recommendations that were used to engage stakeholders to begin to think about disparities and health care quality for minorities.

Following that work, we have our Disparities-Sensitive Measures project for the ambulatory care setting. During this project we did endorse 35 disparity sensitive measures appropriate for the clinician level. You all do have a full copy of this report in your materials.

And finally, most recently, we did a project on cultural competency where we endorsed a framework and 45 preferred practices around cultural competent care. I've provided the executive summary of that report with your materials. We are, of course, happy to provide the full

comprehensive report for that as well.

I'm going to pass over this. This is an outline of the commission paper, which we'll be getting an overview of that paper briefly.

So the main meeting goals today are: first, we want to talk about recommendations for identifying disparity sensitive measures. We also want to then review, once we have those recommendations, on applying that to our current NQF portfolio of measures.

We also want to seek recommendations, again, on methodological considerations for measuring disparities. And finally, an approach on how NQF can really look at measuring disparities prospectively.

So with that, I am -- oh, I just remembered, I did not give my audience and team from Massachusetts General Hospital a chance to introduce themselves. And before you do, I'd just like to allow the folks on

the phone, if there's anyone on the phone who has called in, either as a committee member or as an audience member, I'd like to give them a chance to introduce themselves.

MS. CALVILLO: Can you hear me? Evelyn Calvillo.

MS. MCELVEEN: Okay.

DR. WEISSMAN: Do you want me to repeat it? Joel Weissman, health policy researcher at the Mongan Institute for Health Policy.

DR. BETANCOURT: And Joe Betancourt, Director of the Disparities Solution Center at Mass General Hospital.

DR. CLARK: What's the planned distribution of the white paper once it's completed?

MS. MCELVEEN: We don't have a specific distribution plan. I mean, it's something that will remain intact with the project, obviously. It will be available for the public and the members for viewing. Did

you have a specific question around the distribution of the paper?

DR. CLARK: It was more in the context of how available and what kind of awareness would be made that this exists?

MS. NISHIMI: As part of our process, the paper will be posted and available to NQF members and the public on our website. And the authors are free to publish it in what form they, you know, may wish to submit it to a journal as well.

MS. MCELVEEN: Yes?

DR. CHIN: Could we go back to the meeting goals slide, please?

MS. MCELVEEN: Sure.

DR. CHIN: So my question has to do with sort of both this meeting as well as more generally, this charge to the committee, when you look at the four different bullets there, the first three generally have to do with measurement and the fourth is broader in terms of an approach for addressing

disparities.

One of the very nice things about Joel and Joe's paper was it talked about those issues but then talked a lot then about implementation in terms of, you know, not so much what incentives you might build into the system to reduce disparities as well as issues regarding trying to avoid unintended consequences. And in many ways those latter two parts are at least if not more important than sort of the measurement issues.

And so I'm wondering to what extent are we supposed to concentrate on just, like, just the good measurement issues and the list of disparity measures versus that second and third parts that Joel and Joe also talk about, in terms of incentivizing? And then design a system so that you avoid bad things happening?

MS. MCELVEEN: That's a good question. We have tried our best to sort of lay out on the agenda an appropriate timeline,

timeframe for achieving all of these goals. But recognizing that some kind of rise to the top more than others.

So the main focus is first to really look at the guidelines for identifying disparity sensitive measures and to really concentrate, again, on the methodological considerations around disparities measurement.

We do also have a good portion of our day tomorrow that will allow us to talk about some of the other considerations that you had mentioned and getting a lot of key recommendations on more of the broader issues for disparities measurement. Did you have a comment as well?

DR. BURSTIN: Just one thing to add to that. I think that that fourth one, again, is still, as I read it, still kind of an insider NQF question for you, in some ways. Obviously it's a bigger question than that, but I think in some ways we'd like to have you help us understand as a measure comes to NQF

in any project, cardiovascular, surgery, whatever the case may be, what kind of information should they have to submit on the submission form, for example, about evidence of disparities? What should be submitted by the three-year mark when the measure's been in use to justify it?

It's those kinds of issues where you don't want to have to go back retrospectively, as you're going to help us think through tomorrow, and say this is disparity sensitive, this is disparity sensitive, but instead, think through prospectively how we do this.

So just as an example, the Child Health Quality Committee this past year looked at a measure of low birth weight. And the committee immediately went, ooh, this is a measure really hard to look at in the aggregate, it was a population health level measure, said this is one that should always be stratified.

And it was one of those things, again, made us think that we needed to have some logical process where a committee looks at something and says: this is really high priority. This is a measure that should never be looked at unless you have the strata to be able to look for disparity. So that's kind of what we're thinking, if that helps.

MS. MCELVEEN: Okay. Any more questions? So we're going to move into a presentation from Mass General on the paper.

DR. WEISSMAN: Great. Oh, do we have a clicker for the -- to advance the slides?

MS. MCELVEEN: You can cue me to advance them. I'm the clicker.

DR. WEISSMAN: You're the clicker. All right. Thank you very much. We're very honored to be here today and before I begin I'd like to just acknowledge our co-authors. This was really a team effort of many of the staff and friends of the Disparities Solutions

Center, including Alex Green, Gregg Meyer, Aswita Tan-McGrory, who's in the room with us, Jake Nudel, Jessica Zeidman, and Emilio Carrillo. Next slide?

The purpose of this report, as we understood it, was to provide guidance to the NQF steering committee charged with selection and evaluation of disparity sensitive measures. And it really is all about providing guidance from what we learned from the literature and from our experience in working together on a number of disparity sensitive issues.

The expertise in this room obviously is very impressive and so we don't mean to in any way say that this is the way something has to be done. It's really to raise issues and hopefully to foster some discussion going forward.

The second purpose is to describe methodological approaches to disparities measurement, which we spent a fair amount of

time, and to identify some cross-cutting measurement gaps in disparities.

I also want to just make a couple comments about some terms. We talk there are -- people talk about health disparities which really focus on health status and outcomes and raise a lot of issues about the accountability of the provider.

This report is more focused on health care disparities. We rely on the IOM definition of racial-ethnic disparities in quality that are not due to access-related factors, clinical need, preferences or appropriateness. So it's really more about the provision of care rather than thinking about disparities in general, at least from the perspective of this report.

I think, though, in terms of those bullets that Marshall mentioned earlier, it may be that you want to extend that to reduction of health disparities generally.

Also, early on in the discussions

that we had about beginning this report, we decided to focus on racial, ethnic and language proficiency disparities. We did that not because other disparities with women, the disabled, gay and lesbian populations, there are a number of other populations that exist that have and experience disparities, but we felt like the evidence on racial, ethnic and language disparities were a little bit more developed.

The efforts to develop measures on other populations are in the developmental phase, and we thought that there was plenty to write just on those populations, and so we decided to limit our comments to those. Next slide.

The outline of the report and what we're going to talk about today is we're going to do -- we're going to talk about background on racial/ethnic disparities and data collection. Joe -- we'll do kind of a tag team approach -- Joe will do numbers one and

two.

Then I'll take over and talk about disparities, measures and indicators and some methodological approaches and then hand it back to Joe to wrap up with quality improvement and then public reporting and some of our policy recommendations. Next slide?

DR. BETANCOURT: Great. Well, I want to thank everybody so much for giving us this opportunity to put this together. Hopefully, it was some good bedside reading for you all at some point. It was, I think, went into a lot of different places and again, as Joel mentioned, food for thought.

It's humbling to get a chance to be with you all. It's rare that I think we all get a chance to be together and tackle these issues, so I thank you for your time and look forward to sharing some of our thoughts.

Background, quite simply, in the paper we tried to make the case for why this is important, the impact of disparities on

cost, quality and safety, and then really highlight the fact that when we think about measurement, we clearly see that our ability to really measure disparities is going to be the foundation for any interventions to address them.

And so we talk about the fact that even on the aspect of data collection, race/ethnicity data collection, we have a long way to go. And then taking it to that next step of measurement is what we're looking at, but certainly the key part of the foundation for disparities development and intervention development. Next slide, please?

We tried to put up front in the paper as well a bit about data collection. In this room we have some of the top talents in the people who have been working on this since Day One, particularly Romana and others who, I think, have really set the tone nationally for this issue.

We chose to really lean on the

most recent IOM report on race, ethnicity and language collection, and so we summarize the particular categories and approaches that that report has set forth, but really also endorse efforts of NQF, HRET and other efforts that have, I think broke significant ground here.

We just pasted in here what the selection criteria, or the data collection criteria are and the approach. I will mention that we highlight in the paper the conundrum, upcoming conundrum, I think we're going to face around multiracial populations and how we're going to need to sort that out.

We tried to talk a little about the importance in our learnings over the last five to seven years around data collection and training of registrars, training of folks who are gathering this information, how to collect it sensitively, how to do public awareness campaigns. A lot of information there that I think that we're going to need to highlight.

Joel has done a lot of work around

this topic of indirect estimation and studied it quite a bit. We talk about the gold standard being collecting the information from the patient, but really needing to think about issues around indirect estimation, surname analysis, geocoding and the like that might be used in the shorter term.

And then finally really thinking about how to kind of hardwire this work into where we're going with HIT and meaningful use and all these other issues. So those are some of the key things we wanted to cover. Next slide, please?

We also provided here the criteria for data collection around language proficiency, with the caveat that the IOM really talks about the importance of local customization of language categories so that those might be different depending on where you are. So that was the background and the data collection piece.

I'll turn it back over to Joel,

who will cover, again, sections three and four. And I'll close this out with an overview of kind of lessons learned and some other key policy pieces.

DR. WEISSMAN: Thanks, Joe. Next slide? Actually, back up one and just to follow up with the data collection recommendation, directly reported race, ethnicity and language is the preferred method, as Joe said. We need to solidify and support the infrastructure for race, ethnicity and language data collection. Joe mentioned the clear guidance from IOM.

And then we just felt like, in the short term, where direct report of race/ethnicity is not feasible, that indirect estimation can be put into place immediately and you can actually use it for a lot of population-based analyses. So next slide?

And so now we sort of get into some of the meat of the recommendations of the report, and we wanted to start by saying that

a lot of work has already been done by the NQF. And that when we reviewed that work, you know, we basically came out and said that we certainly endorse the guiding principles from the 2008 NQF report, National Voluntary Consensus Standards for Ambulatory Care.

And they listed these five principles, prevalence, impact of the condition, impact of the quality process, was there a quality gap, and the ease and feasibility of improvement of quality process. And what's sort of not stated there is each of those principles relates to minority populations, not just populations in general.

But that being said, we thought that prevalence and quality gap were two things to really be highlighted. That you would not want to have a condition, whatever that may be, that might be disproportionately prevalent among a minority population and not consider that for some sort of disparity measurement.

And then the quality gap, if we go to the next slide, our recommendation is that if you look at all of the NQF measures, and I guess there are around 700 right now, of quality of care for ambulatory, institution-based settings, disease-specific measures, cross-cutting measures, that we just felt like they should be crosswalked with the literature on known areas of disparities.

In particular, the AHRQ disparities report has a lot of information on that, but other areas of disparity -- other known sources of literature as well. And any of the NQF measures that can be matched to known disparities should basically be considered disparity sensitive measures.

And then finally, we filed this report and we mentioned this in the report in section five, but just to talk about that now is that we ought to integrate these recommendations with the National Priorities Partnership and the Measurement Applications

Partnership because there are a lot of other efforts going on right now to develop measures for health care reform and other purposes that ought to be integrated with these activities. Next slide.

This is something which we thought would help explain how the section three on measure selection is organized and structured, and we think that it's worth thinking about this in terms of an algorithm. This actually came about as well with some of our discussions with NQF.

But if you think about it, first you might think about known disparities that exist either currently or in the past for specific or similar measures that currently exist among those 700 measures.

So you look at the 700 measures and then maybe you've already collected race and ethnicity data and if there are known disparities in the literature or on the data that you collect, then obviously it's a --

then we think it should be a disparities measure.

But there may be examples where no data exist or where the data are not sufficiently stratified by race and ethnicity or the data exists and maybe at this point show no disparities, but it may be that there's a problem with the sample size or maybe there's a problem with actually how we measure it.

And in those cases, we develop some, again, mostly from the literature but some new criteria that we think ought to be considered for thinking about current measures or measure development in the future. And the reason is that based on the experience of our doing research in this area and looking at the literature, where you tend to find disparities are in these categories.

So for example, care with a high degree of discretion: you tend to find more disparities among vulnerable populations than

in care, for example, the contrast would be care that is highly evidence-based with a lot of consensus on what to do.

Some of the work that I've done with Romana looking at the Health Quality Alliance data, you know, aspirin after discharge. If there's -- or with AMI, if there's already a lot of consensus on what to do, you're not going to necessarily find that many disparities.

Communication-sensitive services tend to find more disparities than others. Obviously, especially with patients of limited English proficiency, but also with racial and ethnic minorities that may have various cultural issues to deal with. I would mention here tobacco cessation for congestive heart failure patients.

The third area would be lifestyle changes, diabetes self-management, wherever it's required to try and gauge the patient beyond the doors of the office where the

clinician is seeing that patient face-to-face. If you're looking at measures about lifestyle changes, those tend to be areas where you're likely to find more disparities.

And then outcomes rather than process measures, kind of related to this lifestyle change and patient engagement issue, but again, once you deal with situations in the communities or family life or other kinds of cultural impacts, there are likely to be differences. You're likely to find more differences in outcomes rather than in process measures.

And then we also at the end, and I think this last bullet almost ought to be added as a fifth criteria, and that is to really consider measures along the clinical pathway. A lot of times these process measures, you know, you tend to sort of pick ones that are available and that are well vetted, but then you miss steps along the clinical pathway that really might impact

outcomes down the line.

And the example that we give in the report is renal transplant, where it's not just about referral to specialists, it's about, you know, are they able -- if they get referred for transplant, are they able to get the workup? If they get the workup, are they able to understand many of the bureaucratic -- and follow and be able to afford in some cases, many of the bureaucratic requirements to actually receive a transplant. And then do they actually receive a transplant?

But there are other conditions, we believe, that have a number of steps along the clinical pathway and that it's important to consider other steps. And this is where I think the clinical expertise on this panel would really come into play.

And then the final category that we talk about is where we believe that in the literature has started to show where disparities may exist, but there really are no

good quality measures that exist at this point.

And if I go to the next slide, we call these disparity sentinel measures. And they're created specifically as measures where research has shown that disparities exist. They're developed based on a review of the literature.

And the best example that comes to play, and we do this at Mass General Hospital, is there's been a fair amount of research that has come about recently on pain management in the emergency department, and yet, there really are no good quality measures out there that are being used.

And so at Mass General now, they've started to do some chart reviews. They've started to measure and they started to collect these data. And they're monitoring it just like you would monitor any sort of other sentinel type of event.

So this is where we think that

it's another tool to, I guess, put in the toolbox and to think about measures that should be developed and maybe aren't at this point. All right, next slide.

And then, you know, in trying to -- when we started looking at and reviewing some of the 700 measures for NQF, it was a little overwhelming, and we tried to make some sense about it. And this was a categorization that we came up with that we think may have some value, just in terms of dividing up the types of disparities measures.

It's not really criteria; these are categories, and these are explained in a little bit more detail in the report. But basically ones that focus on practitioner performance, ones that focus or are based on consumer surveys of patient experience, then you go to health care facility performance, ambulatory care-sensitive conditions kind of get their own category, cultural competency and patient-centeredness. There's a lot of

overlap between these categories, but we thought that it would be a useful way to sort of think about it and group some of the measures.

And then what we actually started to do with it that we recommend that perhaps NQF do this in a systematic fashion, if we look at the next slide, is to go through each of these, each of the measures.

And then -- so here's an example where we started with practitioner performance measures, and first of all determine some of its characteristics, so the column three is the type of measure, whether it's condition-specific or cross-cutting, and try and understand a little bit more about the root of the potential disparity.

Whether it really -- whether it seems like the literature suggests that it's something that the provider can do that -- or that may be responsible for some of the disparities, whether it's more about patient

engagement and patient-based, whether it's systemic or whether it has to do with basically affordability issues, such as health insurance. And then think about whether it's structural, process or outcome. Next slide?

So now I want -- oh, yes, we can have questions. Yes, go ahead.

DR. JACOBS: Sorry, just a quick question. I wasn't sure what PM stood for in the -- is it process?

DR. WEISSMAN: Process measure, yes.

DR. JACOBS: Oh, okay. Thank you.

DR. WEISSMAN: Sorry, I know we should have had a little key at the bottom.

DR. JACOBS: I was thinking process but I wasn't sure. Thank you.

DR. WEISSMAN: Process, PM, process measure, O is outcome and then S would be structural.

DR. JACOBS: Okay.

DR. WEISSMAN: Sorry. Now we are

going to do the methodological approaches to disparities measure, how to measure about -- measure and monitor and I was just having a conversation before the meeting started and I actually think that probably 95 percent or more of people that do work in disparities, especially in the clinical setting, don't realize some of the complexities involved with especially tracking things over time.

And I think that's where, you know, it's you don't have kind of one point that stays still. You've got a lot of moving parts. And we cover all these sections in the report. I'm going to go over just a few of them given the time limitations in this presentation.

So if we go to the next slide, the first point, and I think this is potentially one of the more controversial points that we make is that when you're choosing a reference point we're taking the position that the choice of the reference group should be the

historically advantaged group.

There are other groups to consider. The largest group, the group with the best performance and so on, but we think in reviewing the disparities literature that actually there are some perhaps unintended consequences if you don't choose the historically disadvantaged group to the extent that that could drive some resource appropriations later on down the road. Next slide?

And here's the, you know, the class -- I teach this in a class every summer and this is a slide I use that really displays the differences and differences approach. You know, the question is did black-white disparities get better or worse between 2000 and 2010? And here's, you know, you can look at the absolute difference or you can look at the ratio, and the answer is really both.

Depending on how you do it, you could say disparities got worse or disparities

got better. And the arithmetic is in the report. I won't bother going through that now, but if you go to the next slide the take-home message is that absolute and relative changes in disparities can yield different conclusions on whether or not gaps are closing.

And actually, I think you also might, although we don't do it as much, but if you compare institution to institution depending on whether you look at changes over time in terms of relative or absolute measures, you could find the same sorts of issues.

You have similar issues with favorable versus adverse events. Again, you know, did the patient get the service or did they not get the service. You know, one would be a difference maybe of 90 percent versus 92 percent. The other -- or let's say 98 percent versus 96 percent, you know, and the other could be a difference of two versus four.

One could be a difference of 200 percent difference. The other could be a very small percent difference. Again, it's playing with numbers, but it really makes a difference when you report these sorts of things publicly to find out that there's a twofold difference or there's a 2 percent different really makes a difference in terms of how it's seen in public.

And so our recommendation is that generally you need to calculate both types of statistics. If they're consistent with each other, fine, pick the one that's easiest to report. But if they're not, if they conflict and if they give you different messages then I think at least in terms of reporting that that needs to be noted somewhere.

And that it, you know, allow readers to make their own interpretations whether they think in this case a ratio is more important than an absolute difference. Kind of, it depends so much on the particular

clinical condition and the particular measure that's selected. Next slide.

We go into some detail on paired versus summary statistics. By pair-wise comparisons I mean, you know, if you want to look at black-white or Hispanic-white and what happens is, if you have a lot of different subgroups, that does not become very report friendly, especially when you start looking at changes over time and improvements. You get a lot of different comparisons.

I can tell you that I worked with Massachusetts on a state report card and, you know, there are a lot of different measures and the idea is that you want to make it as simple and user-friendly as possible.

And so the big advantage of summary statistics is that they can address these issues by really taking a lot of information and summarizing it in one number. But, you know, unfortunately that has a lot of disadvantages in the sense that it can obscure

important information, in particular, directionality.

What we found in, for example, we use a summary statistic for pay for performance to reduce disparities in Massachusetts, and in fact in a lot of situations, the disadvantaged minority population actually gets better quality of care than the white population for a particular institution. And yet that would show up as a disparity that may not be eligible for example for pay for performance.

So those sorts of things need to be investigated. So pair-wise comparisons we believe using the historically advantaged group as the reference point, should be checked to see if the summary statistic reflects superior care received by the disadvantaged group. And if so, you just need to sort of consider the context of the report and what the relevant policy goals are.

And they at least need to be made

explicitly because what happens with a lot of these disparity indexes is that they really are not as transparent and understandable compared to some of the pair-wise comparisons which are really right out there. So they're useful, but they need to be explained in detail we think. All right, next slide.

The next issue is sample size considerations and, you know, clearly when you, even on a national level, some of our racial-ethnic minorities in large national surveys can result in very small sample sizes. The smaller the numbers, the more likely disparities will reflect chance rather than true differences.

And there are a number of recommendations that we make that are pretty commonly used in both research and in quality reporting. Again, I won't go into them in more detail other than to mention them here.

You can roll up categories. In fact, generally when you look at subdivided

racial and ethnic categories, the idea is to roll them up into those five OMB categories. But sometimes, and you know I've done this myself, minority versus non-minority, but you know, when you do that, of course, again, the danger is it obscures some of the potential differences that may exist among subgroups.

You can use some of the summary statistics or summary indexes that I mentioned. You can also use composite measures for quality measures. In other words, instead of just the classic one is, the composite measures that are reported for the Hospital Quality Alliance, you know, you've got five or six different measures for CHF or five or six different measures for AMI. And you can roll those up into a single composite and that can help with some of the numbers problem. And then you can combine data for multiple years.

That being said, I think the other point that we make in the report is that, you

know, for high stakes reporting you may have to roll up numbers of minorities in order to make them statistically stable. But at the same time, for quality improvement purposes, it may be worthwhile to look at those racial and ethnic minority subgroups to see what's going on.

It can be of interest to your clinical providers and it can be interesting to administration as well in terms of kind of monitoring what's going on. But when the numbers are small, the statistics just say basically that you can't really use them. Next slide.

Risk adjustment and stratification, we spent a fair amount of time on, and you know, basically case-mix adjustment and stratification are ways to avoid some of the --

CO-CHAIR ANDRULIS: Joel?

DR. WEISSMAN: Yes? Sure.

CO-CHAIR ANDRULIS: Could you hold

for just a second? You have a question.

DR. WEISSMAN: Oh, okay.

DR. CUELLAR: I have a question on the subgroups.

DR. WEISSMAN: Yes, in the back?

DR. CUELLAR: Did you factor in also geography? Just I'm from Texas and I know, originally from San Antonio, and when we look at San Antonio versus Houston, for example, San Antonio the Hispanics are like me. I'm a sixth generation, versus Houston where there is a very large population of Central Americans and it's really very distinct in every aspect of health care.

So from that aspect I think that's a very important point, the subgroups is very important. So I don't know if you even looked geographically because I know particularly with the Central Americans, they're really now migrating to different areas of the country.

DR. WEISSMAN: Yes, I mean I think that's, you know, an excellent point and what

we say in there is that, you know, after you get the five OMB categories then there are a lot of approaches that exist to get into a lot more granularity. And it's difficult to settle on a granular list of racial and ethnic categories that work everywhere in the country.

And so the recommendation that we make is exactly what you're saying. Depending on the locality that you should -- after you get, you know, make sure that it rolls up again, but try and figure out what are the categories that make sense for your particular population.

Joe, do you want to --

DR. BETANCOURT: No, that's fine.

DR. WEISSMAN: Okay. So yes, I would say that. So I mentioned that these are ways to avoid punitive effects, especially with pay for performance and other types of high-stakes reporting that affect providers with disproportionately large, poor and

vulnerable populations.

The issues there are that when you tend to risk adjust, first of all, you really obscure all the racial and ethnic or the subgroup differences, and a lot of people are a little bit concerned that it lowers the bar in terms of if, if minorities are receiving poor quality of care, somehow it excuses the provider for that because you're risk-adjusting based on performance elsewhere.

And then it also -- the other disadvantage of risk adjustment is that it allows providers to perhaps focus on the majority population and improve quality of care there without really improving quality of care for disadvantaged populations. So there are some issues.

Stratification, you know, its biggest advantage is that it's more transparent. To be honest, you're still in some ways saying, well, you know, this is the quality of care for minority populations and

we're going to compare those to other settings and that may be better or worse than the majority population. But at least it's obvious in front of you what categories you're looking at.

So for that reason mostly we recommend that stratification should be performed when there's sufficient data to do it. And that risk adjustment, though, can still be appropriate, especially when you've got -- when performance is usually related to outcomes or proxy outcomes that are highly dependent on community factors beyond a provider's control when you're really talking about quality measurement. Next slide.

And now I'm going to turn it back over to Joe.

DR. BETANCOURT: Yes, so just to bring this to a conclusion here, we talked about priorities and options for quality improvement in public reporting and really see NQF's kind of vision very well-aligned with

what we're trying to do with the disparities-related work.

So we're trying to really, with the development of these measures, achieve a variety of things, monitor progress towards disparities reduction, inform consumers and purchasers, stimulate competition, stimulate innovation and promote values.

To Marshall's point, we tried to spend a lot of time kind of picking people's brains and also looking at real world experiences around kind of unintended consequences, untoward consequences from this work, and so we highlighted a couple of different things, particularly issues that have been covered both anecdotally and otherwise around pay for performance and public reporting that have particular relevance to disparities.

So this concept of kind of cherry-picking and lemon-dropping as a way to make your measures look better, the rich get

richer so certain organizations who have more resources are actually doing better around pay for performance.

We talk about teaching to the test and we give the example of a way that people address measures by simply instituting antibiotics for anybody who comes in with a cough absent, you know, to make sure they get their door to needle time right. You know, could that be done around disparities-related issues, gaming the system?

We talk a little bit about the ability of minorities to benefit from kind of this general QI versus targeted QI, and then this growing field of recognition on kind of between and within institution disparities. Next slide, please?

We tried to touch a bit on kind of what do we see out there right now? And on the federal and state side we talked a bit about how certain states, for example, Massachusetts being one of them, has mandated

the collection of race, ethnicity, language and highest level of education for all hospitals and health plans.

And that has been a pretty good experience now, although we haven't done as well around moving towards routine monitoring and measurement. But we do have hospitals that have demonstrated that this information can effectively be collected.

Health plans as well, we'll talk more about health plans in a moment, but the National Health Plan Collaborative, several of the health plans who are here, have done great work in this regard as well. And I think those provide significant lessons learned.

We have seen some Statewide Health Disparities Report Cards, but their primary focus has been just stratifying health outcomes by race-ethnicity, on occasion looking at certain measures, but again, I think there's something to be learned there but perhaps kind of looking more at kind of a

general health outcomes prevalence and epidemiology, less on particular measures. And we might be able to learn, I think more around statewide efforts in that regard. Next slide, please.

From the standpoint of health plans, again, National Health Plan Collaborative being the optimal example, but we do see certain routine collection of race-ethnicity data collection continuing to be a challenge for health plans for a variety of reasons.

There's indirect estimation. We have others who are doing some direct collection, geocoding and surname analysis as well. We have seen in some of our work that local, smaller plans can kind of get data from the state, get data from different places that allow them to provide better profiles of their member race, ethnicity and language as opposed to some of the challenges that larger national plans face. So something to keep in mind and

to explore as you go forward. Next slide.

We provide here an example of the work that Aetna's doing going beyond race-ethnicity data collection to the development of their first Racial and Ethnic Equality dashboard, which has a variety of different elements including disease prevalence and diversity around their geographic market segments, stratification of CAHPS and other quality measures that they are now stratifying now and developing as an annual report. Next slide.

From the standpoint of hospitals, again, we see that hospitals have been -- some hospitals have been able to create routine disparities measurement and monitoring tools, dashboards and/or reports. They have begun primarily by taking off the shelf measures and stratifying them by race, ethnicity and language if that language is available.

And again, to Joel's point, connecting these wires is very, very

challenging. It's not as easy as just, well, let me take the core measures, let me take the data and just kind of connect it. Oftentimes these systems don't speak to each other. They're not connected. Generating these reports, there's a lot of devil in the details, but it has been done.

A particular challenge, I think, for hospitals has been small minority sample size for particular conditions, limiting statistically significant comparisons.

And we also highlight the importance of an appropriate communication strategy as we try to go public with disparities reporting. There's a lot of concern from hospitals and other organizations that if we say that we're doing work in this area, are our patients going to be concerned? Are providers going to feel blamed? And so as we move towards public reporting, communications is going to be really essential. Next slide, please.

We highlight some of the work that we've been doing since 2007 at Mass General Hospital where we have a disparities dashboard release once a year that this is just the executive summary.

It green-lights areas where we don't have any disparities and we stratify our national hospital quality measures, HEDIS outpatient measures. And, as Joel mentioned, we have a sentinel measure around pain management in the emergency room for long bone fracture. So we monitor that.

Orange, lighter areas are where we see disparities nationally and we're exploring them now, mental health and wait time for renal transplantation. We are now stratifying all-cause and ambulatory care-sensitive admissions, CHF readmissions and patient experience.

And then red light are areas where we found disparities and developed interventions, something that we think is

very, very important because as you go public you need to communicate to the community that if you found something, you're doing something about it. Next slide, please?

We also publicly report so anybody could go to MGH Quality And Safety and come and take a look at our equity report, look at our numbers and look at our improvement stories. And again, we can provide that link inasmuch as it might be helpful. Next slide, please.

So I think we have -- we ended with a series of questions related to policy and dissemination. Clearly, we believe that, when we move towards public reporting, standardized measures that are easily understandable and actionable are essential, capitalizing on available measures used for quality reporting is a great place to start. OMB categories used and adapted over time with capacity for local subgroup variation.

And then, again, when we debate

this issue of public reporting how should it be used? Should it be used for payment reimbursement or consumer choice, provider incentives?

And this packaging piece is really, really important. How do we explain disparities to providers, the public, organizations? Think about root causes, link it to quality improvement in a way that doesn't get people to shy away from or be scared of doing work in this area, but instead engaging and really seeing equity as a key part of quality.

So that's the end of our formal presentation. You can go to the next slide. We're happy to answer any questions that you all may have, and thank you so much for your time and attention.

CO-CHAIR ANDRULIS: Thank you, both, for the comprehensive and very, very helpful review. We're going to enter the question-comment stage now, and I'd just ask

you when you raise your cards, make sure I can see them, so please feel free to get started.

And I'll get started with you. I just have a couple of questions about whether you had discussions about a couple points which seem to dance around the edges of your focus around social determinants. And as you touch on that in lifestyle issues, which is, to me, kind of like a broad opening in some ways to that. And whether, because of the close correlation of health literacy to a lot of these issues of culture and language, whether these were discussed and what context you might put it in?

DR. WEISSMAN: Do you want to take that first?

DR. BETANCOURT: Yes. So I mean I think, you know, if you look at these measures that NQF has already they really span the gamut, right? You have something as simple as asthma assessment that Joel highlighted which is a very provider-based kind of piece. And

then you have others more outcome-related where social determinants play a much larger role.

I think what we're recommending is that, you know, these are going to need to be looked at in different ways and some of these that have -- that we would think for certain communities have a greater social-determinant causal connection, then how can, you know, how do we make those disparity-sensitive and what are the appropriate pieces that you need to put in place so that you can correctly gather that?

It's critically important, no doubt, but the measures, again, run the gamut. Asthma assessment is something done in the office. Something like some of these outcomes, diabetes is the example we gave, you know; that requires a lot more work and I think that's something that's going to need to be debated around focus and scope among this committee.

DR. WEISSMAN: I was just going to say, leave it to Dennis to really get at the core issue of the unexplored areas in the report. You know, we did kind of dance around it and address it in a number of different areas, and clearly the social determinants tend to be more important for outcomes, but even for process measures.

I was just reading an article recently about racial and ethnic disparities in TRICARE, the military health system, which is, you know, supposed to be uniform care for everybody, looking at asthma care for kids. And they get sent different rates to different specialists and they get different care even within TRICARE. So even in process measures, social determinants are important.

We make the case that if -- to basically not adjust for social determinants when reporting racial and ethnic disparities, that even though it might be a mitigating factor or an explanatory factor, that it

doesn't say that those disparities don't exist. And that I think especially when reporting to the public, people don't want to say, well, you know, this group may have disparities but if you adjust for health insurance status, the disparity goes away. You know, I don't think that is reality because I think the disparity still exists.

I think the other side of the coin, which we touch on a little bit, is in how much to hold the provider responsible, especially when it comes to high-stakes reporting. And a lot of providers will say, you know, you should adjust my patient population for social determinants because there may be communication issues, there may be transportation issues, there may be all these sorts of things which are more difficult and therefore more costly for me to address.

And there's one recommendation that we kind of threw in there that I've been exploring recently, which says that, if that's

the case, then maybe what we ought to think about doing is risk-adjusting payments to providers.

In other words, if they have a more difficult population, pay the provider more based on their population, but then hold everybody to the same standard in terms of quality outcomes, which is something which I don't think has been tried anywhere, but is something that might be considered in the future.

CO-CHAIR ANDRULIS: Thank you. It's just something that I think as the group goes along that there's clearly more of an emphasis on these issues of social determinants. You can see it reflected in the health care law and it might be something for the discussion along the way.

Why don't we -- for questions why don't we go around the table starting with Elizabeth and we'll just do a round for now and pick up people as they are identified.

DR. JACOBS: Actually, mine's not a question but an offer of a resource. So I noticed that you talked about measuring language in the report and that you weren't sure about the best way to ask it, the two questions, the OMB questions versus preferred language.

And it turns out that Leah Carliner and colleagues have done a really nice paper actually showing the sensitivity and specificity, excuse me, of using those two questions and how to phrase it, and I will email you that reference. I think that would be an excellent thing to include because it's really well done.

CO-CHAIR ANDRULIS: Thank you. Norman?

CO-CHAIR CORA-BRAMBLE: Before you go on, could you make sure that this -- one of the staff persons gets the paper?

DR. JACOBS: You know what? I'll send it to the whole group.

CO-CHAIR CORA-BRAMBLE: Okay.

DR. JACOBS: I'll just send it to the whole group --

CO-CHAIR CORA-BRAMBLE: Okay, great, thanks.

DR. JACOBS: -- the reference so you can all have it if you're interested in it.

CO-CHAIR ANDRULIS: Great, thank you.

DR. OTSUKA: First of all, a selfish comment, I'm just surprised on the dashboard you don't include musculoskeletal care or health as one of the major criteria, especially with our aging population. I mean what, there's -- I'm a pediatric orthopedist but there are, like, 250,000 total hips being done in the American population. So I'm just surprised it's not one of the categories up there with diabetes and asthma and fractures and trauma, et cetera.

DR. BETANCOURT: In which

dashboard are you referring?

DR. OTSUKA: Well, you had one of the dashboards with the six main health determinants, asthma, cardiovascular disease, one of your slides -- it was also with Aetna.

DR. WEISSMAN: Can we go back? Let's just look at it. That one?

DR. OTSUKA: Go back. Go back. There, quality measures. You're measuring --

DR. BETANCOURT: Oops, come back now.

DR. OTSUKA: -- asthma and --

DR. BETANCOURT: So this is just an example of what they're doing.

DR. OTSUKA: Okay.

DR. BETANCOURT: This is just kind of a real world example.

DR. OTSUKA: I thought I saw that in your paper, too, but, you know, maybe I'm mistaken.

DR. BETANCOURT: Yes. No, I don't think we recommended that you should -- I

don't think that we recommended particular areas. We were just -- this is just highlighting what they're doing. But certainly, issues around musculoskeletal when you think about disability and --

DR. OTSUKA: Right.

DR. BETANCOURT: -- mortality and morbidity, I'm -- we're with you on that, without a doubt.

DR. OTSUKA: Okay. And a question, you're talking about measures and quality of care and quality of life, functional outcomes. I think there's got to be a little more focus on the patient.

For example, you talked about some off the shelf measures. Let's just say, for example, pain management. It sounds pretty easy. You know, you give them morphine, pain's gone, pain isn't -- Likert scale zero to whatever. But Asian men metabolize morphine or codeine at different rates so presumably their response would be different.

Their requirements are different.

The other thing, I really impress upon you to focus on the patient. I mean we're here for -- this is the Quality Forum. I mean if we're measuring quality we have to look at the patient and the disparity or the specific population we're looking at has different adherence and compliance. I guess compliance is the old term, but they adhere to treatments differently and that'll certainly affect your quality of care.

For example, if a kid breaks their arm -- I see that a lot -- a Manhattan kid, family would bring their kid in right away. But if the kid was, for example, not to pick on Asians but I guess I can because I'm Asian, an Asian kid from Queens would have banana leaves or some type of ointment placed on it, and that would affect the quality of life. So that brings up issues of adherence, compliance, et cetera.

The last point I want to make is,

again, we're measuring outcomes and quality of care. We have to -- the measures have to be specific for the population. For example, I try to make kids walk and sometimes if I can make a kid walk 10 feet, that's the best thing in the world, okay? But in the eyes of some parents, if a kid never walked before and they're walking 10 feet, that's great, but if their foot's a little turned out or a little turned in that's the worst result in the world.

If you do a total hip in a patient and they can walk more, great. But if they can't sit or if they can't kneel for their religious beliefs or whatever beliefs it is, that's the worst result in the world.

So my point to you is, let's -- I mean, you have all these off the shelf measures that you talked about, but we have to focus on the patient and there's no point sitting here and measuring infection rates, pain management. I think that's a waste of

the brain power in here and your time and my time unless we really focus on the patient and what the crux of the matter is.

And we're here because the populations are diverse, you know? And there's no point in measuring mortality, morbidity. Let's measure quality of life and what specifically we're helping with the quality of life of that patient. Sorry to monopolize time. Thank you.

CO-CHAIR ANDRULIS: Yes, on that point I was -- just an add-on for further discussions down the line, I was wondering whether you folks had ever looked at the Picker -- went to the Picker Patient-Centered Care measures, you know, since there was such an extensive amount of work that was done on that.

DR. BETANCOURT: Yes, I mean on slide -- so I agree with everything you've said and I think that's really the area that we need to push the most on. But if you look

at what's readily available today is the least developed. So I think when we talk about off the shelf it's our recommendation is moving on parallel tracks.

My preference, I guess I'm going to editorialize a bit here, I believe that we need to look at the measures we have by race-ethnicity that go a long way at getting at patient experience because, you know, if you look at HCAHPS alone, right, and you stratify by race and ethnicity, we're seeing some very significant differences by subgroup.

And my sense is that we're leaving a lot of voices out there in the dark that our current mechanisms of really getting at patient experience, getting at some of the issues you mention, are really in their infancy stages.

Making progress in that area, and this is something that we're thinking about from a research standpoint, of really, you know, getting at patient experience in other

ways, whether, you know, outside of surveys, other strategies, I think, is going to be critical. And I think that's something that we want to highlight if we haven't highlighted it enough.

DR OTSUKA: Yes. We obviously have to use what we have. I mean there's only so many outcome measures and surveys that are available, but what I've done in children's orthopedics is actually validate, tried to validate them at least and we should at least make an effort to validate it.

For example, well, I mean there's a general survey I can give to every kid, but I have -- I worked in Los Angeles and I had a big Hispanic or Latino population. I actually validated that study in those kids. And it presumably worked out. It did work out well.

DR. BETANCOURT: Right.

DR. OTSUKA: And I feel good about reporting data or -- with that outcome measure, but I agree with you.

DR. BETANCOURT: Yes, exactly.

CO-CHAIR ANDRULIS: Thank you. Mara?

MS. YOUDELMAN: So I have a couple thoughts, sort of not just on the paper but in a broad concept and one I just want to sort of register as a little bit of a concern. And I completely recognize in having worked on these issues for the last, whatever, however long, focusing on race, ethnicity and language disparities. But I do just want to also recognize that as you guys said, there's a lot of other disparities based on other populations.

And so one thing that I think might be worth some discussion at some point is when we are talking prospectively, how does NQF look at these measures, how do we also think about looking at these measures and developing them for other populations?

And in large part and maybe this is part of my disclosure that I didn't do, but

one of my other hats is I co-chair the Leadership Conference on Civil and Human Rights Healthcare Task Force. And this has been a big issue for this task force of making recommendations on all the different types of populations that might be affected and in large part because people aren't often in just one population group.

And so I think that is just -- it behooves NQF to sort of explain why the focus of this project is specifically on race, ethnicity and language, but also how we can use what we've learned because we have come a long way.

And it is amazing when you guys said, you know, we're better off in race, ethnicity and language than some of these other populations and that's due to this huge work by lots of people, but it's also sort of frustrating that we're not there in some areas, too. So I think it is just an explanation and some background and then

looking how we can address it going forward.

The second piece, and I think you guys addressed this to some extent, too, was on language. I think that we really do have to make sure the focus is on collecting both language proficiency and language data.

And I mention this not because I think this group is -- or the paper was sort of lacking in it, but more because of what we're seeing from policy side of things where because if there's an added cost to doing it, recommendations just came out from the federal government to collect language proficiency but not language.

And I can't see how you can get to disparities and identify if you only have, do you speak English very well, well, not well, not at all. And so for the same reasons I think you guys make the case we need to get more granular data on race and ethnicity.

I think we have to make that exact same case that we need that -- it shouldn't be

granular, but granular data on language and not just a language proficiency standard as we're looking at the measures.

And then the third piece, and really this is just a comment, Joel, on your issue about risk adjusting. We actually have been looking at that from the policy angle and trying to make that exact case that if you're looking at an Accountable Care Organization or medical home or something like that that there should be some recognition that we do need to risk adjust based on, you know, language to pay for interpreters to translate materials or to cover some of those other issues you discussed or others, like the social determinants of health, that if someone's now coming into care that they might need more care to catch up because of the history of disparity.

So I think that sort of dovetails nicely with what you said. It's hard to create, sort of the evidence for that, but I

think we're trying to make that case. And to the extent we can make more of that case through the evidence and through some of the standards that collect it to make the case going forward, again, that would be a good prospective way to sort of go forward.

DR. WEISSMAN: Yes, wouldn't it be nice if ACOs that served minority and disadvantaged populations were well-resourced rather than always --

MS. YOUDELMAN: And that's exactly our case. I mean a lot of it we're seeing, and I think this dovetails, you know, exactly with the policy developments now, and I apologize. I live and breathe this policy stuff.

But that, you know, we're testing all of these new payment systems, and if we're not collecting data and stratifying data by race, ethnicity, language, et cetera, we're going to end up with payment systems that work well for the average and not for anybody else

and we're just going to perpetuate the disparities that we've seen going forward.

So we're at a perfect time I think exactly for this project but just need to keep all of the policy pieces in mind, as much as the standard-development.

CO-CHAIR ANDRULIS: Thank you. Edward? Oh, I'm sorry. I can't see your card there. I'm sorry. Thank you, Donna.

DR. WASHINGTON: Okay, thanks. Joe and Joel, I really appreciate the attention the report pays to sample size as well as to some of the consequences of reporting out for minority-serving institutions. And so the comment I'm going to make has to do with reporting out by institutions that don't have a sufficient sample size to stratify their population or examining disparities, where they might actually look good, be inappropriately labeled as sort of a non-disparity-related institution when in fact they don't have sufficient sample

size.

And I wonder if you might want to address recommendations for reporting out when the sample size isn't sufficient. I'll give the example of the VA health care system. They collect quality and satisfaction data on a sample of patients at each site, and have been publishing report cards for the past three years.

Last year, for the first time with the 2010 report, then they reported satisfaction stratified by race-ethnicity for every single VA facility. And what they did to address the small sample size even in a large -- it is the largest integrated health care system in the country, and despite the large samples sizes available, then only about a third of VA medical centers had a sufficient number of African Americans, for example, to report out indicators, report out performance on individual indicators.

And so what they did was to list

not applicable for the facilities that did not have a sufficient sample size to avoid sort of just targeting the minority-serving facilities. So I wonder if you could comment on recommendations or thoughts about how other systems should approach that issue.

DR. WEISSMAN: I don't have a good answer. I'll tell you that. And I think that what the VA does is an acceptable approach. I think saying not applicable or not having sufficient sample size is not actually making them look good. It's just saying that they don't have the relevant populations.

The same thing happened in Massachusetts except even to a greater extent. We don't have quite as diverse a population as other states in the country. And when we were looking at racial and ethnic disparities in hospitals about, you know, two-thirds or three-quarters of our hospitals really didn't have a diverse population.

And in that case though what

happened was there were still some statistics that were generated that actually made them eligible for incentives, which may have been inappropriate. So I think that we do have to pay attention to how the incentives are structured and how the formulas are structured.

And I also think that there's an issue that hasn't -- that doesn't often get addressed and that is, when they're -- if a program is intending to reduce racial and ethnic disparities, then minority-serving providers that have a lot of the minority populations that you're interested in, there may be some programs that may be targeted towards those institutions because for -- this particular kind of incentive program. So that if the provider doesn't have a lot of minorities then they wouldn't be eligible for those sorts of incentives.

So I think there are a number of ways to go about it. I don't think any of

them have really been explored in great detail and I'm, you know, sensitive to the issue that you raise.

DR. BETANCOURT: I do think we suggested what some organizations are doing that we've seen, which is rolling up to white, non-white, not optimal but, you know, that's one thing, rolling up multiple years.

I mean at the end of the day, I'm a firm believer that, you know, if you have 20 minorities or, you know, 90 percent, that you should be able to show that everybody's getting high-quality care regardless of their background.

And so saying not applicable or sample size too small, I think there are things that we should be able to do that really, you know, put equity at the center of quality measurement regardless of that sample size. So those are some of the things that we've seen on it.

DR. WASHINGTON: Let me just add a

comment. Would you recommend -- so I think non-applicable is actually a good solution if the numbers aren't stable. But would you recommend actually just publishing the numbers, both the numerator as well as the denominator so that people viewing the reports can make their own judgment about the sample size?

DR. WEISSMAN: I would. It looks like Joe and I disagree, but I think if the numbers are not -- and I think we have some statisticians in the room -- if the numbers are not statistically significant, I don't think it's fair to the organization to report the numbers. I don't think it's fair that we should expect the public to make a decision on the statistical stability of the number.

But I also agree with Joe that to the extent possible, you know, maybe we're not doing enough. Maybe we're saying, okay, in this case we're going to clump together two or three years, roll up some of the populations

and report the numbers in that way.

I feel a little uncomfortable in simply reporting, you know, what if they have two or three minorities and -- or what if in a particular group? It can get very anecdotal at that point, and I don't think it would do justice in my own opinion. We don't always agree on everything.

CO-CHAIR ANDRULIS: Thank you.

DR. HAVRANEK: First of all, thanks for this. I thought it was really thoughtful. I found it really useful, and I thought it was really well done. So again, thank you.

I had a few questions related to socioeconomic position. The first is I wonder if you could help us a little bit more with a very operational definition of that? That I think socioeconomic position is something we all understand but when it comes down to actually measuring it at an organizational level, I think we need a little bit more help.

So I think you allude to income as a measure of this, but I think that there are some real practical problems in measuring this in minority populations or using this in minority populations. And I think it has a lot of weaknesses. You know, a lot of minority people have income that they won't report to us because it's, one, it's illegal, or two, it might kick them over into a co-pay that they wouldn't otherwise have.

Second is the opposite problem which is that, you know, we certainly see patients who have little or no income. I'm thinking of a patient I had trouble getting bypass surgery for because he wasn't -- he didn't have a job.

And he didn't have a job because he'd just sold his share of his car dealership and so, you know, I think income doesn't -- has some weaknesses. And I would love some more help measuring socioeconomic position.

Second thing is one of the

problems is that disparities might not exist within an organization because the disparities exist at the door. So and I'm thinking of large academic centers that have -- can report no disparities based on race, socioeconomic position, what have you, but they have, let's say, 5 percent of their patients are minorities or of low socioeconomic position and they are situated in communities or neighborhoods where 20 or 30 percent of the residents of their catchment area are minority patients. And that -- I wonder if you could help us understand if that's a worthwhile thing to measure or to deal with?

And then the final thing is I wondered what you thought about socioeconomic position itself as a basis for disparities, that you kind of touch on this a little bit when you talk about interaction terms, and, you know, it's -- I applaud you for even taking it on because it's so difficult.

But I'm a little bit concerned

that socioeconomic position may be driving more of these disparities than we would like to think. I'm thinking of some anecdotes of patients telling me that their status as having Medicaid, they felt, drove their care at other institutions. In other words, you're a Medicaid patient. Therefore you are X, Y and Z. Therefore I'm doing this to you.

Or, you know, I often see providers react to patients based on things that are strongly related to socioeconomic position. So I mean simple things, dentition for instance, you hear other providers refer to a patient's two or three teeth as a basis for the way they're reacting to them. And so I just wonder what you think and whether or not that's something that this group should be thinking more about? So sorry for the long number of questions.

DR. WEISSMAN: Do you want to start? Great, a lot of meat to chew on there. Let me start with your second point first

about disparities existing at the door. I think it's a very interesting point and I draw back on some research I did as a graduate student, believe it or not, where we were looking at policies about -- what's that? Yes. No, it wasn't that long ago.

We were looking at policies around bad debt and free care, and we found that hospitals in the fairly well to do communities had fairly liberal policies. And the ones that were in the poor communities had much stricter policies. And why? That's because they didn't have a lot of people coming in the door that actually required bad debt and free care.

And I think the same sort of thinking could go around providers in high minority and low minority populations, that we do have to be a little careful about. That if there's a low minority population there aren't that many racial and ethnic minorities to deal with.

They may have -- be better resourced and be able to really address those equity issues and come out looking pretty good, whereas the provider that has, you know, 10 or 20 different languages to deal with and, you know, a very diverse population, it can be very complex and could run into a lot of challenges. So I think that it's a very good point. I don't know what the policy response is, but I think it's something that is worth thinking about.

As far as socioeconomic position, I mean I think there are experts in this room that are -- that know better than I do about how to measure those things. And for example, the literature doesn't even agree on the terminology, you know, and socioeconomic status is the one that most people are familiar with. And, you know, it certainly can be the basis for some disparities.

In terms of measurement, you know, the usual things that people talk about are

income, education, and occupation. Those are the three sort of kind of classic ways to measure socioeconomic status. And I would also say that you might even use some ecological measures, in other words the socioeconomic status of the community in which the patient lives in. I think some community level variables would also be important in terms of identifying the socioeconomic status of the patient.

But, you know, when I teach about this, you know, it's hard to collect all that data. Some of it is unreliable. Income in particular has a lot of non-reporting issues, and you know, probably education or even the education of the parent may be the best single signal that we could get, especially if you think about, again, trying to use one measure to address everything.

But I mean especially if you think about what might impact a patient's use of resources and the way they use resources and

that sort of thing I think would be a good way to go. But it's -- but there's no great way to do it.

DR. BETANCOURT: And I would just add that, you know, there's no doubt that socioeconomic status drives a lot of disparities, but we still know that there's a significant chunk that's unrelated to socioeconomic status. And, you know, as Joel mentioned, there's been -- David Williams has written eloquently about issues related to, you know, wealth and deprivation index. I mean, there's 10 ways to slice this.

In Massachusetts from a very practical standpoint, we collect highest level of education as a proxy for SES and combine that with insurance status to try to get some sense. But even as far as I think we are around monitoring and measurement, we haven't begun to stratify any of that yet. We're collecting it and so we're still at kind of, you know, the early stages of this. So I

think that's going to have to be kind of a best fit choice for the group.

DR. WEISSMAN: I did hear just today that the governor of Wisconsin did not go to college, so I don't know what his socioeconomic status would be, but you'd have to consider occupation in that case.

CO-CHAIR ANDRULIS: Thank you. I think we have enough time to probably cover all the folks who have questions or comments now and that's going to be it, and then we'll be on for a break. Sean?

DR. O'BRIEN: Well, I'd maybe just pause and ask what is the plan? There's question and answer now but then there's going to be a series of recommendations. Will there be time for discussion of each individual recommendation?

CO-CHAIR ANDRULIS: I think ultimately we're going to be looking for some consensus on this, on the points and the recommendations around this paper. And our

discussion will also be formative and bring up other points that we'll consolidate. We're not going to do this kind of at the end of each segment, but there will be some --

DR. O'BRIEN: But this type of discussion --

CO-CHAIR ANDRULIS: -- points to clear.

DR. O'BRIEN: -- is basically the plan for the day? I'm just wondering if I ask a question now or don't ask a question now that we're basically coming back to these issues later in the day in any case? Is that correct?

CO-CHAIR ANDRULIS: Yes. I would ask it now and take advantage of the --

DR. O'BRIEN: Yes. Okay, well, I mean, some of my questions may be a little too much detail or something like that, but I suppose first I would just say --

CO-CHAIR ANDRULIS: Well, in that -- Sean, yes, in that case maybe we want to --

if it's granular, getting more granular detail we are going to have other, many other opportunities so feel free to --

DR. O'BRIEN: Okay. Well, then I guess I just start with a couple comments or questions. I really didn't learn a lot from the report. I appreciated it. A lot of the -- well, I guess one question, one comment I would make is that there's other sources that you drew from in this report that also presented guidelines.

And I think it's probably worth any places where your guidelines were different from other published guidelines. I looked at one that was specifically, I think, published by National Center for Health Statistics related to the methods used for the AHRQ disparities report and it had 14 guidelines. And I look across and there's really a lot of consistency between them and your approach is very similar to theirs and is a very nice delineation of the issues.

But there were a couple where I noticed, okay, well, that the -- one person in this report made a recommendation that was kind of different from what you guys came up with. And I could mention it or not or just come back to them later, but it's probably worth having some focus on why your recommendation was slightly departed from other approaches that were out there.

I think a lot of your recommendations focused on reporting issues and issues, one big issue is when you do things two different ways and you can -- the results can really depend on some subtlety of how you do the analysis, whether you report ratios or differences. And I think one of the recommendations is, well, the only real way to address that is to report things both ways when they contradict one another.

And I think that is -- I agree with that approach. But ultimately when the call for measures goes out they're not really

focusing on well, how do we report these measures? They're really saying what is the measure? And they're saying basically tell us a numerator and a denominator.

And I'm not exactly sure what for something that will be submitted to NQF for endorsement that's related to disparities, what will that look like? Will that be basically here is a measure and we say what the developers are proposing is to basically look at this stratified by subgroups? Or will they be explicit and say we think the ratio of the usage rates for this procedure across this population X and this population Y?

If in a situation where results may depend on whether you're reporting ratios or differences, is there going to be a separate measure that one's looking ratios and one's looking at differences? Are those both going into the same measure? And so I just think there's some -- how some of these recommendations get implemented in kind of the

way things are set up at NQF will take some work.

CO-CHAIR ANDRULIS: Ellen?

MS. WU: So I'm hoping that, I don't know, you guys or the committee or NQF can do something about this issue that's actually come up for many years is around the CAHPS survey? And that obviously it's not a perfect tool for patient satisfaction and we have to get at it different ways.

But certainly one issue is that it's only being done in English and Spanish right now. And there are actually versions, translated versions in other languages that are not being used because it hasn't been certified by NCQA.

So and they're looking at resources, you know, they talk about resources and how to do that, but it feels like fairly simple. I mean, if it's a resource issue in getting the translated versions certified, since they're already translated, issue to

address to be able to -- because, you know, if we take the CAHPS survey and stratify it by race-ethnicity it's not going to yield very much if people who don't speak English and Spanish, who can't read English and Spanish can't fill it out.

So it just feels like low-hanging fruit that we can kind of get at fairly soon. So I don't know if anyone can help with that?

DR. BETANCOURT: No. I mean I couldn't agree with you more. I mean, I think we're struggling with this as a hospital right now because I'm, you know, I think our HCAHPS is only done in English and Spanish certainly.

And we did actually, just as a quick anecdote, we did a couple years back because of our unhappiness with that we did a survey, both telephone, multi-modal -- telephone and at point of care and the like -- around patient experience where we used some tools like the Commonwealth Fund Minority Health Survey, the Kaiser Family Foundation

survey on race and discrimination.

We created our own validated survey to get our patients -- oversampling minority patients. And, you know, we picked up all types of things that HCAHPS never picked up. And so I think, and this gets back to your point as well, which is I strongly feel like we need to push the boundaries there and think of creative ways to get at patient experience.

Now, we're thinking of do we need to do a survey like that once every year, you know, once every two years to get at these voices we're leaving behind? I think it's challenging from a policy standpoint because I think if people's numbers look good they don't want to tinker with them, right?

They don't want to tinker with HCAHPS if everything looks nice and so what's their motivation? But I do believe as advocates we need to continue to push that and figure out ways in which we can do the easy

things like translate some of these surveys.

But also I think it's not only going to be just surveys. I think we're going to need to, you know, we have patient and family councils. We have -- you know, is it focus groups? Is it other ways of getting these voices at the table, again, things that I think we should be doing in parallel.

DR. WEISSMAN: And I would just support the idea of getting other ways of collecting data for the simple reason that it's not only a language issue, but it's also a literacy issue where I think some of our disparities occur and that it's, you know, kind of who's not here raise your hand. It's very difficult to assess what kind of disparities occur among a population that has a literacy problem.

DR. BURSTIN: Just one comment to weigh in before we go to the next one. This is exactly one of those issues that I would hope that we would potentially put in a

parking lot for us to talk about tomorrow.

I think there is a real opportunity potentially, for example, as the NQF evaluation process to say if it's a patient survey, and this is where you guys would come in, is there -- if a certain population rises beyond X percent of the population there should be a validated survey. I mean, if there's ways for us to push that envelope I'd want us to return to that point, but not right this second.

MS. CUELLAR: A couple of comments, going back to the social determinants, I think one of the things that's very important, certainly in the southwest, is lay interventionalists, for example, the Promotoras de Salud and their use. And I think it's a very important point.

Also, as Norman said earlier, as a pharmacist I have to say that really looking at pharmacodynamics is very important, both in the Asian and the black population and in the

Hispanic. Someone may be diagnosed but is poorly controlled.

Also, too, I think it's very important is a first level of health seeking behavior. I'm a hospital pharmacist for many, many years but my brother's a community pharmacist, and I actually sat and watched him one day. And he did 21 patients that he sent literally to the health clinic because coming to the pharmacy was their first step. And either taking blood pressure or whatever, he was the one who intervened. And I think often your community pharmacists are often not looked at.

Also, in the language proficiency, also I think really drilling down and the paper you might be sending us may address this, but I deal with a lot of people who speak Spanish but can't read or write it. So I think that's very, very important. And, you know, measuring the impact of visual aids and unfortunately we do not have enough visual

aids to help with some of those issues.

The other point I think is very important is that lack of diversity and measuring that in organizational leadership. I think measuring that and how they view disparities and the importance of disparities in the organization is very critical, and particularly now where people are looking at quote indigent care versus disparities. That has kind of taken the forefront. So I think that's very important.

And also leadership, not only leadership but the care providers, so I really think those are all important points to discuss.

CO-CHAIR ANDRULIS: Thank you.

DR. CHIN: Thanks also for a fantastic paper. A question for Joel and Joe and maybe also you, Helen, that there's sort of another possibility that you maybe spent two sentences on, but which is somewhat the simple solution that finishes the work in

committee in half a day as opposed to a year, just wanted to get your thoughts in terms of why you didn't explore it further.

Just basically why not just use all the basic measures which are being used more generally for quality improvement and stratifying in this case by race, ethnicity, and language. You could then supplement with just a few things, so, you know, there's existing measures like interpreter services or you mentioned paying for long bone fractures, or Norman's point that there's going to be sort of a research agenda in terms of more disparity-specific things.

But the vast bulk of measures, you know, probably more than 90, 95 percent would currently fall in just the current measures. I mean you talk about disparity sensitive, but I mean they'd really be a subset of the ones that already exist really.

And this approach, I mean, is simple in terms of well, you know, you just

stratify by race, ethnicity, language. It's flexible, so Lourdes' point about well, you know, the key measures may differ depending upon geographic region, population, et cetera, so if you look at everything and then stratifying by, you know, key variables, you know, you'll capture it.

In any case, as Donna was saying, the sample size issue is going to drastically reduce whatever measures you come down to and so it's not going to make that much difference.

And then this sort of important philosophical point that I think one thing we do want to try to avoid doing is this issue of marginalizing disparities. You know, so well, you know, we do our mainline quality improvement and then we have these, you know, disparity measures and then we'll think about disparities for these measures as opposed to thinking about disparities in all of our patients for, you know, everything we do in

quality.

And there's also then this sort of implementation simplicity issue, and this may be a question for you also, Helen, in terms of, you know, the users of this, whether they're policymakers or organizations or something like NQF MAP that Joe is a part of the steering committee for, it's a much simpler argument to say, well, look, you know, we have a couple principles.

Whatever you do stratify it by race, ethnicity, language, add on these additional disparity specific measures. It's a much easier argument to make than us arguing to, you know, a broader committee here's this more complicated process in terms of finding disparity sensitive measures and go through this. And it's a much more involved argument as opposed to two or three simple principles.

So why isn't sort of the simple solution the way to go as opposed to something that's going to be a lot more effort or, you

know, a lot more consultation as opposed to -- you'd be pretty much in the same place probably. So what's the advantage?

DR. BETANCOURT: Marshall, I think -- I mean, I'm a firm believer that -- and I've been an advocate for that particular position that one thing to do, and I think the approach for us to not marginalize, as I say, we should stratify everything by race, ethnicity, and language. So I mean, I think if this -- if what came out of this committee was that recommendation I'd certainly be very happy with that.

So I wouldn't recommend this over that. I think that is the gold standard and that's a way that moves away from this kind of, well, this is the equity bucket here to, you know, everything for every patient who we manage in any way that there should be quality and equity.

So this is a plan B if you will. I do think that the off the shelf measures

which came up a bit before we do miss certain things that we haven't paid attention to. I think that's where we call the sentinel measures might provide opportunity there.

But I think if this committee said, "Well, stratify everything plus we need to look at new areas," that for me would be a very reasonable recommendation.

Now, certainly there's the actionable, feasible lens that needs to be applied, which is that's something that you all need to debate. I'll just say personally, I'm an incrementalist. I do believe that we need to kind of crawl before we walk, before we run, and I think this committee would go a long way by just thinking about developmentally where we could go.

So that might be a great recommendation, the question is how actionable is it? I don't know. That's something that I think will require some exploration from this group.

DR. WEISSMAN: I would also say that the point you raise brings up the tension between some sort of overall measure on everything, on the universe, everything that we collect. You know, how does everybody do versus trying to look at a focused set of measures? And I guess this is your challenge I assume for, you know, what is the goal, the end result for NQF in this exercise?

And so you can imagine that some provider somewhere gets a score for how they do on disparities or equity of care, and is that based on just kind of, you know, all 700 measures? Or is it based on a focused set of measures that people, I think, can focus on and better understand.

And but I think eventually, you know, we do certainly need to collect or to stratify all the measures by race-ethnicity to see what's going on. Maybe there are some that are worth focusing on. I think that's probably your decision.

DR. BURSTIN: If I could just briefly respond to it as well. That's actually part of the reason we've decided it was time to do this project since we had this old set of criteria that picked some and left some on the table. And the real question, one of my questions going into this is does this need to be anything? It should just be routine that NQF would request that all measures be stratified.

I do think the sentinel concept is intriguing because at least then you say, okay, we recognize there's a universe of measures out there, but we know these are areas with known disparities. If nothing else, publicly report these. I mean I think there's some strategies there that probably do get closer to the measure selection piece that, you know, the Measures Application Partnership is trying to think through.

But I'd also be curious to hear Ernest's comments as you think about it. I

mean this was a lot of discussions about the NHQR versus NHDR, just saying let's just take the same quality measures and stratify them rather than coming up with a different set. But -- welcome by the way.

CO-CHAIR ANDRULIS: I'm going to take the prerogative of the co-chair and suggest that we just take a break. We've obviously got a number of questions here. We can come back to the questions after the break because there are five or six more flags that are up and this is going to go on for a while. And Joe's raising his hand, too.

DR. BETANCOURT: I just wanted to say Emilio, one of our co-authors is on the phone. I just wanted to let people know that and Emilio, I don't know if you just wanted to say hello to the group? He just shot me a note as well.

I know he's on and maybe he's on mute. Okay. Well, he's on the line so just, you know, he'll be on to answer questions as

well later with us.

DR. WEISSMAN: Is it possible he --

DR. BETANCOURT: No, he's on. So anyway, we can -- he'll be involved later on for the Q&A.

DR. CORA-BRAMBLE: Joel, you had a question?

DR. WEISSMAN: No, no, I was just wondering if -- I mean if he's on can we not hear him if he's trying to say something? It would be great if he could just sort of chime in even for two minutes? No? Emilio, are you there?

DR. BETANCOURT: Yes, he's there. He just texted me. So all right, well, we can -- I mean the Q&A we can -- just he's available. I just wanted to mention it.

CO-CHAIR ANDRULIS: Okay, before we break I just wanted to have Ernie, if you want to introduce yourself and just there's a point of disclosure also that was requested.

DR. MOY: Oh, okay, got to do the disclosure. Yes. I'm Ernie Moy. I'm with the Agency for Healthcare Research and Quality and I've been working on these various reports since the very first one, and obviously therefore very interested in how this all turns out.

CO-CHAIR ANDRULIS: Okay. We're on break for 10 minutes.

(Whereupon, the above-entitled matter went off the record at 10:43 a.m., and resumed at 11:08 a.m.)

CO-CHAIR ANDRULIS: We have a few folks on the phone that haven't had an opportunity to introduce themselves, and we have a new committee member that's entered as well.

So why don't we start with our present committee member to introduce himself to the group? And we'll also ask you for disclosure.

DR. MCCADE: I'm Bill McCade, and

I was actually on the call earlier until just now. But I didn't do disclosure because I wasn't on early enough for that.

My job is Deputy Provost for Research and Minority Issues at the University of Chicago, and I'm a Professor of Anesthesia and Critical Care there as well.

I'm a director for the ACGME, and an officer in our State Medical Society, and a member of the American Medical Association.

CO-CHAIR ANDRULIS: Thank you. Dawn, could you introduce yourself too?

MS. FITZGERALD: Yes, hi, this is Dawn Fitzgerald. I'm with Qsource, which is a Tennessee quality improvement organization.

CO-CHAIR ANDRULIS: Great. And Evelyn, could you also introduce yourself?

MS. CALVILLO: Evelyn Calvillo with California State University Los Angeles.

CO-CHAIR ANDRULIS: And I'm sorry, we're also asking for disclosures, any word on disclosures from you?

MS. CALVILLO: I have no disclosures.

CO-CHAIR ANDRULIS: And Dawn?

MS. FITZGERALD: I have no disclosures.

CO-CHAIR ANDRULIS: Okay, is that all, or am I covering everybody who's on the phone? Anybody else? Okay, very good.

So let's pick it up with Mary, and continue our conversation, discussion, questions.

DR. MARYLAND: Thanks so much. Let me first add my thank you to John and Joel, very comprehensive, very thought provoking.

I would like to ask that as we look at our work over the next day and a half, that we think about how to share some information very specifically.

There's lots of talent and best practices that are out there that we don't know about. Some kind of way, and maybe it's

a parking lot issue, how we share best practices and/or additional resources such as language and those nuances that might come up as we continue to work together.

I would also ask that we think about beyond our work now. What impact will this work have on patient outcomes, patient care, the next generations of providers?

How do we share it so that the next group in 10 or 15 years is not sitting here doing this?

Is there a way to incorporate looking at diversity, disparities in such a way that rather than being an add-on which always translates into more money, more work, more effort, who's going to pay for it, as part of what we do in the normal course of providing quality health care, which I believe is the ultimate goal.

And as long as we keep it in a separate bucket, we create a divisive that I believe doesn't get us where we're trying to

go. So I would ask that we think about how to put that lens on.

And finally I would ask, how we think about educating the future both patients, how do they come to visit a provider, and get the best when they leave on that visit.

I think of the opportunity where a person who did not have English as a primary language needed to be referred for mental health services, and then added and by the way, someone who speaks Spanish. And so now, what does it require to get that person to that service?

Similarly, if you go to a physician's office or provider's office as a nurse practitioner, I have to add that, when you come to the office for that first visit someone automatically asks you about your co-pay.

Someone should automatically ask what you need to maximize your visit in that

facility, then we start moving the needle away from thinking of only disparities.

CO-CHAIR ANDRULIS: One of the points that you raise about disparities, actually I think also has resonance broadly and fundamentally with what we're about here and that is, as I've had many conversations, the issue of culture, while it's race ethnic specific, where's the connection?

Everybody has culture, you know, there is a cultural connection, and there is a language connection to, and gradience.

And rather than seeing other than how it fits within a broad mainstream, so that also for providers organizations it becomes an important consideration. So Colette?

DR. EDWARDS: I wanted to add my accolades about the report and also piggyback on what Mary has said, Marshall has said, Mara and Joe, in terms of I think it would be fantastic and very important for some very simple things to come out of this committee.

From the standpoint of I want to piggyback on, there was a recommendation with regard to integration with NPP and MAP, that I think it's really important that we not add to how overwhelming this area can be, how confusing it can be, add to more measures when we haven't even done the basics yet.

And really do that integration, and try to be as simple and efficient as possible so that we can focus on getting some things done as opposed to being caught in the quagmire of the perfect measures or the endless list of measures.

And also recognize kind of in a real world perspective that it's going to need to be relatively simple and something that people can really act on.

And I think that the timing is crucial with the swirling that's going on with regard to medical home, and ACOs and the changing reimbursement that this is a huge opportunity to try to embed some of what we're

talking about.

And really embed it to Mary's point, because otherwise we're retrofitting after the fact with less impact, less efficiency, and we just are getting to be further behind.

DR. LU: Yes, this is Francis Lu. And again I want to reiterate a great big thanks for the tremendous report that you put together.

And I just have a fairly simple question perhaps and that is around the issue of understanding our patient population going forward.

I understand the focus on racial and ethnic and language, there are disparities related to race, ethnicity, and language in what you've said in terms of where the research has been.

But I'm just concerned that this kind of focus might inadvertently lead to a certain amount of generalization and

stereotyping, and will not give us the more detailed information that we need to address where the disparities might exist for specific populations.

And I'm sure you're well aware of this problem. So for example, amongst the Asian Americans, you know, the median income is quite high, education level is quite high when you look at that generally speaking.

But when you look at specific populations like Cambodian or Mon, that's where the disparities and income issues come about.

And I know you mentioned about how to, you know, that that's an important issue to think about in terms of how do we drill down into those sub-ethnic groups.

But it seems to me that there needs to be, we need to maybe strive in the next day and a half to reach some consensus on how to approach that issue.

Because I think if we don't, it's

going to be left to everybody doing it whatever way they want, and we'll never really capture that information it seems to me, if we don't speak about that specifically.

And then one more thing is the issue of capturing data in addition to race, ethnicity, and language, but the other measures such as age and gender and geography, I think have been, you know, the AHRQ reports have spoken about disparities related to those cultural identity variables, let alone years in U.S., sexual orientation.

Are things that again I wonder if maybe NQF already has made decisions on this that I'm not aware of.

But in terms collecting that kind of information routinely, so that we can analyze and stratify and subdivide the information to get more specific information that might be critically important to understand the specific problems we're dealing with.

So it's just a general comment that, you know, maybe for further discussion.

MR. EPSTEIN: For our growing perspective discussions parking lot, am I correct in understanding that cultural competency issues will be explored further into our steering committee work? That's a question.

Additionally, I support one of Dennis's opening questions concerning health literacy which has not yet been addressed.

Also, given that the public comment period for the proposed HHS data collections standards is currently taking place, and especially since people with disabilities are apparently going to be considered first as I understand it in LGBT issues, will be potentially be considered at a later time. At least the LGBT work is going to be more pilot as I understand it.

I think these are important opportunities that need to be considered at

some time. For the present, does the mass general team have any comments to offer?

DR. BETANCOURT: I think just particularly on the last one. I mean I think there's a lot of opportunity for identifying and adjusting disparities in other groups and I think we've touched on that.

I mean, I know I can comment personally that we're doing a lot of work on the issue of disability, but it's been a huge learning curve for us.

You know, there's such a spectrum of disability. Some people don't want to be called disabled. There's a lot of knowledge acquisition that I think needs to happen around each one of these different issues.

And I think, you know, I think Mara mentioned that we've come a long way and I think it's because we've given it a lot of attention.

And I don't see any of these as more or less important, I just see them as

where we are in terms of stage of evolution of knowledge.

It's not to say that we don't have good knowledge, but my sense of it is as you talk to people that we need to do more to figure out what are the measures? How do we ask these questions?

Some of the work that we've already done around race ethnicity, but I think that that's all, you know, should be for the committee's consideration.

DR. MOY: Okay, great report. In particular, I appreciate the flexibility I think that was built into the appreciation of different kind of methods. And I think related to that, I was going to put two things perhaps onto the parking lot.

One is the notion of I think it's important to consider the purpose of the measurement, because different methodologies are probably appropriate for different purposes.

I think in the disparaties report, we've kind of evolved what we think is reasonable methodologies for what we do, which is reporting at the national level across different populations.

But we're often asked, well can we do this for quality improvement or pay-for-performance and just take this? And we'd say no, don't do that.

So I do think that that needs to be something that we consider as we discuss measurement issues related to disparities.

And the other one is an issue that we grapple with all the time, which is developing methodologies for tracking disparities' changes over time.

And I think that also is something that's important because we often will see, you know, larger disparities that are getting smaller quickly.

And maybe you don't want to pay so much attention to those as opposed to places

where there are few disparities, but they look like they're starting to pop up. Just the temporal trends and other methodological issue to throw in there.

MS. TING: Oh, thank you for a great report, Grace Ting. And I think I really wanted to echo Mary's comment of we're really trying to institutionalize.

And to that end I think that a lot of the proposed measures that you discussed and examples had identified the different root causes of health disparities.

And so I think one of the viewpoint that's missing that Leonard, you know, brought up is that I would like to see some discussion and potential measures surrounding the systems class capability as a whole. And do how many you certify interpreters and translators do we have, you know, throughout the country? Are people regularly assessing reading level?

Again, I don't know that these

measures exist, but I think that if we're able to look at the industry and the system structure as a whole, we can be able to identify, you know, is that some part we need to address some attention in addressing the clinical measures? That's one thing.

And then the other thing I would like to encourage everyone to continue to think about, is to try to select measures that are going to be as cost cutting to as many stakeholders as possible.

So from the health plan perspectives, many of the ambulatory care measures are really valid as well, but how we get the information is really dependent on how well people code in terms of diagnosis and claims.

So what I have found in recent experience through our grant project is that the medical groups that we're working with have a much better sense of who their patients are and what clinical conditions they're

experiencing. But by the time we're looking at the administrative data, we're much less able to report accurately.

So if we could include that as a discussion that as people across the spectrum are measuring this that we can also look at the linkages of how data's transmitting between us and the standardization of it, I think that would be very helpful. Thank you.

DR. CLARK: So I would like to join everyone in congratulating you. I thought it was a great report and really enjoyed reading it.

The question I have beyond, you know, what to measure and how to measure. I'm just wondering, and maybe, you know, for the MGH group and also for NQF, as part of this project is there an aspect of it which would be a measurement of the impact of the what to measure and how to measure that will be included? And that will be in some time frame, that will be visible to those of us?

Because I think as important as deciding what to measure and how to measure, the bigger question is, so what will be the impact of doing that on all of the issues that have been discussed here early? And is that nested in here somewhere?

DR. BURSTIN: It's actually a great question. There are several ways we could do that. We would certainly be able to keep an eye prospectively on which measures when they're submitted for maintenance, which is a requirement for all measures come back to NQF in three years.

We'd be able to see how many of them, in fact, provide data if that's one of your recommendations, that are stratified. See if there are disparities, and as they keep coming back see if there is a reduction in those disparities. I think that's the first point.

The overall impact of having NQF endorse measures is a broader topic. We've got

an evaluation being done right now that's beginning to track that. We can track it because they tend to be used in federal programs and things like that.

We don't often know, for example, Grace, how many of those measures are picked up in health plans, things like that.

But I think the more we can identify which measures you guys think are the sentinel measures, or just broadly how often are we, in fact, getting measures stratified with the results available at maintenance, I think that would be a first good step. But other suggestions are very welcome.

DR. WEISSMAN: I also just wanted to comment on Grace's comment that in the report I think we didn't pay as much attention as perhaps we should have to system or structural measures.

And I think particularly as somebody mentioned the, you know, the idea of medical homes, and that's a great example of

an evolving concept where some of the early measures are really much more structural than process or patient experience.

And as it's evolving and we're getting we're experience with it that field is moving more towards the process outcome and patient experience measures.

And I think you could draw a parallel to that with disparity sensitive measures that it may be, in fact, important to collect some structural characteristics of provider groups.

You know, Dennis has an instrument on this, other instruments exist about the cultural competency of the organization as a way to kind of start and just, almost as guidelines for the organization to adhere to in order to work on their disparities. So that might be something worth considering.

DR. JACOBS: My comments will be very brief. I just wanted to follow up on something that Ellen said earlier, which is

that there may be measures for disparities that actually aren't capturing disparities in all populations.

So the example she gave of like not having CAHPS and multiple languages, is one way where you're looking at, it's an existing measure but it's not really capturing everything you can look at, so thinking more about broadening your idea of there are some ways in which we're not capturing disparities.

It may not be that there's no measurement exists, but the measurement exists but it's not being applied to all populations.

And maybe when we look at these measures that there are measures out there that already exist and we're going to adapt them, maybe one of the processes we want to go to is say well, can this measure disparities for all populations that we're interested in measuring disparities for?

Maybe we need to make recommendations about how to expand them in

certain populations.

DR. HASNAIN-WYNIA: Thanks. I also want to thank Joel and Joe for the report which was very, very comprehensive, and all the co-authors as well. I wanted to really speak out in support of Collette's comments regarding simplicity.

And I think that it's going to be very important for this committee going forward, to really articulate how, what the work we're doing here is different than what was done vis-a-vis NQF a few years ago.

And I think that's going to be important in terms of the policy environment right now, versus what it was three or four years ago, five years ago.

So simplicity is very important and I'm worried that we not come forward with another, you know, 700 measures or so.

I think the importance of disparity sensitive measures versus sentinel measures, all of that I think is a very

important conversation to be had, but I think at the end of the day we need to be thinking about the end users and who is going to be taking up these measures.

And, you know, one of my roles as the lead for the Aligning Forces for Quality Evaluation, which is looking at improving quality in 17 markets throughout the United States and equity is one of the key components. This is funded by the Robert Wood Johnson Foundation.

And I will tell you that the value statement to providers around collecting race, ethnicity data, language not so much, they're a little bit more clear about that.

But the collection of race and ethnicity data is one thing, but what they really need information on is about the utility of the data.

So what I'm really worried about is coming up with a list of many, many more measures with the end users throwing up their

arms saying, what are we doing?

Because we haven't done as you said, even the simple things that many of us around the table assume at this point in time should be taking place. There hasn't been a tidal wave of change. Let's face it, it is very incremental.

And I think we're still crawling, and we shouldn't be crawling, we should be walking at least, you know, Joe, to use your words.

So again, just to reiterate and reenforce kind of thinking about what we want to come out with at the end of the day and how we communicate that, and how we communicate how what we're doing here now differs from what was done five years ago.

CO-CHAIR ANDRULIS: Good, great. Thank you all for those wonderful comments. Very thoughtful and obviously, you know, that paper has been a source of great stimulation for us as a committee and without a doubt. So

thank you again, for all your work.

And from here we're going to go from the forest into the trees as we'll start to look at pieces of your report.

And if you turn to your agenda, you'll see that we're starting with Sections 3.a and 5.c, and there are questions that NQF staff have put to us with regard to the specific sections.

If you look at the, what has now become the new 1045 selection criteria on, should NQF focus on prevalence and quality gap, as the criteria to select disparity sensitive measures and assume that other more general criteria are necessary for all measures.

And then the paper recommendation around known disparities that do not exist a set of more specific objective criteria should be applied. And then the attended questions to that.

What I'd like to ask is if Joel

and Joe, and for that matter, Evelyn, could kind of just give us a quick kind of recap on these points. Just a few minutes of your time to take us back to these sections and your thinking.

But just to put us back in that place, and then we'll begin the round of issues and questions.

And we'll start with Marshall when the time comes, and Romana. Your flag's still up. We won't go to Romana. Marshall, I guess will start.

Marshall, I just want to have them start with a little synopsis. If you could bring us back to those sections.

DR. CHIN: Just this one point first that, you know, when you get very micro very quickly, and I think that like Romana and Collette and Grace and others raised a number of really important points that are fundamental in terms of the practicality and simplicity issue.

And I still don't have a great sense in terms of the overall charge and purpose that we can get very micro and we need to get micro eventually, but in terms of us thinking about not missing the big picture, I'm not sure we're there yet.

I was at this meeting, or conference call, last Friday, it was sort of a similar group looking at quality improvement and reimbursement and incentives and QI.

And towards the end of the meeting, Carolyn Clancy made this point that she basically said well, details are important, but guys, don't miss the big picture in terms of the recommendations.

And I'm wondering, you know, are we in danger of doing that? That we can come up with some nice micro recommendations, but in terms of this basic issue about the practicality, usability, the practical points, you know, Mary, Collette, Grace, and Romana, all sort of, you know, made these points.

I'm not sure exactly in my mind so the road map in terms of what the product is that we're looking for that gets us there.

So in other words, the very micro I guess, yes, is necessary, but I'm afraid that we're not necessarily, that doesn't necessarily jibe also within the big picture.

And so again maybe clarity from, you know, Helen's had this experience, thought a lot about it with AHRQ -- to the extent that you're divorcing sort of measure selection from then implementation, usability et al., in some ways that's kind of artificial.

And so I'm wondering what advice do you have to us in terms of thinking about this issue of making sure that what we do isn't just a sort of a nice stand-alone product, but it's something that's going to be -- it'll be used and be able to make an impact?

DR. BURSTIN: I'll start and see if Nicole or Rob or anybody else wants to add

anything. I think those are good points and I don't want us to jump to the trees if you feel like we haven't set up the principles for the forest.

I mean, if that's important, and I did hear some really good comments. I know staff's taking notes, but it would be useful to have a quick synopsis of what are those key principles, that would be very useful.

I do think this committee was constructed in this way quite intentionally. There's a lot of end users here. There's a lot of implementers here. And there's a lot of smart people who've been thinking about the research aspects of it.

So I think our hope is that by bringing this group together, you would come up with some recommendations that are actionable. We can also make them actionable in terms of our criteria.

One of our criteria is usability, and two of our criteria are usability and

feasibility.

I think it would be useful perhaps to have this committee look through, think about it through the lens of the four criteria that we use to evaluate all measures. To think about, are there specific aspects here that are especially important to consider, importance to measure and report around the measure gap, for example, the evidence for the specific populations we've already talked about, you know, how often do you have the evidence, not have the evidence of known disparities, came up as one of the big issues.

We've talked about, you know, we haven't really touched very much at all on sort of reliability and validity, some of the issues Sean was just sort of teeing up for us earlier, of the measures around scientific acceptability, and then usability and feasibility keep coming up.

But I do think it would be useful to try to stay high and then dive deep when

you think we're ready.

CO-CHAIR ANDRULIS: Helen, I was wondering if it might be possible perhaps to describe what's happened with previous reports and the way they've been used, because I'm sure they ran into the same forest/tree issues that we're facing, and it's a variation on a theme.

DR. BURSTIN: Yes, it's a good point. I mean, I can give you one example. Palliative care, for example, several years ago, really still, really very, very few measures that were out there on palliative care, end of life care.

We did some initial work, developed a framework, actually Robin was still around for some of this early work, developed a set of palliative care practices. They actually became the basis of the Accreditation Standards the Joint Commission uses to accredit palliative care programs.

We're now at the point where a lot

of those measures have been developed based on those practices. They've just been submitted to us for a palliative care endorsement project.

So I think we tend to see that sort of trajectory there of what we hope some of this early thinking does. I also think -- and I told this example to a couple folks at the break.

I mean, recently in our Cardiovascular Committee, the chair of the Committee sent back any measure that didn't include the data stratified by disparities. It says it's required for maintenance.

It's on the form. It's blank, sent them back. Please send the measure back when you can actually provide these data. Amazingly enough, it took about a week and everybody submitted their data.

So I do think, again, we do have a role here we can play in terms of helping to be that pushing force, of saying this is now

the new rules.

And I think especially, Var and I were talking about this earlier, the policy environment has changed so much that I think there's now a whole lot more wind at our back than there was five years ago when the initial Disparities Committee met. But, I don't know.

CO-CHAIR ANDRULIS: I mean it also comes back to one of the questions that was on the conference call, and that is: is NQF asking the right questions?

And this is the forum for it, you know, there could be discussion around what are these mega issues that might be, and this larger world that might be considered in the context of moving this to an agenda that would be broadly applicable, or applicable to the audiences that you'd want it to resonate with?

So I think that's fair game. I think the issues around getting into the trees a bit, well, I think there's going to be a back and forth on it. You know, I think

you're going go into the trees and then the forest is going to pop up again.

So I don't see it as necessarily, now, oh, we're delving into such detail, such micro that we're going to lose scope of where we're headed more broadly. Instead, I think it might actually inform where we go more broadly, because it will probably raise certain questions, elevate certain questions to that level of, well, what does this mean in a broader context, you know, if that makes sense.

You look a little puzzled.

DR. CHIN: I guess the way I think about simplistically is thinking about three different areas. One is measure selection. The other is creating the incentive systems so they matter. And then third is designing systems in a way that they lead to the right things as opposed to unintended negative consequences.

And in some way they convey, this

discussion about sort of measures, if it's divorced from the latter two, in some ways it's dangerous, because you can have great measures but then terrible incentives and terrible systems.

But some of the issues about well, what makes a good measure and what makes sort of a good format in the simplicity, someone's having that discussion separate from the implementation and the practical use, to me doesn't make sense.

DR. EDWARDS: Does that then become like the example that was given for the Cardiovascular Committee, that those have to be taken into account or come back to us after it's done. Is that --

DR. CHIN: Well, see, like if the paper, Joe and Joel's paper, they sort of snuck this in there that, a good chunk of the paper was talking about these really critical implementation issues, because you have to, in some ways.

Otherwise, it's sort of like well, you know, we create measures or recommend measures, and then they can be used or misused, and that's where the action occurs in terms of whether they either are helpful or else not.

And so that's why to me it seems still a little bit divorced. And I think it's probably what, like Grace and Clint were talking about as, you know, like big health plans in terms of, what's going to be something that's going be, again, practically useful as they're thinking about it as implementers?

MS. TING: Right, so my experience is, I really like the National Health Disparities Report and I read it faithfully every year.

But, you know, I think that, I suppose, is a little too broad, and to say well, how can I take these even, even the state level trends and translate it in situ

into practice; I think that's the challenge.

And I think I would like to see this group and the measures we select, you know, definitely either bridge that or provide some examples, a la NCQA, saying okay, as we're accrediting you, here are some of the things that will qualify for this measures.

It needs to be a little bit more solid, I think, then these current measures.

CO-CHAIR ANDRULIS: Marshall, do you think it's a matter of a guidance for implementation that should be accompanying this?

Like what do you do with, should this move forward, the information, not that they're separated from it, there's a connection between the two.

DR. CHIN: Well, I think that's what Romana was sort of hinting at with her statement that we don't want to have a report that's a great report that then just sits on a shelf.

Not to say that prior reports just sit on the shelf, they've really done, I think, a nice job of helping push the field.

But as Clint says, we are in a great position where it is a different political environment where we can have more ambitious goals, in a sense.

And so I think sort of think about this at this time so that the product is more likely to be user-friendly, and there's a shorter distance between, you know, product on the shelf and then actual implementation.

So I guess it's sort of broad, but I think that probably does inform discussions.

For example, if the micro is such that we could spend the rest of the next day and a half talking about the micro and having really great discussions, but at the end of the day, don't have enough in terms of in the linkage to the, you know, for the implementation and real world practicality.

So I guess I'm just arguing that,

perhaps more so than the current agenda, that needs to be just brought up to the fore by you and Denise, to make sure that we do have that connection.

So at the end of, like tomorrow, we don't have necessarily just the answers to these questions, but answers to these questions but embed it within the context of the practicality, the simplicity, the usability, the implementation.

MS. TING: Yes. So I, well, I was going to say I definitely second that in that I feel that the incentive piece, and creating the rightness in this, will be so critical.

The 45 guidelines, for example, I'm pushing and pushing internally within my organization, but a lot of times, you know, their response is, we don't really have money. So unless it's mandated, you know.

CO-CHAIR ANDRULIS: Mara?

MS. YOUDELMAN: I think though, that it goes back, Marshall, to your first

question was, is the overall recommendation of this group going to be that you have to do stratification for all 700 quality measures that exist?

And if it is, that's taking us in a very different direction than if we're saying, we're looking for specific measures that are addressing disparities.

And so I wonder if that's almost like the global question that we have to start with is, you know, and to some degree it is simplistic. Because, you know, you just have to do this for every single quality measure that's out there.

And then we have to figure out what that actually means, and then we go into the weeds a little bit. But, I mean, we were talking, Helen and Romana at the break, are we at that point where we can do that?

And I think, yes. I mean, I think we've got enough of the basis in the law, and the policy, and in the backbone of NQF, and

what you said with the cardiovascular folks of, you know, if you say it has to happen, they're going to find a way to have to happen.

So is that sort of that preliminary discussion that we have to have? Are we really talking about, this is going to become universal, that's our recommendation, and now we're giving you the road map how to do it?

Or are we really putting out a call for specific measures and just sort of taking that little pieces, here we go from what we did five years ago?

CO-CHAIR ANDRULIS: Sean?

DR. O'BRIEN: Yes. I was going to throw out just one tentative idea for ways to narrow and focus the scope of the steering committee.

I mean, it's probably important to think about broadening the scope at the same time, but in my mind I think you can distinguish this looking at population level

disparities or the folks as looking across geographic areas and across time, and just kind of looking to identify potential disparities.

That's a little bit different from looking across providers and different other levels, other units of health care providers where there may be raising more issues related to attribution.

That it seems like that would be a more specific area where you could actually do something that's different from what other groups are doing to basically what issues arise and how and we're actually not just looking at measuring disparities, but actually comparing across units.

And sample size issues that people have mentioned are going to be, you know, really come into focus.

Issues of case mix adjustment and stratification, those issues are going to come into focus, because all the issues of

interpretation and attribution are really going to be brought into play there in a way they wouldn't if you were just doing a population-level focus. And I think there'll be plenty of areas to actually make recommendations.

MS. MCELVEEN: Two things. First, you've obviously given us a lot to think about and consider. And so we were first reviewing the paper to see how much of the paper addresses the issues that you've raised.

Secondly, my other question was, do you feel that, as a group, that you really are looking to make recommendations more on a system and structure level?

And then based on those recommendations, talk about how the measures would be addressed?

So, in other words, providing recommendations on how a system, or broader recommendations on how the system should be set up, or how it should look for implementing

the measures properly. Is that kind of the sense of what the group is saying or -- does that make sense?

CO-CHAIR ANDRULIS: Go ahead, Romana.

DR. HASNAIN-WYNIA: I think that's one approach. I do think that -- you know, I'm in fear of contradicting myself.

So I'll use Mara's example of, so if we're thinking about, you know, kind of focusing on the systems level, I think that that would be good on the one hand.

But it shouldn't be at the expense of completely not having a conversation about a measure that should be, you know, at least discussed and voted on, vetted through this committee.

So, you know, currently as Mara pointed out in the new regs, we have English proficiency but nothing about language spoken.

So it's really hard to kind of, from a system perspective, target the

resources needed to address disparities based on language if we only have proficiency information.

So I think that, you know, one goal of this committee could be to address that gap in the current regs through the measures that we discuss.

So, you know, that's an example. So I do think we need to strike a balance, and I also kind of share Marshall's, I don't know if it's a concern or just his point, that I think a little bit more guidance in terms of how we proceed would, I think, benefit the entire committee in terms of making the best of our time around a larger framing, around, you know, where are we going to spend the next two hours.

And before we really, you know, do we need to really dig down into the weeds right now or do we need to have that framing discussion right now?

And make that a very explicit

goal. And I'm not quite sure what the process is, Helen, and maybe you can speak to that, but maybe that's something that the committee needs to come to an agreement upon before we actually do get into the weeds.

CO-CHAIR ANDRULIS: Okay. Well, it seems to me there are kind of two, putting it simplistically, two paths going here.

One is around, almost sounds like you're talking about guidance for the group for the next period regarding the work ahead of us.

But also coming back to this guidance of what this means for the field, you know, and how you help them, so it's almost a two-part.

But I think, you know, correct me if I'm wrong, Marshall, but it seems like there are three levels that we're talking about in terms of addressing these measures.

One is kind of a level of sentinel measures or other measures. The other one,

the 700 measures.

The third is application of more broadly, coming back to the cardiovascular example of, well, it doesn't matter, it should be considered in whatever your view, you know, beyond the 700 measures.

And so, I think a point of consideration both for NQF and the group is: well, which way are we going on this? Are we looking to apply a broad-based, broadly applicable approach or are we saying well, let's go with, you know, go with something that's --

DR. CHIN: Well, there's also maybe a practical answer to that, so if you start with the 700 because of the sample size issues that Sean, and I think Donna, had raised originally, you rapidly come down to the very small number anyway.

So in practice, you're dealing with a very small number anyway. But there are the existing measures, the measures that

exist that are really disparity-specific in terms of, you know, ACE inhibitors, interpreter services, you can say, well, you know, that's a very disparity-specific measure that exists right now.

And then you have, you know, the field of, well, disparity measures that we want to have, but they don't exist now that we need to find. So, I mean, that's what it comes out to practically.

You know, a small set of the 700, which, you know, I guess it's going to be determined by the sample size alone. The things that we all know in terms of disparity specific things.

And then the ideal, that, you know, we need to have a call for proposals for in terms of developing important measures that don't exist now.

DR. BURSTIN: I guess my sense of it is, we were going to actually do all three. And I don't see of us doing one or the other.

I think there's, you know, I think we very much want to have a sense, and if you look through these questions they get kind of deep, but they're also pretty broad.

Should we just say, you know, one of the requirements could be, based on sample size, can you provide stratified data by three-year maintenance? It could be as simple as saying, yes, on all measures. That's part of the determination we'd like to hear from you.

But I think, you know, one question might be as we go through this, maybe as we work through each of these issues we could talk about sort of the inside the measurement space, how it applies, and maybe a little bit about sort of outside the measurement space, the implementation, the implications of doing X and what needs to happen.

That might be one way to do it, but I think our hope very much is you're going

to help us write this call for measures on the cross-cutting measures.

And so as we're going through this process, we hope you'll come up with what we hope are the measurement concepts that we are going to want to bring in, what are that, what is, what are, what is that small set of cross-cutting measures that you think would really add value?

Because currently we really have almost none within the portfolio. And the question is, what would be useful out there?

And we've heard, certainly, that there's great interest in picking up some of those measures if they existed, in terms of picking them up from limitation.

CO-CHAIR ANDRULIS: Okay, so would you like to proceed?

MS. NISHIMI: So, let me just try and reflect on what I heard here, and recap on how we might proceed, really teeing off of what Helen just said, and see if that's

amenable to the group.

I heard that there is a desire to look at some big picture issues first, some of the system's approaches, you know, the general direction the committee thinks perhaps implementation should take and the like.

There's also the notion though that drilling down to some of these questions, which is what Helen just proposed, perhaps should be looked at through the different lenses so that we look at these questions now as they've been laid out in this agenda, section by section.

And we first discussed what the committee's thoughts are about the big picture, any implementation concerns surrounding that and the like there. And then we perhaps narrow it down to specific applications.

So when you're talking about population level reporting, when you're talking about sentinel, and by population

reporting I'm referring to the existing NQF population of metrics, I'm sorry, mixed things up, you know, inappropriate use of the word population there. That we look at in terms of the notion of sentinel measures and that we look at it in terms of where we would like things to go.

So that we systematically look at these questions but through two different lens. And that will then allow us to march through the different sections rather than have sort of a large free floating thing up in front.

And then frankly my concerns will be we will then start repeating ourselves when we got to the questions again. So does that approach seem fair?

Yes, thanks Helen. Then why don't we just proceed that way, Dennis?

CO-CHAIR ANDRULIS: So would it be okay, Joe, Joel, if you could take us back to these sections and just give us a little just

brief review?

DR. WEISSMAN: Sure. I mean just the next slide I think would be enough, because you already know about this one. Oh, sorry, the one after this, yes.

We were also trying to think operationally about what the charge of the group was, and we were trying to think of the different measures in different buckets.

And so we kind of thought of the measures that currently exist. And I think in previous conversations there are some measures that are already stratified by race and ethnicity, and there either are or are not disparities evident in those.

And, you know, at least our suggestion that the, at the outset was to take those as disparity measures.

Now I think in light of some of the conversation about whether or not there should be a relatively focus group or whether you should go through everything, is something

that you all are probably going to have to decide.

But then I think the other criteria that we mentioned that are worth considering are, there may be some focused measures that have clear evidence of disparities that you really want to put in as a, basically to profile an organization.

But then we should think about other kinds of scenarios, other situations where, that are worth exploring and that we may not be looking at, at this point.

And these other criteria come into play, you know, high degree of discretion, and I would put in kind of evolving technologies.

I mean, you know, I think the literature is replete with examples of disparities of evolving technologies where there's a lot of discretion.

Some of the early work that I did with Arnie Epstein on coronary artery bypass

graft and PCI is a good example of that.

Communication sensitive services is we're finding more and more that those are the ones that persist as having disparities and so on.

So this is just a way of, you know, we can't answer how big a bread box to look at, but if you wanted to kind of think about the measures as being in different buckets with and without evidence of disparities, with or without measures stratified by race and ethnicity, and then where to look for other measures that might not be there especially as you entertain measure development recommendations that might, these might be helpful to look and think about.

Joe, you want to add anything?

MS. NISHIMI: Okay. So in terms of some of the global questions I heard the committee raise around this section, what I heard was, create simplicity about what should

be measured. And one of the solutions offered was to just require everything be, you know, looked at in the portfolio by race and ethnicity.

So that this notion of a subset if you will, of disparity sensitive measures has been overtaken by events, OBE, as we used to say at VA.

And so I think that's one of the big picture recommendations that the committee, you know, should opine on and it will be very helpful.

I mean as I think Marshall or someone said, you know, maybe we could get rid of part of our agenda in one fell swoop by just saying, you know, yes, we don't need to apply all these existing criteria. So some discussion for the staff will be useful.

DR. O'BRIEN: Well, this isn't exactly following up on that comment because I'm just looking at the slide.

And what I see there, the one

point that's up there that I feel is the most like problematic, not problematic in a bad way, but in terms of having issues surrounding it, is the recommendation to focus on outcomes rather than process measures.

I see those as obviously valuable, but they're the most, the ones that bring in the most issues related to risk adjustment. And I feel like there's probably, everyone in the room has different ideas about the issue of risk adjustment.

I just feel like it's a, so if we come up with a recommendation that yes, we think we should be in the area of disparities, we need to be focusing more on outcomes, have we really thought that one through or not?

When I've been on NQF activities at some of these outcomes where the issues of attribution are unclear that lead to the most debate and protracted discussion and like uncertainty about what the measure's trying to do.

DR. WEISSMAN: I have just one quick comment on that, and that is that I think that the issue with outcomes often comes back to the accountability and responsibility issue about, is a provider accountable or responsible for outcomes once the patient leaves their door?

And I think that discussion is evolving. And a great example of that is hospital readmissions where, you know, hospitals have said in the past, we're not responsible for readmissions.

Well, I think the world is changing. And that there is a certain responsibility.

We haven't figured out exactly how that's going to be accomplished, but there is a certain responsibility to reach out in the community, to engage the patient in ways that reduce those sorts of adverse outcomes.

So I agree with you, and at the same time I think that we have to, I at least

see an evolving discussion around those issues.

DR. BETANCOURT: And I would just add too, I mean we tried to, I think for some this works better than for others.

So we gave the example of flu shot, right? So you could look at flu shot by race and ethnicity and see no disparities in offering a flu shot to the patient. But then you look at actual receipt of a flu shot by a patient, patient getting the flu shot.

And unless you identify, you know, people can very easily check the box, mammogram offered, I mean this is the major kind of hub of the work around cultural competency, is that we can say, yes, I offered this but the patient refused. That's the very easy way out.

If you don't look at the receipt of services, you'll never know that there's a problem and you won't figure out ways to solve it, I believe.

So I think for some there's risk adjustment required, they're more complex. For others, I think they become ways for us to identify areas for innovation, for more work, for interpreters, for language, and for information in a low level health literacy, for assuring that an interpreter is present.

I think there's other systematic things you could learn about that. So there's probably several of them. We give as an example, flu shot, that lends itself well to outcome measurement as a way of driving improvement.

And I think we've done that around diabetes, where we just took some very big slices.

We didn't get caught up in who's attribution it was, we just know that we need to better with like, you know, diabetics, because we're doing less well for them.

Now then you start to experiment and figure out what you need to do and it runs

the gamut between public health issues and medication inherent, so that's my perspective on it.

DR. BURSTIN: I'll just make one comment, that NQF already updated our evaluation criteria this past year, did a whole task force on evidence guidance and specifically recommended only measures that are outcomes or a process measure so they clear a link to outcomes.

And so we've already said, distal measures too far from the outcome, won't get through regardless of whether it's disparities or not.

So I think one caution for us going forward is I don't think we need to spend a whole lot of time on issues, that is sort of across all measures. A good measure is a good measure regardless of the purpose. We wouldn't want a measure that's so distal process wise. But it's a good point.

CO-CHAIR ANDRULIS: And our work

on looking at organizations, what they were doing process structure wise, that's the same question that came up over and over again, how can you link this to outcomes?

If we have the presence of X, you know, a procedure of some sort, what does it mean in the context of outcomes?

And that was what was missing, organizations were looking at it only from the presence of the process or the structure, not looking at it as the link to outcome.

And so we'll move over to Elizabeth and then Romana.

DR. JACOBS: We've been having a discussion here so you'll probably hear both our thoughts in this conversation.

So we were just wondering about the 700 measures that are already out there. Would it be worthwhile then to think about which of those would be disparity sensitive measures? I mean I think that's what you proposed.

But one of the ways in thinking about making this simplistic is making sure that we're not asking people, sometimes it's really hard as we talk about sample size issues and language issues as sometimes it's hard to stratify across all those measures.

And do we want to simplify it by actually making it easier on people, in addition to saying well, of those 700 measures maybe there are 10 or 20 that aren't captured there that we think are really important.

So I do think that there is, I agree with the simplicity, but I also don't want to miss important things that would be important to look at if we don't think outside those 700 measures. So I'll just stop.

DR. HASNAIN-WYNIA: I actually have a direct question, and if I missed this in the report and it was there, I'm sorry.

But of the 700 measures, so this is on page 19 of the report, of the 700 measures of quality of care for both

ambulatory and institution based, of those 700 how many in the last disparity specific committee were endorsed as disparity sensitive?

MS. MCELVEEN: Thirty-five.

DR. HASNAIN-WYNIA: Thirty-five?

DR. WEISSMAN: They were just ambulatory though, right?

DR. HASNAIN-WYNIA: They were just ambulatory though, right.

DR. YOUDELMAN: How many ambulatory, and what percentage is disparity sensitive?

DR. BURSTIN: Sort of a point at time estimate, which is why we wanted to revisit this. It's about four years ago, my guess is it probably at about 150 or 200, but just a guesstimate, but again, there's many more measures now.

DR. HASNAIN-WYNIA: So I also just wanted clarity in terms of the question that Robyn posed, which is, you know, one of the

most simple things that we could do is say, and I don't know if I'm interpreting the recommendation correctly or, you know, what we're supposed to deliberate.

But one recommendation being that we just say across all the 700 measures, we put a recommendation out there that says that we stratified. Period. So that's one.

Then the second layer is, that we want to make sure that we're not missing sentinel disparities measures that we, you know, that we absolutely should be focusing on, which we to date have not, basically brought to the table.

So that those are the two, I mean I'm very much oversimplifying right now, but those are kind of the two broad strokes that we're contemplating.

And then Liz offered a little bit more nuance in terms of going back to the 700, and we talked about the 35 or so ambulatory disparities sensitive measures.

We don't have inpatient, but Joel for example, just, you know, brought up the example of the readmission measure which is a really important one.

It's very timely. I mean there's a clear policy link to it. There's an incentive, I mean there's or a disincentive, or however you want to look at it.

So I guess what I'm asking for is clarity of the question, and did I frame what Robyn was asking us to contemplate, accurately?

MS. NISHIMI: From my perspective, yes.

CO-CHAIR ANDRULIS: Yes, getting back to the larger picture versus the detail. One of the I think attendant questions is, could you use some subset, sentinel indicators, or another group for guidance?

You know, these are more concrete, recognized measures that are more broadly accepted or understood.

And perhaps a charge that could be put on the table is, could you then build an agenda around making sure that there is a practical application that would then resonate, again come back to the field?

And field say, okay, now I understand how to use this, what importance it has, and perhaps how it has broader effect in terms of other measures.

DR. BETANCOURT: I just wanted to comment on this 700, you know, look at them all or not. Of course, I can't come up with a good example now, but I remember as we started to go through them there were some that just when you look at them, you just said this doesn't really have, and I don't have a good example, but I remember multiple saying, yes this, there's no there there.

But it may be worth maybe at some point for the committee to get a snapshot of some of these. We don't have them accessible here.

Because you might be able to more clearly see whether, you know, which path is worth going at, because you might say, well these, I also think the actionable, feasible lens is worth thinking about as well as you mentioned.

CO-CHAIR ANDRULIS: Mara and then Romana, you're back on. No, Romana.

MS. YOUDELMAN: I guess another way of looking at this is instead of an opt-in, which is sort of the framework you've put, do we do an opt-out?

And I mean I'm sort of going back as an advocate, and I don't know the practicalities of how this works in the field, so, you know, I leave it to the rest of you for that reality check.

But maybe what we're really thinking of is we want to include as many as possible. And so we should be starting with that threshold of yes, all 700 should be in.

But maybe there is some subset.

And maybe it's, you know, I don't know how many, but that shouldn't be in for whatever rationale, but there's a really good rationale for why they shouldn't be in as opposed to trying to find ways to get, you know, put them in the bucket. It's really how do we take them out of the bucket.

And I think the basis for that at least in my thinking is probably based on, I'd rather push as far as we can go. We're at a good crossroads.

Too, what I hate about existing disparities research is it always seems to be focused on the same darn conditions because the standards and the basis is there, so we keep getting that data over and over again, and we don't really have, you know, we really don't it expanding as much I'd like it to.

And so I guess that's just possibly a different way of thinking of it. Is there any way to do it as, we all are in until there's a really good rationale for

taking them out?

CO-CHAIR ANDRULIS: Collette?

DR. EDWARDS: So I think kind of big picture, looking at all 700 is very appealing but from a practical standpoint it just is, just thinking about it and not even having to do it is overwhelming.

And looking at it from the health plan perspective and also looking at it from the hospital perspective, physician in the office perspective, it's just not going to work.

So I think having it out there as the goal, but then having either something staged that has some rationality with some statement of, we recognize ideally it should be 700, but we also recognize that it's not going to happen. And therefore, we've come to this subset and here's the reason.

And the four in that second bullet, the four sub-bullets are very appealing from the standpoint of, I think

they're very actionable, kind of one-on-one at the practitioner level, they resonate at the health plan level. And they're also tied to what's going on related to consumers, as well as medical home, ACOs.

It just really lines up with a lot of things that are kind of converging right now. So does that become a set of filters for some subset of the measures?

CO-CHAIR ANDRULIS: The guy whose tent fell over, Marshall.

DR. CHIN: This is much for Helen again in terms of, so in practice I don't think anyone's going to be doing 700 measures. And so there are other NQF initiatives like MAP where they're going to try to come up with a subset, which I'm gathering is going to be more than what are like 40-100.

So again, a system of discussion is already happening so that in so much as moot, that, you know, no one is going to say, well, do all 700. But there are wider

umbrella NQF initiatives that is already cutting it down for us in terms of for the general population.

So it gets back to Romana's point that, in some ways it may be we're saying well, do race, ethnicity and language on whatever generic subset are going to be done, but then don't forget about these disparity, sentinel or disparity specific things, which you haven't included but, you know, we as a committee think we have added value in terms of saying you should do these also.

DR. BURSTIN: What Marshall's referring to is that there is another effort in NQF now called the Measures Application Partnership, which is trying to at least make recommendations on the selection of measures for various pay-for-performance and/or public reporting programs.

They're struggling right now trying to figure out exactly what those criteria are. We've been working with them to

think that through.

And I don't think any of us are saying, therefore the result of this committee is to stratify all 700. I'm trying to think about more principles going forward. So as measures come back up to us for maintenance, you know, the principle is you must always just stay strong.

Every measure that comes back to NQF for maintenance must have data on the measures in use and the available date on disparities.

And then potentially you follow something like this to say, if you're going to publicly report or use these measures for pay-for-performance, you would want to stratify them.

I mean NQF endorsed measures aren't endorsed just for the purpose of QI. You can use any measure you want for QI, but if somebody's going to pull up a measure and use it, select if for payment in a renal

bundle payment program or use it for payment in a hospital program then boy, based on the data we've seen, this is one that should be considered stratified before you do your payment, things like that.

I think it takes it to a different level. I don't necessarily want this committee to just say, blanket all 700 disparity sensitive, I think that would feel sort of false and incredibly impractical.

DR. CHIN: I think it's a critical issue because here's where the lever is, I mean money. And so we're talking like either actual pay-for-performance, reimbursement or else public reporting, that's where there's power.

So that's where we really need to be careful in terms of making sure we get it right, terms of recommendations on the disparity measures because that's where it's going to have an impact.

DR. BURSTIN: Currently there's no

requirement that have the measures that CMS or others pick up for payment, that they stratify or look at any of that. And I think the timing is right, given the fact that those data are now increasingly available.

That could be a pretty important lever that we've never had before, which I think was part of the idea of bringing this to you again five years later.

MS. YOUDELMAN: I agree. I mean I wasn't expecting there to be like a trigger date, and all of a sudden on X date you have to, you know, do it.

I do like the idea of, you know, as things come up for review, and I guess I would also suggest, is it practical to say as a new standard is coming through, you need to evaluate whether you can do it from the get-go.

Because at least in other data collection it's easier to do it from the get-go, rather than go back a couple years later

and fix it. So that might be another principle, is one, as existing standards are coming up for renewal, what's the lens on which they are evaluated?

And then two, as a new standard is being developed, can you make sure that you're developing it in such a way that you are addressing the disparities issues?

CO-CHAIR ANDRULIS: Yes, and Collette, coming back to your point. It doesn't necessarily preclude using those four points there as guidance for health plans or others, it's just we're thinking a broader scheme. But in terms of interpretation and guidance for application, then that could easily apply.

Norman?

DR. OTSUKA: I'd like to say that orthopedic surgeons, or our society has been at the forefront in diversity, but I must say that I disagree with you, Helen.

I don't think the wind is

necessarily behind our sails, at least for orthopedic surgeons.

Because we do a needs assessment and we did one, I forget, five or eight years ago, and culturally competent care was low, reimbursement, you know, other advocate issues were high up.

And we redid that needs assessment recently, two or three years ago, and still culturally competent care, although we set up a diversity board and have done teaching, have done education, marketing.

So my point to you is that, I mean I don't want to keep the measures too plain or, I mean I agree with simplicity but they have to be good, you know, garbage data in, garbage data out.

So maybe one of the motives of our measures would be to raise awareness of culturally competent care and health care disparities amongst our grass roots surgeons and physicians. And I think that would be

another motive for some measures that we do.

MS. NISHIMI: Well, I will say that I still haven't heard clarity yet on which way the committee feels. And so I don't know if you want to literally run the table or not. I just, I don't think that we have clear guidance on how to handle this.

Or if we don't want all 723 in there, what are our specific exclusionary criteria if you will going to be? Are we going to go with these four, some other add on to this four or something? I think we just need to be a little bit more crisp.

CO-CHAIR ANDRULIS: You know, I think, again I open it up to the table too, but my interpretation is we're talking about, do we come up with something that's more, as I said more broadly applicable?

Or do we select out certain subset and use that as a starting point for organizations, for plans that would recognize these as indicators that, okay, we know these,

we can use these to build on for guidance?

I feel, I mean I think the question on the table is, do we want to go broadly, or do we want to go more narrowly? Or do we want to do some hybrid of some sort that would provide the potential to consider, and parameters for considering other measures while you set up a design that considers specific measures to begin with? Does that capture it, Robyn?

MS. NISHIMI: Yes, I think that gets most of it.

CO-CHAIR ANDRULIS: Luther?

DR. CLARK: Yes, I think I agree with what you just said. My sort of position listening to this is that we'd want to be more inclusive.

And while 700 measures are a lot of measures, I mean there are 700 measures. So that there should be a reason not to include them rather than looking for reasons to include them.

But at the same time, it's important to really, you know, highlight those that are particularly important or valuable.

So I think we can do both so that whoever created the 700 measures, there are the 700, so you don't want to give automatic outs, but realizing that it's not going to be valuable for all of these.

And I was thinking in terms of the potential maybe to even detect disparities as to where they might not be apparent. And along those lines, I wanted a clarification. I wanted a point in the slide.

So the, when you speak to data exists and show no disparities, is that data, it does not show disparities or it shows that there are no disparities?

DR. WEISSMAN: No data exists on disparities, is that that statement?

DR. CLARK: Yes, but that's, I'm saying that you may not show disparities that exist because they weren't detected.

DR. WEISSMAN: Right.

DR. CLARK: And that's where, you know, my sort of tilt for being more inclusive, because it is possible that you, there's no reason to suspect or think there are disparities but, in fact, they do exist for reasons that may not be apparent.

DR. WEISSMAN: Yes, I think you can go a little, and we were driven a little crazy about thinking of all the different possibilities.

PARTICIPANT: Go to the mike.

DR. WEISSMAN: Oh, I'm sorry. That there probably are a lot of different possibilities for how you look at this.

But there are, you know, starting from the simplest, there are some measures for which are already stratified and declare data, and you start from there.

And then after that you've got a lot of different scenarios. And I think what we were pointing out here is that there may be

some situations where either the measure has not been stratified by race and ethnicity, or maybe it has and it's not showing disparities.

And yet there's a suspicion that actual disparities should miss, maybe exist that for some reason or another this measure is missing something that one might suspect based on either these other criteria or based on other literature that something exists.

So I think there are a lot of different ways to go beyond that. I'm sorry if that wasn't as clear as it could be.

DR. MCCADE: I actually want to echo what Luther just said. My suspicion is that of the 700 measures, not all are going to be equally important and have equal impact on outcomes.

And I think if you want to try to find something that is going to be relevant, as Romana just spoke about and what Marshall was going towards before, I think what the practicing physician wants to see is that when

I do something, it actually has an impact on the care that I provide.

And so maybe that should be a stratification criteria of the 700. The ability to actually have an impact in the measurable fashion that physicians can even see, and perhaps be encouraged to continue culturally based thinking in terms of what they do in their practices. And I think that may be one approach to do this.

And I think the idea of creating sentinel measures is also important, and although that's not what we're focusing on at this second, it's on the slide so I thought I'd go there.

When I first read it I wasn't quite sure as to what was called for at that point, but now I think I'm developing a better understanding for it after hearing a bit of discussion and thinking about it a little bit more deeply.

And that there are some things

that potentially the 700 measures don't account for. And I think this committee has a lot of expertise and could potentially be very useful in trying to identify those particular areas.

And as the call comes along to create a request for these additional measures, we could be very instrumental in helping to identify specific things that could be important in that regard, both impact as well as visibility and measure in a responsible way.

DR. WASHINGTON: So going back to what Marshall said about thinking about how these measures are used. I think that part of the challenge is that they're used many different ways. And that different health care systems are at very different stages in both their gathering and reporting of the data and how they use it.

I mean despite sort of the best efforts to admonish organizations on how to

use the status, some may potentially misuse it and sort of target resources, since most organizations don't have new resources, then addressing disparities which is obviously a goal, might come at the cost of shifting resources from some other area.

And so they'll make the disparities maybe for the targeted measures go away, but if you don't continue to then monitor them as they then shift to other things, those disparities will come back.

So I think it's important in thinking about the point about measures that show no disparities, that though they show no disparities issues, disparities may recur in future years.

And so I would not necessarily solely use that as a criteria for including versus not including stratification for a particular measure.

The other thing is I really want to reiterate that I think there's a very

important role for adding measures, particularly the four bullets that are up on the slide, and thinking about organizations that have gotten it that they have to measure report on disparities. Then the next challenge is how to translate the gap in reducing process measures in reducing the disparities and outcome measures.

And I think some of those missing steps, some of the missing links have to do with communication, and have to do with some of the other things that we're not currently measuring. So I would encourage us to move forward with encouraging some of those new measures.

MS. CUELLAR: I just wanted to mention when you're looking at measurement as well, the government has put out the carrot out there for electronic medical records, and one of the things we might consider is partnering in as this moves forward.

Because from assessment tools to

providing education, to providing things that are health literate appropriate, that's a potential route for this and there is a lot of grants out there.

A lot of the health organizations are taking advantage of this carrot that has been put out there for monies to become more, as a country going towards the total electronic medical record, and starting from admission all the way to the continuum of care.

I think that's a great avenue that perhaps partnering with them, might provide not only an avenue but of capturing the data as well.

CO-CHAIR ANDRULIS: Elizabeth, and then Marshall?

DR. JACOBS: So I thought Marshall was next. I probably don't need to say this but there's a lot of ideas going around here.

I really want to act on something I think either Mary or Collette said about,

we don't want to make people more negative about thinking about these issues around providing equitable care, and we can do that if what we ask for is so burdensome.

So I keep hearing 700 measures and I'm going, oh my god, how are we going to narrow that down? Because it's just, we have to somehow figure out a way to bring it down to a way that it would be more user friendly to people.

So it wouldn't be necessarily perceived as more of a burden and again kind of marginalized disparities again as we've talked about, or make them viewed more negatively or continue to perpetuate that view. That is not something we should care about, or it's just another thing we have to report. We don't, it's a burden.

DR. CHIN: So as of a point of information in the comment that maybe Helen, and maybe Joe, can comment upon, you know, so this Measures Application Partnership that NQF

is working on was charged by Congress as part of an ACA to recommend performance measures for reimbursement and public reporting.

And so that, you know, one of the first things it's designed to do is to basically create common measures across like all the federal agencies, so like HRSA and NCMS, et cetera.

Which you know if that does work, you know, will be like a 800 pound elephant in the room in terms of, you know, getting everyone on the same page, and the private partners, private insurers would follow also in terms of everyone would love to have some uniform data set.

So if it does take off, I mean it will be very powerful. And Liz like in the subgroup, on the commission subgroup, it's a very similar list of criteria that people have been throwing around that, it looks like it's slide two or three, it's the last one Joel had, you know, like prevalence and

actionability and morbidity, et cetera.

These are exactly the same, you know, terms that they've been throwing around now as like, at least in our subgroup trying to come up with an, you know, criteria for thinking about the measures that can be chosen.

So again this issue of trying not to like duplicate and reinvent the wheel. I, you know, like Helen and Joe's on the steering, overall coordinate committee for this MAP, if it does seem like this was going to work in terms of there be consensus, and all these big players at the table in terms of, you know, being buy-in for this.

In some ways it would be nice to be able to wait until, you know, there was that list of 45 measures or 100 measures as opposed to being up to 700.

My guess is that a lot of the measures that we'll be coming up with, really are going to the usual suspects. Because like

the last meeting we had, the usual suspects are the usual suspects for good reason. I mean if the prevalent conditions that cause a lot more morbidity, that existing measures, that are feasible to measure, my guess is then that a lot of them are going to be sort of versed in with the ones that Mary talked about.

But in terms of this committee, the big hole is going to be the things that aren't on the list of 700 right now.

So the things that do become in some ways particularly, you know, our venue that other people won't think about because it's just not as high priority, so mainly things like interpreter services, and limited health literacy measures.

Or some of the potentially, like the pathway things that, like you can get for example, a transplant, and there are some nice examples, there's a lot of pathway communication things that are on the list that

may not be coming to mind initially in terms of more broadly people thinking about performance measures.

So part of the question for Joe and for you Helen is, well, you know, how do we minimize duplicating the wheel that again, if it's going to be sort of big work by some ways, you know, bigger umbrella that, I've got to come up with the same measures at least for part of the subset.

You know, it would be a lot of work for us to duplicate what will be 80 percent of the same in terms of the existing measures versus us thinking about some of the things that would fall through the cracks by the other committees, not because they're bad committees but because that's not their priority, thinking about disparities.

So is it a mistake to say well, you know, let's not duplicate what's going to be the duplication anyway in terms of what they're going to come up with, and of course

think about the very specific things.

Or is that dangerous in terms of saying well, we can't rely upon the MAP coming up with the measures and issues that we would think would be important? And so that we should go through this big process of going through the 700 and thinking about, you know, what are the things that are directly relevant for us.

I mean I just don't want us to do a lot of work that's going to be basically wasted work potentially.

DR. BURSTIN: I'll start. Joe, if you want to add in anything, please do so. This MAP is new. We are just beginning this process. We've just literally done a draft set of criteria. You don't see a draft of course that they've been started, I think they're through a clinician group.

I still think, and again I don't have any great need to go back through each of the 700 measures and do this determination.

If that's not necessary, that's not necessary.

That's my first question to you guys. Do we need to do any of that? I mean that's the real issue here. We did it for ambulatory care. We didn't do it for anything else.

Having just that in ESRD committee sure feels like to me, ESRD is one of those places you'd want to say boy, if you get a look at these measures, make sure that the CROWNWeb, the system CMS uses for ERSD includes the data to be able to stratify.

So I'm happy to have you guys think this through however it makes the most sense in terms of the lens. It's a condition. Is it somebody's cross-cutting issues that have been raised?

I would not rely on the MAP to come up with the key ways we want to look at the measures that come to NQF for endorsement around disparities.

I still think that 's going to be,

they're relying on this process for the selection. It's still the next step in the process.

I do, however, think it's critical, and I know that's a big portion of tomorrow, that I definitely want this committee to tell us, what are those kind of sentinel cross-cutting disparities measures and cultural competency measures that we should bring in? That's not an either/or to me, that's an and, and we need to do both.

I still want to be able to hold people's feet to the fire to be very honest, when a measure's up for maintenance. What do they have to give to NQF? What do they have to show us that they've at least been thinking about the disparities issue?

Because if we don't, it won't come up. I mean if it wasn't for people sort of really pushing on the issue and insisting on it, I mean do you want NQF endorsing measures of low birth weight and not having somebody

raise the spectra of saying boy, if there was ever a measure you'd want to stratify, this is it.

It's just really a question for you as we go through our process, that's endorsement of measures is still the first step before selection of measures.

And I don't want it to be viewed as something that's an either/or. But you could at the same point say that as a principle whatever gets prioritized by the National Quality Strategy or by the, you know, Measures Application Partnership, should definitely be stratified.

I mean that's one approach, saying the measures that are prioritized should always be stratified, maybe that's one different approach to looking at this.

CO-CHAIR ANDRULIS: Joe, do you have any comments?

DR. BETANCOURT: No, I think that's exactly right. I mean I think the MAP

is, you know, trying to find its sea legs. And I stood on that. This has occupied a lot of my oxygen around measurement in the last, you know, couple months.

But I think now that this process is underway that we have at least a paper. I'd like to offer to present it to that group so that they can at least wrap their mind, I mean they're dealing with a lot of very, we're still at probably 50,000 feet, and this is probably down a bit.

So I'm committed to and I think certainly Helen as well, committed to keeping it in there. But I think it's going to more an integration as it goes, and I think we'll do a good job at trying to prevent reinvention. But it's going to be a very iterated process. That's a long road yet, a couple years perhaps.

DR. WEISSMAN: And I guess this is almost as much of a question as a comment, but in the same sense that there are 700 or so

measures that currently exist in NQF, and MAP is coming up with a subset of those.

So just because the 700 exist doesn't mean that an ACO or other organization has to select all those, but NQF is coming up with a recommended subset that are really important and that should be used in a lot of different situations.

And my understanding is that there is a parallel charge here. That if your organization is interested in either reporting on disparities, or incenting on disparities behavior, here is a subset of measures that you might think about using. Is that a good way to state it?

And almost by definition then, you're saying that the charge of this committee is really to come up with a subset that would be a focused subset that would represent the experience in disparities. It wouldn't be perfect. It wouldn't be everything, but it would be something to focus

on.

DR. BETANCOURT: You know, say our bringing up the 700 was more from the standpoint of, you know, it's a spectrum, right, for consideration for the committee.

On one end, you know, you're thorough. You go through all. You've got them all and you pick, you know, and there's a whole spectrum of different things I think you could do there.

I don't think it's, you know, our firm recommendation that you start and look at all the 700. I think the committee, it's up to you all to decide which way to go, but that's, you know, from the most thorough to a couple of different approaches in between is what's I think at your disposable.

CO-CHAIR ANDRULIS: Ellen? Well, I think that's the question we're going to come back to, you know, that's what we'll probably vote on in some way, shape or form, is your point. Yes, Ellen?

MS. WU: I just want to say I really appreciate Helen's approach to this and it feels, I mean, you know, 700 how ever many numbers of measures, I don't know if it's necessarily the point in that more so, whatever standards are established it somehow changes the culture and the way we do the work and how we look at it.

So it's not an added effort, right, to do this disparities as work. It's part of a health systems kind of every day process.

And I don't, you know, I think the way you articulate it in terms of new measures coming in or whatever, always having that lens on it is great.

So how ever we get there, through the measures or whatever other standards that are set, I think that's really where I would like to see, the kind of the eye on the prize.

You know, it could be two, but if it changes the system in a dramatic way, that

would be huge, right, in just how they do every day processes. And I don't know what that is, but I just --

CO-CHAIR ANDRULIS: Okay. I think Elizabeth, I think you had your, oh, and now it's down. Oh, thank you. Donna and then Romana?

(Off microphone discussion)

CO-CHAIR ANDRULIS: Pardon, and then break for lunch. And Ernie and that's it.

DR. WASHINGTON: So 700's a big number for us to sort of wrap our, at least for me there it's sort of wrap my mind around. And so I was looking again at the framework in this paper, which I think is really a useful starting point that thinking about structure process and outcome, sort of where the 700 measures fit in that framework. And also more importantly, thinking about the roots of the disparities.

So if we really want something

actionable, if the end result of this whole process is not just better measurement but taking that better measurement and using it to eliminate disparities, then maybe the approach is to think about which things do we have the greatest impact to change.

And so, for example, you guys list the roots of disparities as provider based, patient based, system based, or related to health insurance. We know a lot is happening with the Affordable Care Act.

I think we've agreed that some of the patient based social determinant, it's extremely important to stratify on those things, but perhaps it's the provider based and health care system based that the health care systems had to reporting this stuff out, have the greatest ability to change.

And so as a starting point I would recommend looking at those system based and provider based measures. And then also, a mix of both the condition specific and cross-

cutting.

The thing I like about the cross-cutting measures is their potential to perhaps influence multiple measures.

CO-CHAIR ANDRULIS: Where are we? Ernie, Ernie?

DR. MOY: Okay, I think I had two comments. One was hearing a lot of things around the table I think that were meant to try to protect the measure user, and the 700 comes to mind. But what are we going to do about that?

But I think many of your recommendations we're talking about are actually for the measurer suppliers, then the measurer developers.

I don't think anybody uses all 700 measures, and if they do they take out which ones they want to use, right.

And so I don't really have a problem with saying well, for the measure developers that they should consider

stratification of these measures. Generically, it's not a burden on the measure users, they still get to pick what they want to use.

I did like the notion that was raised, which was I think the notion of a starter set for disparities measurement. That would kind of hone things out.

But the more I thought about it I had some reservations, which are simple observations looking at disparities and it's variation, is that there are certainly some disparities that are highly prevalent across geographical locations, but they vary tremendously.

And there are simply some places where they have less of one disparity and more of another disparity. And so I hesitate to restrict the starter set because it may not truly capture all the disparities that are there.

I think maybe an alternative

approach is to specify criteria for an organization to think about when they're selecting disparities measurement for themselves.

Or even a generic statement, yes, users should consider the issue of disparities, which is a very general kind of comment. But that would be a recommendation I would favor it more.

CO-CHAIR ANDRULIS: It goes back to that guidance for how you select, and how you don't just go by the starter set. You know, choose and learn how to use it for your organization, and what measures would be relevant for your setting. Yes, very simple, and a good starter set for one place, would not be a good starter place for some other place.

DR. MOY: Yes, good starter set for some other place.

CO-CHAIR ANDRULIS: Absolutely. Romana, I think you, no? Marshall, are you

back up again?

DR. CHIN: Yes, just a question for like, who does what? And what does the work in this committee, be it versus NQF's fantastic staff versus no one doing the details and it's a more general recommendations.

In other words, is the goal at the end of the day to have then a list of, for example, let's say if it turns up there were 30,40 disparity measures that, you know, is the end product.

You know, which involves, for example, if we went through like MGH's recommendations, that's a really extensive process that is going to know, like well, is that the goal? And if that's so, then what does this committee do versus NQF staff?

So I guess it's about the charge. What is the end product and what specifically do the people at the table expect it to do, versus NQF's staff versus, you know, some

other entity?

MS. MCELVEEN: Sure. Well, the goal was, and Mass General proposed obviously this algorithm and measure selection. Taking that information and applying it to all 700 measures would not be a charge of the committee.

That would be something that NQF staff would do initially. If then we come to a resolution say okay, out of all 700 we found 50, we then would present that to the group and say okay, this is, you know, the output of our work, what do you think? That's all I have.

MS. NISHIMI: Right. So having said that, that's why we're focused on having you try and identify the criteria that you want us to filter these measures through.

We've heard support for the general categorizations scheme, you know, process, outcome, structure, provider, system based, you know, patient experience. So I

think that part is clear in our mind.

And maybe we could take the break for lunch now, and come back and hone in on some of those criteria Mass General has teed up.

You know, that there be a primary focus on prevalence and the quality gap that I heard, you know, some question about also. The degree to which it's actionable. Some of that's going to fall out again, the categorization scheme, but we need the committee's best input at this point, really, you know.

Robyn, Nicole, Kristen, Elisa, if you focus on prevalence and quality gap that's going to float the most important ones to the top. We can categorize them out when they come back, or if there are other things that you need us, filters that you need us to apply.

So if you can think on that notion, when we come back we can address that

and then start addressing some of the other micro issues that we need to get through today.

CO-CHAIR ANDRULIS: So we'll take a 30-second break for lunch. I think lunch is on outside.

MS. NISHIMI: Yes, so I think the notion was for, you know, folks to take a break, return whatever calls you need to do, get your lunch, and then come back in here.

CO-CHAIR ANDRULIS: Come back and we'll pick it up after a bit.

MS. NISHIMI: Because we will have a working lunch.

CO-CHAIR ANDRULIS: Absolutely.

(WHEREUPON, the meeting in the foregoing matter went off the record at 12:43 p.m. for lunch and back on the record at 1:12 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:12 p.m.)

CO-CHAIR ANDRULIS: So to start off the afternoon, what we want to do is I want to draw your attention to the PowerPoint that's up on the questions that we and NQSF in particular have put together that are meant to be a digest of a capturing of the conversation that we have had as well as we're, they would like this advisory group to go.

Please, yes.

DR. NUNEZ-SMITH: Okay, so I was just trying to recap, and before we talk about this slide think a little bit about summarizing what I thought I heard over the first half of the session.

And I walked away thinking that we as a community are sort of tasked with three domains of work.

And the first one I thought of was, is to come up with some overarching principles, perhaps some position statements.

And the notes I took were, one, potentially all prioritized measures should be stratified.

That all measures coming up for recertification should be stratified, but also thinking about how disparities might be distinct entities within institutions, and therefore the uptake of measures may vary by institutions and systems. So that one of our charge is to think of some broad position statements such as those.

And then separate from that it's really perhaps criteria development more than measure identification per se in that other two domains, which was thinking about a starter set around the existing measures that we're sort of calling the 700 right now.

And then thinking about criteria for sentinel measure development and those ideas that might be new, but within a context of ability to implement successfully, uptake, be actionable and feasible.

But I just wanted to make sure that that was sort of the consensus of the group as to where we were headed into the afternoon, and if I'm off then to be sort of redirected back on track.

MS. NISHIMI: Okay, if I could just clarify something. We're going to get to issues of stratification in some of these principles when we get to the long list of questions that's on the second page of the agenda.

So I agree that we need to come up with some principles on what we need to, what our expectations are for reporting and how they should be reporting.

I don't want to launch down that path right now and try and identify principles, because that's part of the afternoon's work is to identify these principles not just around stratification but also about reference population, et cetera.

DR. NUNEZ-SMITH: So is it fair

then to say that we are sort of doing both things which is coming up with these overarching principles?

MS. NISHIMI: Yes, there were overarching, cross cutting recommendations.

DR. NUNEZ-SMITH: As well as specific criteria around measure identification, measure development?

MS. NISHIMI: Right.

DR. HASNAIN-WYNIA: I'm being very simplistic, but I want to get back to the 700 and I'm trying to understand.

Of the 700, 135 in the last round, the ambulatory disparity, 30 -- no, I'm sorry. There were 135 or a hundred and something ambulatory measures total at the time. Of those, 35 came forward and were endorsed by NQF, right.

DR. BURSTIN: Came forward and were designated as disparity sensitive.

DR. HASNAIN-WYNIA: Disparity sensitive, okay. So what we have of the, kind

of if we substract that number, we still have quite a few measures that are not ambulatory. So they're either inpatient, long-term care, et cetera, or they're ambulatory again?

MS. NISHIMI: Or additional ambulatory.

DR. HASNAIN-WYNIA: Or additional ambulatory. So what I was trying to do is see if we could get to a point where we know that we've already vetted a number of measures, but it doesn't look like that happened because there's an additional list.

DR. BURSTIN: All measures are now going through measure maintenance at this point, so that that's why this is an opportune time for us to think through, what would we want to see as those measures are re-vetted in addition to what we want de novo measures to bring to the table, whether they're the disparities sentinel affects the direction we go or the more cross cutting ones, what would you want to see them bring to the table?

DR. HASNAIN-WYNIA: Okay, so I was trying to make sense of kind of what we were working with when we, in terms of answering this question for you all about whether the staff should be reviewing the existing 700-plus measures to identify those that are disparity sensitive, which does not identify the sentinel disparities measures, only the disparity sensitive, right. Okay.

MS. McELVEEN: That's correct, and thank you, Romana, for leading us into our next discussion point.

So what we wanted to do as quickly over the break, we just came up with a few basic questions to kind of figure out where the committee is in terms of consensus around some basic ideas that we need to iron out before we move any further.

So the first was just mentioned. It's up on the screen. Should NQF staff review the existing portfolio, and that's the 700-plus measures, to identify a disparity

sensitive subset?

MS. NISHIMI: Show of hands.

MS. McELVEEN: Show of hands.

MS. NISHIMI: Yes, okay.

MS. McELVEEN: Okay, so that's a yes. And if the committee agrees with that the next question is what criteria should be used?

Recommended within the commission paper are the criteria of prevalence and quality gap. If you agree with that we will use that. And in addition, is there anything else besides the prevalence and quality gap that should be considered?

DR. MOY: Can I ask a clarifying question? By prevalence do you mean overall prevalence, prevalence in whites, prevalence in blacks, differential prevalence, and the same for quality gap?

MS. NISHIMI: I'm trying to remember what the papers specified. I think the paper was prevalence within, so something

that was more prevalent in a minority population would rise to the top.

DR. MOY: Delta prevalence and same for quality, delta quality?

MS. NISHIMI: Yes. Assuming, yes.

DR. MOY: So greatest of whatever tracked populations there are.

MS. NISHIMI: Right. Another clarifying question?

DR. McCADE: So under anything else, I think there are a couple of things that should be considered. One of them is that is the measurement easy to obtain, or not easy to obtain but not burdensome to obtain for practicing physicians or for systems if a record.

And the other thing is prevalence is an interesting concept, but maybe we should look at potentially financial impact in terms of the ability to change cost of care with respect to the intervention that's made.

MS. NISHIMI: Okay, with respect to the former, I think that was actually part of some of the original criteria. So I think applying that filter is not an issue if the committee agrees.

With respect to the latter, I don't think we have data that we're going to be able to, I don't disagree that that would be a good filter, it's just that I don't think we have the data to be able to apply that filter.

DR. HAVRANEK: Just as an additional criteria, I wonder if the idea of public impact should be brought in. That there are some conditions that have more impact on the public that capture people's imagination more than other conditions.

So if you compare disparities in mortality rates for breast cancer, I think it has more of a public impact than does differences of mortality or hospitalization rates say for heart failure.

And, you know, I think one of the things that's been missing in this discussion is a sense of what the community feels is important.

That, you know, it's one thing to come up with a list of stuff with a bunch of people in suits in a expensive hotel in Washington. It's another thing to think about what people in minority communities think is important. And I'd just really like to see that issue and that voice brought into this.

CO-CHAIR ANDRULIS: It's a question of measures in context. You know, what are the contexts that are relevant to the way this information coming from NQF would be received.

So community context, political context, social context, other contexts that are key to interest, understanding, application.

DR. HAVRANEK: Yes, I think that's right. I think that, you know, if those whole

field is going to move together then there has to be some, you know, ground swell of support for it. There has to be some sort of underlying support for it.

This can't be a set of esoteric measures that are interesting to, you know, quality wonks. I mean I think this has to be something that captures the imagination of people who are, sort of push their physicians and hospitals and all that other sort of stuff to make some fundamental changes.

MS. NISHIMI: Okay, so impact was one of the original criteria, so it sounds like --

DR. HAVRANEK: I guess I'm advocating for making that be one of the fundamental things.

MS. NISHIMI: So does anyone, can we have a show of hands? Does the committee agree that impact should be a criterion that's considered when this NQF staff applies its filter?

DR. WEISSMAN: Can I make a comment? We went through some of the same discussions. The impact, usability, feasibility, all those sorts of things, they already exist. That's how measures getting to NQF.

The principles, and I don't know how different principles are from criteria but so be it, that were from the 2008 report, each one of them there's some overlap but they referred to impact on the minority population, right, feasibility and ability to affect quality processes on the minority population.

So and we think that's important too, but just to make sure that the discussion goes in an efficient way, I think people ought to realize that there's overall impact which gets a measure already selected, which we assume all these have already gone through that filter.

And then there's this additional one about, you know, impact on the minority

population, ability to improve quality of care in the minority population and so on. That's what these principles are focused on. And we just were emphasizing prevalence and gap but the other ones make sense. It's not that they don't make sense.

MS. NISHIMI: Thanks. No, that's important, so all appreciated.

CO-CHAIR ANDRULIS: Ellen and then Marcella.

MS. WU: Just to be able to learn from past experience, what were the criteria that was used for the ambulatory care screening and how did that work out?

MS. NISHIMI: The four were quality gap, prevalence, the impact of the condition, impact of the quality process and actionable, ease and feasibility of improving the quality process.

How did it work out? It depends on where you sat. Some people thought there were too measures, some people thought there

weren't enough. So I can't provide you anymore guidance other than it really depended. Some people thought 35 was just way too many and others were unhappy that it wasn't a broader set.

CO-CHAIR ANDRULIS: Marcella, then Mary, then Marshall.

DR. NUNEZ-SMITH: My first comment is really linked to that because I was going to ask if there's any utility in thinking about a target number for us as a group as to what was a good size for our starter set or whether we wanted to revisit that later.

But there probably is some range that is reasonable, and I also was thinking around that 30 to 40 that the ambulatory team came up with.

And then my other thought was, I mean I agree that we should probably be starting whether it's the delta, but just for the record, to say that the point was made several times this morning that often where

there is no delta existing currently doesn't mean that's not an important place for us to look in the future for disparities or that we just might not have good enough data in that area yet.

So to have some caveat where, for us to consider that there might not currently be a delta, but there is either overwhelming impact or something else that makes us think that's an important criteria or standard.

DR. MARYLAND: So I guess I would ask because I hear what's the impact, how do we overlay a lens of what's the state of the science that could impact what could be a very emotionally charged issue?

As an example, in disparities one might think about prostate cancer incidents and the issue of whether or not you should screen, and there's lots of debate.

So what is the state of the science for it, and we have to be I would think socially responsible in terms of

thinking about cost benefit analysis. Do we advocate for what we think the public wants to hear as we should do something versus has the needle moved and we now should be thinking of what does the science say we should do?

And I guess as we think about these things, I want us to consider what's the best in terms of what we know scientifically in discussing what could be a very emotionally charged issue.

DR. CHIN: I just drew a blank. Oh, so the question really having to do with like what is practical from the staff's perspective that I think like all the items that Joel and Joe had on the slide are great elements. Some of them start getting hard to operationalize.

So for example, prevalence and quality, very straightforward I think for a first, the staff to go through. Something like impact is tougher.

And I agree with Mary that it

should not be politically based, at least the data we get, but based upon, you know, the science. Or something like actionability, that's another one where it's not as obvious as prevalence or quality gap.

So it's so much what is feasible for the staff. So it may turn out, for example, that if you still get to a reasonable number just using the simple ones like prevalence and quality and then, you know, if a reasonable number are then sent back to this group, then the group can probably, you know, fumble around with the ones that are hard to operationalize like actionability and impact and all.

But if there's still a large number then maybe you guys need to then include us in the filter.

MS. NISHIMI: That would certainly be the approach that I think we want to take even if you hadn't recommended it.

DR. CHIN: Okay. I mean another

issue too is like, and everyone will remember this when the ILM made recommendations recently to AHRQ in terms of equality in disparity reports, one of the major concerns with the kind of reports is that we didn't have prioritization.

So that you basically have 200 measures and so no one knew where to start. And so that's still potentially the danger here unless, and particularly the impact part is brought in.

That, you know, prevalence and quality are one thing but, for example, if there's a big quality gap in, you know, measuring A1c for diabetic patients, that's probably a less important measure even if it's a big gap, than something that has a more of a direct public health effect like, you know, an actual outcome as an example.

CO-CHAIR ANDRULIS: Francis and then Mara.

DR. LU: Just a very simple

question. I may have missed it obviously, but the five criteria that were used in the ambulatory report that Helen, not Helen but Ro mentioned earlier, I'm just wondering could those be the five criteria that we use here as well? I guess I'm missing something there.

MS. NISHIMI: Well, and that's the recommendation from MGH, was that on balance we should focus on prevalence and incidents as we screen now across this portfolio.

Recognizing that when NQF endorses measures it considers a lot of these other things writ large, so some emphasis.

DR. BURSTIN: Especially as the endorsement criteria have been getting harder and harder over the last several years, a lot of these other things are part and parcel impact as one of the must-pass criteria and for all measures.

So it doesn't need to be in this dataset. I think a criteria is what the MGH -

-

MS. NISHIMI: Yes, I think the impact on viewing it from the disadvantaged population is slightly different. So but let me see if I can -- oh well, you had two other people and then let me see if I can frame the question.

MS. YOUDELMAN: Well, and it might be a similar question. So when the ambulatory measures were evaluated and you got down to 35, was it because the other hundred really weren't at all relevant to disparities or it was sort of a prioritization or, that's what I'm sort of struggling with is how to understand like why 35 made it and a hundred didn't versus, you know, yes, that.

MS. NISHIMI: Yes, it was more or less a prioritization. I mean, you know, folks didn't feel that however all 20 or 12 or how many diabetes measures were in the set at that time, all had to be considered disparity sensitive. In fact, I don't even know if any

of the diabetes ones are in there. I'm just making this up relatively speaking.

So once they created a set they winnowed down and said, you know, if you're going to measure one aspect of X care, you know, these one or two measures of a set that's, you know, 12 or 15 or 20 are the most important issues.

MS. YOUDELMAN: So let's just use 20 diabetes' as an example, and regardless, 20 of X. So if there's 20 of X do they always have to report on 20 of X or do they get to pick and say, I'm doing two of X and reporting?

MS. NISHIMI: Who chooses what to implement is a separate issue.

MS. YOUDELMAN: Okay, because I guess I'm also trying to figure out, if you have 20 and, you know, the question is, is it become a huge, you know, obstacle for folks versus we're sort of moving forward with electronic health records and everything?

So if you're collecting race, ethnicity and language once it should be available for how ever many measures you have. But again I think that might be sort of my simplistic understanding of it.

But I'm trying to sort of grapple with this, you know, not putting a burden on it, not creating a negative impression of addressing disparities, but also are we really almost stratifying ourselves more than we have to by saying we're only going to pick and choose certain measures?

MS. NISHIMI: We're not going to pick and choose what people should implement. Folks who are implementing would pick and choose amongst this.

But we are winnowing down the 700 to say, you know, for the time being right now because we don't have, you know, all of our systems in place to make this seamless as part of, you know, an analysis for any entity to just push button because, you know, they

collect race and ethnicity.

They collect all the data elements for these measures so, you know, doing one is, and pushing the button is no different than pushing the button and, you know, getting a read-out for 50.

We're not at that place yet, so what we're doing is identifying the measures that you may wish to look at when you're looking at disparities within your system.

MS. YOUDELMAN: So is part of the evaluation then, if someone else is requiring it for implementation currently that it should rise higher for us in terms of prioritizing because it's being used then therefore we should have this data collected?

Like if there's 20 disparities measures and CMS is requiring two, should we be thinking about saying well, those two should have race, ethnicity, language because they're already being required elsewhere and the other 18 aren't. Like is that a way to

splice this or not?

MS. NISHIMI: Well, yes and no, because we don't select the measures. At the end of the day CMS does. We make recommendations through this new process called the MAP, but at the end of the day the end users, the health plans, CMS and others select their measures.

I think the key is also to think about we do have a usability criteria and which is explicitly about is the measure in use. Is it having an impact in terms of QIR accountability functions?

So that's why in some ways I keep coming back to the idea of saying it's really when these measures come back up that I find it more intriguing and interesting than as they're up for review and you're looking at the overall measure, the impact, the importance, the size of acceptability, et cetera, that you then include this lens rather than just simply retrospectively without the

benefit of another committee looking at the measure.

MS. TING: Thanks. And I don't know, the usability is really good segue in that I think beyond that, I would like to propose that we also consider how many stakeholders in the continuum health care systems are impacted, you know.

So not to say that the ambulatory or inpatient by themself in a silo isn't important, but to extend that to the extent that you have a measure that impacts several key stakeholders who can all report on and I think it would make it more powerful. So I would like us, you know, to consider that possibility.

And then the other thing is -- had a brain leakage moment. Oh, in terms of ROI and how you were saying that we might not have the data, I would also like to propose thinking about the measure in terms of not so much actionability, but in terms of turnaround

time.

So for example, if we were to look at a measure like immunization or inpatient readmission, you know, post procedure, those tend to be a little bit more episodic.

And if you change the system or have intervention you should be able to see improvement more quickly than say trying to impact hemoglobin A1c score, which is a chronic disease and requires a lot of constant vigilance on the part of the doctor, the patient and so on. So you might not see that needle move as easily.

And in my again practical insight to a situation, a lot of times the senior leadership in order to commit to interventions, dollars, whatever, they want to see a quick result.

I mean it doesn't have to be return on money, but they want to see improvement in quality.

So if we pick all these long-term

measurements that are going to be very hard to move, you might lose kind of the power. So I'd like at least to have, consider a mix of some short term versus long term. That's all, thank you.

DR. McCADE: So the white paper we reviewed actually makes a point of this particular aspect of it.

And it deals with the feasibility and actionability I guess in that, and I'm probably certain this wasn't done in the first screening, that the impact of patient or practitioners and systems that take care of predominantly minority population are not adversely affected by the measures that we choose.

And I don't want to put a disproportionate burden on people who just happen to have very large minority practices because it may not be as actionable for them as it may be for other people.

Maybe that criteria should be

taken into account in your selection of the 700.

DR. MOY: This is just a practical consideration, which is that you might also want to look separately across these outcome measures and look separately across different settings, because you typically see bigger disparities gaps for instance in outcomes and processes and you'd bias towards outcomes then if you just took a straight line kind of approach.

And similarly you tend to see bigger disparities in outpatient settings than inpatient settings. So again that would be a bias if you just took a straight line approach.

MS. NISHIMI: Yes, and I agree. So I think the staff probably has enough guidance on the extra criteria, the types of things that the steering committee would like to have considered as they screen the 700-plus.

Primarily, you know, it's like we're just going to do at first, cut a prevalence and quality and then we're going to look to the extent to which we can assess the other criteria that were discussed here.

And, you know, obviously the work product comes back to the committee for deliberation and discussion.

Okay, why was this one in this one not this one had data this one didn't, you know, I don't know, went with my gut. You know, I mean there's going to be those kinds of discussions but I think the staff has the sense of where to proceed with that.

So then the next question, and I think that everyone's in agreement but that we need a show of hands on is, MGH proposed a categorization scheme that focused on, you know, looking at whether, going to Ernie's point and that was a great segue, structure, process and outcome measures, they divided up under patient, experience or so, then they had

provider levels, system level, et cetera.

So was that categorization scheme that they proposed is the committee comfortable with that? Because what we would then do is array the "disparities sensitive set", in quotes, in that categorization scheme and would allow you to see how many structure, how many process, how many outcomes you had where there might be gaps, et cetera.

DR. JACOBS: It would be useful if you could point out on what page that's listed so we can look at it again, please. Thank you.

MS. McELVEEN: Sure. I was just looking in it. So that's actually Section 3.d. on the paper and it's Page 24 in the comprehensive report.

CO-CHAIR ANDRULIS: Sean, did you have --

DR. O'BRIEN: Well, I didn't mean to hamper the progress because this is going back. I just wanted to throw out one other

which is looking at sample size and precision issues when you're doing the screening.

It's not all captured in the prevalence proportions that are very close and is not amenable for the stratification and that will be a practical issue.

MS. NISHIMI: Well, and I think that's a practical issue that we'll get when we identify the principles because that is one issue, sample size.

CO-CHAIR ANDRULIS: Mara?

MS. YOUDELMAN: If you guys could provide a little bit more clarity, and maybe this isn't part of what you guys would do as an analysis, if it's not let me know.

But the root of the potential disparity and when something is sort of patient, the patient's at the root of it, I guess that was a little confusing to me.

And if you guys aren't going to do with it then forget it, but if it is possibly part of this evaluation that you guys are

going to do I'd like to know more.

DR. BETANCOURT: I think, you know, we try to say, so and the easiest example is asthma assessment. When is an asthma assessment done? That is clearly something that the provider initiates.

If there was an issue of communication, right, so some measures are more kind of patient based and those are more related to communication and/or self-management issues.

And then there were system based measures which are more, you know, are systems in place, interpreter services and the like.

So it's kind of a rough way to think about these things. Some are more clear cut than others, but it was our way to kind of think about okay, really particularly for foundation for interventions, kind of what are these measures telling us about where we might intervene in the future?

CO-CHAIR ANDRULIS: I had a

question that may, I think it's in here but I just wanted to hear from you folks maybe clarify it a little bit.

But when I read the section, to me one of the dominant issues, challenges is around care coordination, kind of connecting these. While I know it seems to be implied in here, I was wondering where you might see that fitting in your framework context.

DR. WEISSMAN: There probably are a bunch of care coordination measures and they could fit in any one of these categories. For example, patient experience is very important.

Assessment of care coordination, it could be other things in terms of what practitioners do. There could be measures within the hospital. We said that there'd probably be a lot of overlap.

I think the idea isn't, and by the way I know that some people, these are not additional criteria. This is just a way of you've got 700, you know, if you come up with

a small set whether you want to call it a starter set or something, you might just go back to this categorization and say boy, do we have some in each of these categories? And if not maybe you ought to rethink it.

That's really I think, you know, the utility of this categorization. I don't think you ought to read too much into it.

MS. YOUDELMAN: Can I just go back, because the reason I asked the question and I don't know how to be delicate about this so I just won't, is to me sort of having the root of potential disparity being the patient it's worrisome to me.

And I understand that there could be circumstances I guess, but in general it feels to me like it's almost, you know, the blame the patient kind of a thing.

So and I'm not sure that should be part of this discussion, I mean to the extent that we're measuring what the provider did to do the intervention to get the test done to do

whatever that's what we're evaluating. And I guess it is somewhat process versus outcome.

If some of the outcome is related to the patient has not complied maybe, but it's just a little bit worrisome to me to I guess to see it so sort of starkly in black and white.

And Ellen's agreeing with me because we just had this conversation across the table and decided who was going to raise it. So she can be indelicate with me too.

DR. BETANCOURT: Can I just say though I think it's less, it's not on blame. This is about potential for intervention, right?

So if you're going to do a quality improvement intervention on you see that you have low numbers of asthma assessment by docs and you know that you need to do some, you know, report carding, auditing, feedback and the like to get the rates of asthma assessment up, if you see that it's something that's more

subject to communication then what the patient based piece tells you, is that a coach or a navigator might be able to help the patient deal with some of the barriers to kind of achieve that quality.

So it's less a blame and more of a kind of if you were to improve here where would you improve?

MS. YOUDELMAN: Well, then maybe it's just a coding and could be called communication or something else. But I think it's just by calling it "patient" to me it's sort of like some of it's on the provider doing, some of the patient doing.

And so that just to me is, you know, from the advocate perspective I just don't like it when we sort of put on a pass. It purely is semantics.

DR. BETANCOURT: In looking at this I think the title is poor. It's not the root. It's the, you could call it the implementation lever or the lever for change,

right. I mean if it's with the position with the patient, you know.

MS. YOUDELMAN: Then again if it's a communication issue it's having the right systems in place. It's not the responsibility of the patient to bring the interpreter. The communication services should be in place at the provider level.

So I think it is just a framing issue, but at least from I think the advocate's perspective or the, you know, patient whatever role I'm supposed to be representing, I think it's the naming and the coding more than anything.

MS. TING: Maybe we can call it intervention audience or intervention target audience. That might sound better? No?

MS. YOUDELMAN: No, because I think if it is a bad communication then it's a communications systems issue. And so it is, do you have an interpreter, do you have a navigator, do you have a whatever?

It's not really about where the -- I don't know. I'll let you guys figure it out and wordsmith it. Those are just the concerns that I'll raise.

MS. TING: If I may, so overall I like the categories very much. The only other thing that once again I must ask us to maybe consider is adding another category for more of the system type of metrics for the health care system, you know, as a whole infrastructure stuff.

CO-CHAIR ANDRULIS: Anyone else?

MS. NISHIMI: We got the feedback. Mass General, MGH, you know, consider what was done here perhaps think of some labeling, but overall directionally, I think that we're good to go on this one.

CO-CHAIR ANDRULIS: Okay. We have another set of questions related to these issues.

MS. McELVEEN: So again we're asking for a show of hands. First question

is, should NQF adopt the disparaties sentinel measures approach?

MS. NISHIMI: Anyone who agrees yes, if you could raise your hand.

MS. McELVEEN: And everyone is clear on, okay. So yes, if that's the case should this be applied retrospectively to the entire portfolio of measures, so to the 700-plus measures?

(Off microphone discussion)

MS. NISHIMI: Right. This would be an additional lens as we go through. Anyone, committee members who think that yes, this should be an additional lens.

MS. YOUDELMAN: What's the result? So additional lens being something that --

MS. NISHIMI: Is both disparities sensitive and should be considered a sentinel measure.

CO-CHAIR ANDRULIS: So the idea is to review the set of 700 measures and pull out those that you would consider sentinel.

MS. NISHIMI: So let's imagine we've got all 700 measures and we checked off, you know, all these that are sensitive.

There will probably be a subset or there could be a subset if we look at through the lens of ones that are also sentinel. And the question is do we want to retrospectively look at the portfolio or is this a notion that we should consider going forward?

DR. BETANCOURT: I just wanted to make a point of clarification, because I think this should be disparities sensitive. Because the review of the 700 is disparities sensitivity. Sentinel is new measure development.

CO-CHAIR ANDRULIS: It seems like the language should be disparities sensitive measures.

DR. BETANCOURT: The top one should be sensitive, the bottom one is new measure development which would be sentinel measures.

CO-CHAIR ANDRULIS: The bullet should read disparities sensitive measures and then pull out sentinel measures from it.

DR. HASNAIN-WYNIA: So sentinel measures are the new measures. So when we're reviewing, when the NQF staff is reviewing the 700-plus measures, you're pulling out disparities sensitive measures.

And we're looking at the sentinel measures through a completely, well, through a different lens.

MS. NISHIMI: Well, no, but some of the sensitive measures could also be sentinel, considered sentinel.

DR. HASNAIN-WYNIA: So sentinel, can I get clarification from how you're defining them, please?

DR. BETANCOURT: I think we were thinking of sentinel as kind of new measure development, yes. I mean, you know, this is semantics again.

Sentinel means something that

becomes a watch post for you, right? But the thinking about the first part in our framework, sensitivity is a review of the 700.

And then sentinel is wow, you know, we have disparities in pain management and the ED and there's nothing like that at all in the 700. That should be a disparities sentinel measure that is something new that's developed.

MS. NISHIMI: Right. And what the point I am making is there may well be measures that you consider to be sentinel that, and there won't be very many, but there may well be them within the disparities sensitive set. Just because it's new doesn't mean it's sentinel and vice versa.

DR. WEISSMAN: Just to sort of beat the semantic issue into the ground, sentinel has really two very distinct definitions.

One is, you know, when we think of sentinel events, those are those stand-out

events. That's not how we were using it here.

We were using sentinel in terms of like the Army sentinel, you know, watching over things. But as being sort of thinking about the future and really thinking about them as developmental in the future.

These are areas where, you know, the literature suggests something's going on. There are no measures out there, no good measures out there, and so we've got to think about perhaps developing or asking measure developers to look at those areas.

You could use them either way you want, but I'm just saying is that the way that it's in this report, and I apologize if it's not clear. It's that sort of developmental issue, not sort of the stand-out sentinel event type of issue. That may, you may want to have different terminology.

MS. NISHIMI: That actually is an important distinction because that's not a distinction that I think we interpreted the

report to mean.

CO-CHAIR ANDRULIS: Elizabeth and then Marcella and then Mary.

DR. JACOBS: I have a question about that, because as we talked about and you say in your report and looking at the disparities measures. you said one way to do it is to look at data we already know.

We already know that there are disparities in certain things, and so as we look at measures that we already know there are disparities in.

But some of those measures, and if you use that criterion impact and prevalence, some of the 700 may not fall out even though they could be these measures we should develop and these sentinel measures. Do you see what I'm saying?

So I think I have the same confusion you do, Robyn, which is that I think there is a way to look at those 700 as maybe

not being things that already exist that we know but could potentially be measures.

So I just, that's where I think we could develop new things and look at, take that lens to look at the existing 700 too.

DR. NUNEZ-SMITH: So the framework that I was thinking in that was slightly different.

So I thought that we were looking through the 700 for disparity sensitive that we would then prioritize, so a little bit different, and that any de novo measurement creation would be sentinel.

But then following up on Elizabeth's point, which was one of thoughts I had earlier, which was in the first pass there may be measures that don't meet our criteria for, you know, delta prevalence, delta quality, but we suspect might be important in the future.

So how do we build in a way to revisit those whatever we call them be it

disparities sensitive --

DR. JACOBS: The almost sentinel.

DR. NUNEZ-SMITH: -- or almost sentinel, the quasi-sentinel, the emerging sentinel.

MS. YOUDELMAN: Maybe it's this way. So the first pass of the 700 is disparities sensitive, and then you go back to any one that didn't make that cut and you check if they should be adapted as a sentinel one or a new one becomes a sentinel one, right?

You can't be both disparities sensitive and disparities sentinel. It's an either/or, but there might be some --

MS. NISHIMI: Well, that's the question.

MS. YOUDELMAN: Right, but there might be some, well, I guess I'm going with the either/or, but I think that there might be existing NQF standards that are not disparities sensitive that could become a

sentinel.

MS. NISHIMI: Is the committee -- Mary, you had your --

DR. MARYLAND: And I guess what I would encourage us to do in addition to looking at a glossary and definition of terms down the road, is to be clear that we are not referring to the sentinel as we're all familiar with the Joint Commission, because that raises us to a different bar and everyone would say of course.

So we either need to choose different language or be so exquisitely clear that we're not having this problem.

DR. WASHINGTON: All right, I think we're just arguing semantics. If you look back at the prior slide, what we voted on were, were additional criteria needed.

And it sounds like what people are saying, and I totally agree with that, is that these additional criteria are needed, we just should not call them sentinel.

But I would argue, apply this in addition to the other two, and particularly for the reason that Liz said that there may be areas in which disparities have not been assessed.

So overall quality measures where disparities haven't been assessed, but if you looked they would be there because they fit into one of these categories.

MS. NISHIMI: Okay, so just to clarify then. So we've got the 700-plus measures, the staff screens them with the original lens, identifies a disparities sensitive subset.

The staff then screens those that haven't been placed into that bin with the four criteria that are here that are listed up to see if, in fact, there are sentinel measures.

Is that what the committee is agreeing to? Does anyone object to the approach that I just played out?

DR. O'BRIEN: Why not just screen them all with all of the criteria at the same time?

MS. NISHIMI: Well, how we operationalized the recommendation I think we --

DR. O'BRIEN: Something that I thought that was in there that was important was, moving forward you're looking for new measures.

If you apply kind of the usual framework that's out there or that we might come up with or they may have come up with, you may have missed the boat. There may be areas that are not being addressed when you apply the criteria that are out there, because maybe they're in smaller populations or areas that haven't been previously the focus of performance of quality measurement.

But that there needs to be a separate effort to take what's in the literature to identify those other leftover

things that are being missed and not be a separate effort looking at sentinel measures, so they could be moving forward, sentinel measures and something else.

DR. CHIN: Can you state it one more time?

MS. NISHIMI: So we've agreed to screening 700 measures to identify the disparities sensitive set.

So the question is whether we should take those that aren't identified as disparities sensitive, but in fact may be indicative of the characteristics for sentinel measures, and take a pass at leaning at those.

Recognizing that we're going to get to the question of guidance that we give to future measures, that's another question. Right now it's just what we do in terms of sentinel measures and the existing set.

CO-CHAIR ANDRULIS: Ernie, then William, and Ellen's leaning into her --

DR. MOY: This isn't about the

process. This is just about the criteria.

Another approach that people have chosen are trying to perhaps ensure that there are measures that are specific to specific populations.

So maybe trying to make sure that in this quote "sentinel set" there are measures that are relevant to the American Indian population specifically like alcohol or domestic violence. That's just another criteria that could be applied in addition to these four.

DR. McCADE: So my thinking about this sentinel sensitive sort of question, I guess it was confusing me before and now it's a little clearer, at least it was until we had this discussion, is the fact that in order to do the review to determine a sentinel event one would have to do literature searches to find out what measures or what disparities exist. And then identify them as not being in the 700 that exist right now.

And then having either an RFA or whatever a quest for additional measures that you submit outward to identify what measures should be used.

Now that's how I thought that you intended the sentinel event to be done. And that's not a re-review of the 700 that you excluded. Once you excluded them they're excluded.

And maybe if they become important later on then they may fit the criteria that the data that is now available could suggest that you could use an existing measure that would require a re-review of the whole 700 set or how many other there will be in the future.

But the idea of sentinel I think is an entirely new sort of description of a measure that exists, because there is nonexistent measure and there is data suggesting that there potentially should be a measure.

CO-CHAIR ANDRULIS: Joe, do you,

Joe, want to say anything about?

DR. BETANCOURT: That's a very articulate way of describing, I think, what we were looking at with the exception of using the term "sentinel event", because we're not talking about events but that's okay. And everything else is perfect.

CO-CHAIR ANDRULIS: Luther?

DR. CLARK: Yes, Robyn, quick question. Do you have any sense of what these numbers are likely to look like, I mean being familiar with the 700?

MS. NISHIMI: No.

DR. CLARK: In terms of disparities sensitive and sentinel, no?

MS. NISHIMI: No.

DR. CLARK: So it could be all or --

MS. NISHIMI: Really I mean I'd shoot from the hip if I could, but I can't even get there.

MS. McELVEEN: So the question --

(Off microphone discussion)

MS. McELVEEN: So just to reiterate, recap, we won't apply the sentinel measures to the portfolio, is that right?

(Off microphone discussion)

MS. McELVEEN: Okay, so let me just take a step back for a second. So the paper recommends again certain categories of evidence to determine disparities sensitivity.

And within those categories, that's where they make a recommendation around disparities sentinel and they also list out those four sub-bullet criteria that are shown on the screen?

PARTICIPANT: For sensitive.

MS. McELVEEN: For sensitive.

PARTICIPANT: Sensitive.

MS. McELVEEN: Okay, so according to the paper, where no data exists on disparities for a particular measure or where data exists but shows no disparities, their suggestion is to then apply those four sub-

bullets that are shown on the screen to identify those measures as disparities sensitive.

Then if a known disparity exists but no quality measure exists, that's when we term those measures as sentinel. Does that make sense?

DR. CLARK: I think we're moving away from the word "sentinel" because it has other meanings for providers and institutions than it's being used here.

DR. EDWARDS: So let's say that you come to 35. For everything else of the other 700 you are going to apply those four sub-bullets, is that what you're saying? Am I understanding that correctly? Was that the plan?

MS. NISHIMI: That's what we were asking the committee.

DR. EDWARDS: That's what I understood and I vote for yes for that process.

DR. NUNEZ-SMITH: Because my understanding was that for the 700, to have the staff go through. Disparities sensitive is the only lens through which they would be evaluating those 700, and then a whole other conversation begins separate about novel measures, right, for where there is, for the last situation where there are data suggesting a disparity but there is currently no measure.

So I'm just saying my understanding was the staff go through once, 700 for disparities sensitive, not applying the previously known as sentinel criteria.

DR. EDWARDS: But weren't you asking the question of did we want you to go through for the remaining, go through those four sub-bullets? That's the question on the table, okay, and I'm voting yes for that.

MS. TING: But then in addition, it's possible to also still do the review of brand new conditions, or emerging conditions. So for disparity exists but no measure --

MS. NISHIMI: Yes, we're not talking about going forward right now.

(Off microphone discussion)

MS. YOUDELMAN: So I've gone through my 700 and I've identified 350 that are disparities sensitive. All right, whatever, 40. Pick your number. We know there's 35 and ambulatory care, okay, so get to 40. Of the remaining, can something be sentinel if not sensitive? That's what I'm trying to figure out.

Do you have to do a second lens, and it's not currently care with a high degree of discretion, communication sensitive, lifestyle changes, outcomes rather than process, but it's still a measure that if it was developed slightly differently when it comes up for reevaluation as opposed to from scratch, it would become sentinel.

DR. WEISSMAN: Can I, well, Joe wants to say there's another step, but I like the linear thinking in trying to put things

into buckets and I think that makes a lot of sense.

And here's one way to think about it that you, you know, apply this the prevalence and quality gap and you've identified some disparities sensitive measures. And then you've got a bunch of others that maybe don't fit those filters.

But for example, it may be a communication intensive measure and you might consider that as a potentially, a disparities sensitive measure in addition.

But then I think you're also conflating the two tasks that NQF is trying to do. One is to come up with a list of disparities sensitive measures among the measures that they already have, and then, you know, there's new measure development and where do those new measures, disparities sensitive measures come from?

They come from this idea that there are things in the literature for which

no measures currently exist. And those are what we're calling sentinel, bad word, exploratory, developmental, you know, I think ought to maybe settle on a different word right now.

And that's a different task entirely I think, because when NQF recommends their starter set or whatever they want to do about disparities sensitive measures, there are not going to be any sentinel measures in there because sentinel measures don't exist.

It's the RFA process that William mentioned. Does that help a little bit maybe? And Joe, do you want to --

DR. BETANCOURT: I think the slide that we had that was the algorithm, I think it does a good job of this. Because at Step 1, you know, let's say, let's dial back ten years.

There's a measure on inhaled corticosteroids for pediatric asthmatics. We know there's a disparity there. That gets

pulled out right there. We know that's like disparity measure. That's one.

Then the second is, well, there's something that's that there's no disparities, but if we look at that filter there, those four points, yes, this could be something that, you know, if looked at could be sensitive as well. Pull that out as well.

And then the separate task which Joel's mentioning is, you know, after you've done that with the 700 then there's an RFA process that's saying well, you know, the NQF concludes that there are these disparities.

There's no measures at all, and we need to go through the sausage kind of making of developing those new measures that have a more disparities focus to them. That's what we were trying to convey.

CO-CHAIR ANDRULIS: Colette, is your -- Romana?

DR. HASNAIN-WYNIA: I was actually going to let this go, but then I went back.

So I'm actually quite confused now, because I'm looking at Page 22, so help me clarify, please.

Page 22, prior to the section 3.c, the last paragraph and how disparities sentinel measures are distinguished from disparities sensitive measures. And as I read this I am thinking that what you just proposed is a different definition.

DR. BETANCOURT: All I can, I mean if you look at what we've done at Mass General, we didn't have, we know that disparities, we call our disparities sentinel measures right now, pain management in the emergency room because by race, ethnicity which we did not have any measure that we were reporting to anybody or to anything.

And so we created, we began with a chart and then began to create our own internal measure for that. That's we called it.

So we stratify our core measures,

HEDIS, CAHPS, you know, all that stuff because number one, we think that's important.

We think that evidence shows us that many of those are disparities sensitive, and so the sentinel piece is something that it's a completely new measure development. That's what we were trying. And if we didn't communicate it clearly, I'm trying to communicate it clearly now. That's exactly what we're trying to do.

DR. WEISSMAN: I would listen to what Joe is trying to explain and --

DR. BETANCOURT: Yes, if it got jammed up then that would, does that make sense?

DR. HASNAIN-WYNIA: Yes, so can I just, should this last paragraph then be edited and rewritten to clarify?

DR. BETANCOURT: Sure.

DR. HASNAIN-WYNIA: Okay. All right, because that ended up being quite confusing to me because I thought that these

were not necessarily new measures, but were measures that were specific to measuring disparities.

Not just disparities sensitive, which are general quality measures that we can look at through a disparities lens, but these were, that sentinel as defined here were very disparities sensitive, or disparities specific for helping organizations recognize potential disparities.

DR. WEISSMAN: We apologize for --

DR. HASNAIN-WYNIA: Okay, all right.

MS. YOUDELMAN: So can I just build on that and clarify, because I think I've now got it too, at least I hope I do.

So an NQF standard already is collect race, ethnicity and language data. And you may have another standard, a hemoglobin A1b, c, whatever the heck that is, H1, whatever.

So both of these could be

disparities sensitive under this framework even though one is very explicit to disparities, which was collect this data, and one is a more generic, get your A1b -- stop laughing at me. Thank you.

Get your A1c, but since we can stratify it by race and ethnicity it's also disparities sensitive.

So we may have different levels of disparities sensitive and some that are really explicit and happen that way and some that we think are just sort of getting there. Do I have it?

So we don't have a hierarchy within NQF standards where are some are really great and some are not. They can all be disparities sensitive.

MS. McELVEEN: So now that we have an understanding of the algorithm, let's just recap. Everyone agrees we should review the entire portfolio to identify them as disparities sensitive, correct? Okay.

And everyone agrees with the current definition of disparities sentinel. That's a completely new measure. I'm seeing some head nods.

DR. O'BRIEN: I'll say something that I feel we're voting on this, is once you take a vote on it and then the NQF staff feel compelled to move forward and carry it out, trickles down.

I feel like there's a lot of confusion and I feel like I heard ideas that I thought sounded great, but also the overwhelming sense that there's confusion. And now we're going to vote on something potentially while there's still confusion, and there may be everyone have a different idea in their mind about what sentinel means and what we're distinguishing here and distinguishing there.

Vote on it now and now it's, your stuck with it and it's going to affect the way all the measures get developed moving forward.

And it seems to me, I'm not trying to like stop any progress, but it seems like you'd want to have this watertight before you make a vote.

CO-CHAIR ANDRULIS: Well, I think there's some clarity around the idea of examining the 700 measures in terms of sensitivity.

I think the term "sentinel" should be removed. And you refer to "novel", emerging, other, but just get rid of the word "sentinel" as it's really confusing if we're kind of kicking back on it. So it's novel, emerging, or heretofore not identified or whatever.

DR. JACOBS: So what we're voting on now is that the 700 would be reviewed, but with an eye of what we know are existing quality measures for disparities and those that would mean the other four criterion. Okay.

And then in addition to those

we're going to think about new measures. We'll just call them new measures.

CO-CHAIR ANDRULIS: Right, yes. Are those up, or questions, Donna?

DR. WASHINGTON: I was just going to offer similar clarification. That in essence what you're saying is that the definition for disparities sensitive are the first two major bullets up there. One of which includes the known disparities and the other which includes these four criteria.

CO-CHAIR ANDRULIS: Yes.

DR. McCADE: And as part of this then, the staff's work, in terms of helping us to figure out what sentinel measures we should take or whatever word we're going to call it, novel measures we should take, are they going to provide us with literature of one type or are we going to use our own expertise to help to develop literature?

I'm not sure which is the case to develop the literature.

MS. NISHIMI: You're going to use your own expertise to make recommendations to staff on what areas you think would be appropriate.

DR. O'BRIEN: Could I ask a question?

CO-CHAIR ANDRULIS: Absolutely.

DR. O'BRIEN: So was your intention, so disparities sensitive, classifying measures as disparities sensitive has been done in the past. It was done five years ago apparently.

And I thought that this group was doing something different, which was not just looking at existing measures through the lens of disparities, but actually saying, how do we address measurement of disparities in health care?

And that I thought it wasn't that the new measures would all fall into this bucket of what you were calling sentinel. It was rather that moving forward there would be

a call for measures.

And that wasn't to mean that everything that would be in that call for measures is going to be the sentinel measures, but sentinel measures was a second bucket of new measures. Is there one bucket of new measures or two in what you're proposing?

DR. WEISSMAN: So was your intention, disparity sensitive classifying measures, disparity sensitive has been done in the past, was done five years ago apparently.

And I thought that this group was doing different, which was not just looking at existing measures through the lens of disparities. But actually saying how do we address measurement of disparities, in health care.

And I thought it wasn't just the new measures would all fall into this bucket of what you are calling sentinel, is rather that moving forward there would be a call for measures.

And that doesn't mean that everything would be in that call for measures would be in the sentinel measures, the sentinel measures would be a second bucket, but sentinel measures was a second bucket of new measures.

Is there one bucket of new measures or two in what you're proposing?

DR. WEISSMAN: Well actually, 2008, was the report, was only about ambulatory care measures. And they came up with a set of principles that are very similar to criteria, although the principles applied, even the impact kinds of things applied.

They said, you know, impact on minority populations, so there, there is some overlap. But they came up with the principles, but they came up with 35 disparity sensitive ambulatory care measures.

Now there, not only are there a ton more ambulatory measures, but, you know, there are other institutions involved,

hospitals, nursing homes, and so on. So now you've got these 723 measures.

So I think, what we were asked to do was kind of, you know, repeat this process were those principles, are they good candidates for criteria? And should there be new criteria considered, and so on.

And so what we've done, and we said, well, you know, those are pretty good criteria and especially prevalence, and quality gap are the ones you really want to focus on.

So as Robin was saying, you know, you go through all those and the first cut for disparities sensitive is prevalence and quality gap, for disparity sensitive.

The next cut is for the ones that don't make those cuts, you go and you apply those other four criteria to see if any stand out, I like to think of the communication ones.

There may be ones that there's not

an obvious disparity gap right now but, you know, you ought to think about those as being disparity sensitive. And then, relying on the expertise of this group, you may want to put out some RFAs saying you know what, this is an important area in disparities but we have no quality measure among those 723, 723's not enough

So we need some, let's develop those, and then, I don't know if that answers that question so far, and then a separate question would be what gets done, with these measures that you all identify?

And I think that's where the role of NQF versus NCQA, you know, and who does what with what, is maybe getting a little conflated. You may just want to suggest a set of, you know, like a starter set, that somebody could use to characterize or profile an organization as having equitable care.

Maybe you want to add some modules to be sensitive to some of the geographic,

differences in the country, right?

You might want to do that, you might want to make all those kinds of recommendations, but you're really getting into another territory there about how they're actually going to be used.

And that's why I think, Robin, kept saying we're going to list a bunch of measures. It's up to other people, other organizations to select which ones they want but we're going to say if you want to profile your organization in terms of it's equitable care these are some disparity sensitive measures, that you might look at. Sorry to go on for so long but I don't know, does that help?

MS. NISHIMI: Well and, but to go to your point, Sean, we will move into the kind of guidance that we can give organizations. On the criteria or the principals or the recommendations, on what they should use as they identify, which of

those measures are better or worse for them to implement. We haven't gotten to that part of our --

DR. BETANCOURT: It might be good to do a clinical example, I'm just kind of trying to think through my head of a clinic example.

Helen mentioned the end-stage renal disease bucket, right? So lets just say there's 20 measures in end-stage renal disease. And you look at them and you say well, you know, we know there are disparities in end-stage renal disease.

So there's three of them that are about, you know, some aspects of referral to end-stage, to being listed for renal transplantation. We know there are disparities there, those three get pulled out.

But then there's like five others where we have a suspicion that, you know, that around communication issues or did the patient receive information that they understood, that

are there but there are no disparities.

But those are root causes for these disparities and those would be filter Number 2 which is okay let's pull those out.

But then we know that there's a lot research on trust. I'm just throwing this out, mistrust and mistrust is a big issue but there's nothing in that whole 35 on end-stage renal disease that is related to trust, and that might be some kind if sentinel measure, around trust measurement, you know.

So I'm trying to, I'm kind of trying to walk through these three buckets that, you know, the new measure development is something around trust, nothing there, you know, some that are obvious disparities.

Those are takers and then there's a couple that there's no disparity yet but there's enough evidence to suggest that are communication sensitive, disparity sensitive.

DR. WEISSMAN: Where Joe's coming from is at MGH, we developed, or they

developed a, what they call a sentinel measure around pain management right? So there was no vetted measurement now we have this, what we're calling a sentinel measurement at MGH which is around pain management.

That terminology doesn't apply to NQF because once you have that, if you have that as a measure it's no longer exploratory, it's no longer, you know, it's one of your measures.

So I think that's where, you know, we have to make that distinction also. In terms of what an individual provider, you know, that's forward thinking like MGH might be, versus you know this is a national body that will have vetted approved quality measures.

So then this sentinel measure is no longer sentinel it would become one of the measures. It's more like the RFA process.

CO-CHAIR ANDRULIS: No flags up, okay, we're ready for the next section then.

Did we vote?

MS. NISHIMI: I don't think we need to vote on this one, I think we kind of, we got it.

CO-CHAIR ANDRULIS: We got it.

MS. NISHIMI: We got the message.

CO-CHAIR ANDRULIS: So what we'd like to do is, like I said before, the new 2 o'clock we'll start the mythological approaches to disparity measurement.

That's Section 4 and since there are nine sections here, we've got about 10 minutes each.

MS. NISHIMI: What we're really looking for here is MGH provided some discrete recommendations. You know, for instance around stratification, how you should proceed in terms of a reference point, et cetera, it's this is the second page of your --

CO-CHAIR ANDRULIS: Second, it's on Page 8 of your agenda.

MS. NISHIMI: Page 3.

CO-CHAIR ANDRULIS: I'm sorry page 8 up here, sorry Page 3 of your agenda.

MS. NISHIMI: Page 3 of your agenda so these are the kind of principles that we're looking for the group to recommend.

CO-CHAIR ANDRULIS: And so you'll see that each of these have questions attached to them, so we can, shall we start off down the, unless there are some procedural points? I can't believe there would be any procedural points in this group.

MS. McELVEEN: There's also, sorry, there's also a table in the full report on Page 46, just so everyone has all of their reference materials.

DR. CLARK: No, I just think there might be a typo, shouldn't it be an advantaged group not disadvantaged?

CO-CHAIR ANDRULIS: Historically advantaged group, should the reference point be the historically advantaged group?

DR. CLARK: Not disadvantaged?

CO-CHAIR ANDRULIS: Not disadvantaged, that's right.

DR. CLARK: Okay.

CO-CHAIR ANDRULIS: So the floor's open for discussion, on let's see if we can go down these point by point for the time being. Francis do you want to start us off?

DR. LU: Again a very simple or simplistic question but is that entirely clear to everybody, the historically advantaged group?

CO-CHAIR ANDRULIS: Don't you folks define advantaged within the paper, or do you just kind of naturally assume that everybody knows what advantaged is?

DR. WEISSMAN: Yes I think so.

CO-CHAIR ANDRULIS: Do you use Paula Braveman?

DR. WEISSMAN: She did a really good article a few years back on defining what disparities are, and I think you know to this day there are still many well respected bodies

that are taking, you know, the largest group or the best possible group or a benchmark and so on.

And I think these are all, you know, relatively valid choices, but it just seemed to us that if you're really talking about equity and disparities reduction that, you know, there are too many instances where you get unexpected results.

And what you're really interested in is in comparing the minority historically disadvantaged population against the advantaged population. And I mean, and to put a label on it, it's usually the white non-Hispanic group that gets chosen for that.

And in terms of race and ethnicity and, you know, I think that was in our thinking, but really based on Paula Braveman's article.

CO-CHAIR ANDRULIS: Won't that have, it seems to me that would have regional implications though. For example or

circumstantial implications that just like, you know, if you're in Appalachia, you know, that's the white population would be kind of a, that would be a confusing reference point for advantaged group, you know.

So I'm just wondering whether that would require, while Paula Braveman's overall approach might be valuable, there might need to be some further language refinement to that, beyond just simply an overall statement about, it's white population generally.

DR. WEISSMAN: Yes, I mean I think that's where some of the social determinates come into play. Socioeconomic status that people mentioned earlier, but we're focusing on the racial ethnic issue.

CO-CHAIR ANDRULIS: Comments? Sean.

DR. O'BRIEN: Do you know how they do it in HRQ reports and would they --

DR. WEISSMAN: Ernie's right there, ask Ernie, don't ask --

DR. MOY: Yes I think, I don't. So we do it more, we just compare it all to whites or non-Hispanic whites or, you know, high income or high education, and so it's fixed across all of our comparison's and I didn't know, it wasn't clear to me if that was meant to be true here to or if this was going to vary from measure to measure.

DR. WEISSMAN: I think we would basically follow that recommendation, and that would contrast with, I'm forgetting the other example, the only other example I'd given here.

But I think it was a CDC example where you know they used a, in their disparities index, their summary statistic, they always took the best performing group and that's a real contrast, you know, in terms of what we're saying.

And it should be recognized, as a contrast, but I would think that we would go with the high, you know, in terms of SES we'd

go with the highest income, in term of race ethnicity we'd go with the white non-Hispanics as the reference group.

With the understanding that in certain parts of the country there may be some granularity that might be more appropriate.

CO-CHAIR ANDRULIS: Francis.

DR. LU: Yes, so I'm just wondering would it be of benefit to align with what has been done at AHRQ, so that it just makes that clear with the caveats that Dennis brought up about the local issues, you know, that need to be taken into account.

But just so that, because I'm just concerned that historically advantaged group from a measurement point of view may be a little ambiguous to some people or it's not entirely clear. Whereas what Ernie mentioned just before seems to be a lot clearer.

DR. MARYLAND: And maybe actually listen to the discussion that could become one of our tenets that we use standardized things

rather than new ones, because it tends to probably unnecessarily complicate the issue.

CO-CHAIR ANDRULIS: Are there comments on this? Next is absolute versus relative disparities and favorable versus adverse measures. Questions are, should both absolute and relative statistics be calculated and should public reporting of disparities calculate statistics, using both favorable and adverse events.

MS. WU: Well I'm just assuming, that it goes without saying that there's explanations included with the statistics, it's not just a book of statistics that don't, I don't know, who to answer, I don't know who I'm directing this question to.

Given that there could be potentially conflicting statistics that are reported out there needs to be a narrative that goes along with it, yes.

DR. WEISSMAN: Yes, and I would just, just to clarify the recommendation, first of

all, I would alter the second sentence in the sub bullet so, should, oops, let's go back. They should both be, one said calculated the other says reporting.

They should, I think the recommendations they should be calculated, they should both be calculated. Both relative and adverse and favorable.

Sorry, relative and absolute and favorable and adverse should be both be calculated. And then I think the idea would be see if they give a consistent measure.

And it should also be clear from that little graph I showed, that at a single point in time, it doesn't make that much difference in terms of relative and absolute. It's over time that you could come up with different conclusions.

The Trivedi article was a classic case and, Ken Keppel and I ended up writing to the editor of the New England Journal of Medicine because, you know, we found very

different conclusions by using a different set of measures.

The favorable and adverse issue you can get, that has two issues involved, one is you can get a different answer over time. But the other is that the reporting to the media, the public perception can be very different, and that's a similar but, it's a similar but different issue.

And again, you know, you may want to calculate both and see what kind of message you're trying to send and if it's very consistent. Then pick one because the simplicity of the report is pretty important. But these are complicated issues unfortunately.

CO-CHAIR ANDRULIS: Yes.

MS. YOUDELMAN: Okay, so you're not recommending to report absolutes, both the absolute and relative statistics, so there's not really a direction which one to report, you're saying calculate both but not

recommending to report both.

MS. WU: See Page 30. The recommendation that they put in their report is close to the question. The recommendation was, both absolute and relative to be calculated and if it leads to conflicting conclusions, both should be presented allowing the reader to make their own interpretation. So I think I agree with that.

MS. YOUDELMAN: Assuming that there's some kind of narrative?

MS. WU: As an explanation of why there's a difference, yes. So was that what you wanted to know?

CO-CHAIR ANDRULIS: That's part of the discussion, I guess the question to me that comes up is if one of them shows an adverse effect then, more of an adverse effect so you, let's say the case where you've got, you know, both lines going, both lines going down, but the rate for whites is much stronger than the rates for other populations.

Might someone say, hey we're all doing well here. Versus somebody else saying, hey you know this is, the race is continuing but somebody's falling further behind even though you know we are looking to improve to this level, what's the level we're trying to get to, this rate is much slower among this population, the rate's much slower among this population.

DR. WEISSMAN: I agree absolutely Dennis, it's if you can, I mean even just two lines, you know, on a graph, and they could be, you know, very low, very high, they could go in different directions, they could go at different rates.

It's amazing how much just two lines can vary so much and so there, that's where I don't think there's a right answer. And I think there's a context and a certain value.

And I think the other example I gave, was the Warner article that had, you know, I thought a very inflammatory title about heart

disease among blacks and whites in New York state, saying it got much worse and in fact the African American rate, had been in at a fairly small level and tripled, compared to, you know, an improvement compared to only a doubling improvement in the whites. But the disparity how as they measured it widened and so they said things got worse.

I don't know, you know, I mean, I think that's for people to decide and it is somewhat of a value judgement, and it's, I just think if you come up with different conclusions over time as to whether things are getting better or worse, you probably ought to let, I think reasonable people will disagree.

DR. EDWARDS: I had a couple questions an an intense concern. The one question had to do with, are we going to have any comment, or is anyone aware of anything, that there's any consensus around, as to what's the threshold percentage for something to actually be called a disparity.

We wrestled with that at CIGNA, when we were trying to decide what we were going to go after. We never found anything so we had to do our kind of best judgement and move forward.

So that was one question I had, and then I'm very concerned when we start talking about leaving the reader to make their own interpretation.

This is really complicated, look how long it took us to be able to be even to tell you, hey what we thought was going to be a simple task.

And then when I watch the news and I can watch two different channels and it's like, are they reporting on the same story? So that makes me, I don't know what the answer is, but that makes me very, very concerned when you're talking about something this important, and this emotionally charged, and I guess I'm certainly not a statistician.

And I understand the whole point of if you look at it this way, it looks good, if

looks at, it looks bad, but what's real, is there really, what is real, is it better or is not better?

And if there's a question I vote for us to go with the worst case scenario conclusion rather than people deciding to selectively interpret it that things are fabulous. When the people, you know the patient at the end of the story is not any better off.

CO-CHAIR ANDRULIS: Yes I think that's what the point I was trying to get at also with you, Joel, is that there is disparities in context you can bring up.

If the trend is down for everybody, that's fair game to cite that, but if the gap is widening, as the trend is down, then there is something else that needs to be considered there.

And that, I think that measurement should be the focus of the discussion, rather than the, what could be interpreted by other

people locally to say, everybody's doing better so, you know, it's problem solved, or as being solved.

Now maybe it is, but I think it needs another layer, I think what you want to do is stimulate discussion around it to consider well, what's going on, maybe it is okay, but it seems to me it's a flag of some sort, may not be a red flag but it's a flag of some sort.

DR. EDWARDS: And the dollars follow, that's the thing.

CO-CHAIR ANDRULIS: We're going to Ernie and then Luther and then Marshall, and then Romama, and then Donna and then, no sorry, and then William. Okay so you got that order Ernie, just zig zag, I think Ernie you were next and then Luther, Romana then. Okay.

DR. MOY: A common corollary to this one, because I agree that reasonable people can choose to disagree, based upon these findings, is to show the actual rates

themselves.

As opposed to just showing the difference or the relative rate and then people can look at those rates themselves and have a little bit more information from which to make a decision.

So I believe there's a corollary, to not only just show difference or relative rates but also to show the actually rates.

DR. CLARK: So I guess I'm next, I have the one same concern because there is a question, I think it may, it makes an assumption about the sophistication and the motives of the interpreter, and I think that may not always be in the best interest of what we're trying to do.

So it might be that one way to approach this is to give some guidance as to when relative risk and absolute risk should be used.

Because if it was just left to the discretion of the interpreter then that would

concern me. And I think that would create some potentially problematic situations.

DR. CHIN: I'm wondering if it may be possible to be even more proscriptive, you leave a little wiggle room, for example, of issue of relative and absolute as opposed to say, well example present both. But I think one thing that may be worthwhile for staff to look at is the recent IRM report to AHRQ regarding the quality in disparities reports. There's one chapter that was contracted out to a communications firm, where they came up with, you know, very explicit guidelines on, you know, basic stuff which was really important.

So things like the titles on slides, you know how you label X axis, Y axis, how you label the different lines in a graph, but sort of a standardization approach.

And it seems like these five different things are all sort of that ilk, where I think you've identified some really

key issues. But I wonder, I mean a lot of comments have just been raised, there be like too much leeway, you're giving.

In terms of like the judgement for example to relative and absolute risk seems to be a clear one where, the can really be misused.

And so, unless there's reasons otherwise, I would think that then both might be the way to go, but I wonder if it is maybe possible to have like, you know, a document that essentially codifies some of these recommendations and makes them feel explicit.

DR. HASNAIN-WYNIA: I would echo that, what Marshall just said, and I think also, I mean, there's just, the thing that's striking to me here is that, you know, the goal here is to improve quality and reduce disparities.

And we know that if we only focus, at times, on generic quality improvement, we do see that the gaps remain because of the

differences in the relative rates of improvement.

So I do think that, just to support what Marshall has said, that it's important that we show both. I think it's important to show that there may be improvements but those improvements my not be addressing the disparity, which is our argument for addressing the disparity in quality improvement.

Because without having both sources of data we'll miss one or the other side of that picture. But I think that just ultimately, it's this notion of, we have to be able to target the disparity and where we see that is when we see, you know, if we're doing two groups and we see both groups improving but we still see the gap remaining or getting worse.

So I think it's an argument for providing the data, that shows that you have to target the disparity to address the

disparity.

DR. WEISSMAN: I mean I agree with that but I just want to note that, you know, you used words like gap, disparity, improvement, and each one of those, when you operationalize them in terms of the actual rates, could mean different things to different people.

And I think that's, you know, where the, I wouldn't even call them statistical, I mean, they're like math issues, right? I mean that's where, you know, when you're trying to look at differences in differences or differences in ratios over time, and the numbers just behave funny, and it depends on if they're a big number or a small number.

I mean, Dennis, was talking about, you know, a general improvement over time, but the gap gets bigger. Well maybe the gap gets bigger but maybe the ratio gets smaller.

You know, and then what does that mean, do we say that the disparities improved

or got worse? And I think that the danger is when this does move towards high stakes issues, like reporting, that, you know, the tendency is we want to, you know, I talk about this report card that I worked on from Massachusetts, we wanted to give a thumbs up or a thumbs down, I mean that's what we really want.

But sometimes maybe that you don't want either of those things when the math goes in different directions, you know, if you're not willing to sort of give the actually rates, like Ernie said.

Ernie, I mean, I think that's great, you can give that information but I just don't think the public would be able to, you know, do anything with it.

DR. MOY: If you want the QRD approach to this, is if the two absolute and relative disagree we say no change.

DR. WEISSMAN: Yes, could do that too.

DR. WASHINGTON: Just to reiterate, I agree I think that rates should be given because if the choice is left to the person reporting it out, they may report absolute statistics one year, relative a different year. And so then it's difficult to track over time, and do those calculations.

So I would report both the rates as well as both quality and disparity, well if you give the rates you kind of collect your own, quality and disparity.

MS. McELVEEN: Sean, Marshall. Sean.

DR. O'BRIEN: I guess there's some applications where a single number is needed and, I mean, what do you do in those situations, you're right, I don't know if there are any interventions out there that are, incentivizing hospital level improvements in disparities.

But if there's a pay for performance context you need a single number to rank a provider and decide who to reward and who not

to reward. So if in some cases you can't avoid this issue and what would you do in that situation?

DR. WEISSMAN: Is that a rhetorical question?

DR. O'BRIEN: Well no.

CO-CHAIR CORA-BRAMBLE: I didn't know if you wanted an answer or if you were just making a comment.

DR. O'BRIEN: Well, because I think at some point someone's, we are going to be ask to vote on this issue, do we agree with it or not? But I guess in order, I think the answer depends on a context and that recommendation can't work out in all scenarios, because there are scenarios where you need one number, I think.

I think there's some areas where you need one number, not two, so maybe I guess I'm suggesting that it needs some type of qualification. And then on a related question, typically the NQF measures have

specifications.

And usually it's a single number, and you're specifying a numerator and a denominator, there are a lot of different reporting issues and sometimes NQF doesn't really address some of the reporting aspects, but sometimes they do make recommendations about, well present the number and account for imprecision and other aspects of how it gets reported.

I'm trying to get my head around what will the specifications look like for the types of measures we're talking about? Are we talking about a set of if we have a domain that has five populations, we're talking about five numerators and five denominators?

Or are we talking about a specification that actually includes, you take, you take this proportion for this group, and you divide it by this proportion and get the ratio and would that actually that level of detail be incorporated into the

specifications that some TAP and Steering committee and eventually others will either vote up or down on. I don't know what everyone else had in mind for that little of detail or not.

CO-CHAIR CORA-BRAMBLE: I don't know if we're ready to get to that level of detail, Joel, and I'll let you comment, you'll be the author of the, you know, I don't know that we're at that point. My sense is that depending on what measure you chose, you know, you almost got different results.

DR. WEISSMAN: Yes, I mean, I don't think we were trying to get to that level of detail, I guess we leave it up to you all to figure that out.

I think there certainly are times when you want to get to one number, especially of you're going to use it in an incentive program. And you could conceivably build in a lot of this complexity into that, I mean, I think, you know, Ernie's idea, recommendation

that, you know, if the absolute and relative, this would be an improvement score right?

If you think about a pay for performance program that not only looks at benchmarks and thresholds, but improvement over time, this is where it would come into play.

And then you would, the question is would you look at the change in the gap or the change in the ratio, and if they conflict you might just say, you know, no change.

That's one possibility, you know, the same idea of making reports over time, state report card and national report card, that Ernie does and again if the things conflict over time you might just say, you know, indeterminate.

CO-CHAIR CORA-BRAMBLE: I think the issue is when we get to, if there are performance incentives, or of one is going to get a differential in terms of per member, per month, or something based on certain outcomes

then the clarity in terms of the measure is going be imperative.

But I don't know that we can determine that now, but I do agree with you that there needs to be clarity.

MS. McELVEEN: Ernest, did you have a comment?

CO-CHAIR CORA-BRAMBLE: Did you have one?

CO-CHAIR ANDRULIS: Any conclusions?

MS. FITZGERALD: This is Dawn, if possible, on the phone, can get in the queue?

CO-CHAIR ANDRULIS: Absolutely, Dawn, you're on.

MS. FITZGERALD: Okay, great, I don't know when to raise my hand if you can't see it, so I'm just going to jump in where I feel most relevant.

I would like to go back to the original comment, that I believe it was Ernie made, about a recommendation itself that, and while I agree with the recommendations in

terms of the provision of both statistics, particularly when there's conflicting solutions.

I think a more relevant piece of information is the need and desire to have the actual rate displayed, the tending over time, such that, regardless of what statistic is used, one could make it their own calculations for the alternative if possible.

And that also resolves the issue I think where we kind of got into a discussion around implementation and what one would use for sort of a payment incentive program.

You know it's best when you're giving evidence at program that you're fairly transparent in that calculation. And so, you know, providing the rates is the only real way for, to get to true transparency and how you calculate the three when the unintended rational might be for using over the other.

CO-CHAIR ANDRULIS: So at least part of the discussion is around use of the

absolute versus relative, is there a sense of both? Is there a sense of waiting one way or the other, could we put a vote to?

Is there general agreement on trend, using trend information where possible, and also, well, all information is appropriate to use whether there should be a weight toward that showing that kind of optimal disparity.

Or more direct specific significant disparity should one continue to exist, as featuring it not necessarily coming to the conclusion about the feature.

Making sure that becomes a point of focus, rather than being able to be, the information being able to be diluted in some way shape or form, than being picked off.

CO-CHAIR CORA-BRAMBLE: I would go on record saying that I think the trend comment most accurately captures I think where we want to go.

I don't know that we have the level of granularity to say that one method is

better than the other. I think that whatever method is chosen you have to stick with it so then you can determine if over time you're making an impact. And that's, you know, purely from sort of a quality improvement perspective.

CO-CHAIR ANDRULIS: Great, silence is assent, everybody agrees, okay.

(Off microphone comment)

CO-CHAIR ANDRULIS: Then I'll offer, which do you prefer, do you prefer using trend information and then highlighting, should there be a disparities, a significant disparities issue that's still present?

That that be featured in the context of additional information, rather than just laying out, I don't, that's option one.

Option two is just as Joel had originally suggested, you provide the information, if there's a conflict or there's something that varies, that you just let it be interpreted by the audiences that are using

the information.

DR. HAVRANEK: That's not an either or is it? I mean you can have, you can say that trend data should be shown, absolute differences should be shown, and relative differences should be shown.

CO-CHAIR ANDRULIS: Right, yes. I mean they can be blended but it depends on what, I guess, the guidance and interpretation of the information comes into play.

Because I think it still opens the possibility of somebody saying, well things look better. But at least, you know, you can't force it, but at least you can say these are some of the points to consider, you know, just because the trend line's down doesn't necessarily mean that the disparities are going --

DR. WEISSMAN: Dennis, can I suggest that everybody just look at Page 30, while they're thinking about this. And in a sense that has, you know, that's got the individual

rates for each of the populations, it shows a change over time and it displays how, you know, did it get better or worse, well both.

So that, I mean, that really gives trends, it gives rates, and it gives the disparity over time, the change in disparity, depending on how you define that. And so everybody burns that into their head.

That's what we're talking about and I wish I had a better recommendation, but you know, and this is just one example, you know, based on numbers and sort of a mid range of percentages.

CO-CHAIR ANDRULIS: Okay, Grace and then Ellen and then Marshall.

MS. TING: So just based on the experiences at my company as well as those of other health plans at the National Health Plan Collaborative, we do, those of us that do track health disparity matrix, do track both absolute as well as relative.

And we do look for trends, and then

the other thing is that we also look for favorable as well as negative comparison because the favorable minority outcomes means that there is no disparity among Caucasians.

So the best practice right now is just set the best preforming benchmark, and then measure up to that, so I think that these are very solid recommendations.

CO-CHAIR ANDRULIS: Ellen.

MS. WU: Turning to Page 30, I looked at my notes here, are we talking at all or noting a statistical significance, because there can be a difference, but it's not significant right?

DR. WEISSMAN: Yes, well I would ask Sean, to weigh in on that. You know, that's my one comment about statistical significance, is that, I guess I have two comments.

One is that it's very dependent on sample size, so at AETNA or WellPoint everything is statistically significant I

would guess. But pretty close, and so I think you have to be careful with statistical significance.

And the other thing is that some of these disparities indexes, which we haven't even gotten to yet, don't have well defined statistical properties. And so it becomes quite complicated to even calculate a confidence interval which would give statistical significance.

Again, unfortunately I don't have a strong recommendation, you know, a lot of this gets into judgement, discussion, context and that kind of thing.

MS. TING: So a follow up to that, so if I may comment on that based on our experience do track statistical significance. When we indicate those points, you know, by little up and down arrows, but we also do look at, in terms of general size of a population.

So you might have an area where huge

statistical significance but very very few people, or another area with only one or two percent absolute, but a very distinct, very statistically solid difference, and has a very large population like the county of Los Angeles.

And it makes more sense to then focus resources to try to move the bigger populations. So I think there really, you measure, but then you have to be smart about the data too, and sometimes you won't know what conclusions to draw until you actually sit there and massage it a little bit and reflect on it.

DR. WEISSMAN: There are other measures like the mean, or the standard errors or the differences, as a proportion of the standard errors. I'm forgetting what they are but there are some other measures you can look at too.

CO-CHAIR ANDRULIS: Norman.

DR. OTSUKA: It's a clinically

significant difference, I mean a 20 percent difference in muscle strength may not be clinically significant in gait for example.

So rather than say statistically significant, I'd rather say a clinical significant. And that number may be very different from a statistically different scenario.

CO-CHAIR ANDRULIS: I think the order is, Ellen, and then, Edward, and then Luther and then William. William do you want to go first?

DR. MCCADE: Yes.

CO-CHAIR ANDRULIS: Go ahead.

DR. MCCADE: So I just have a fundamental question about this whole discussion in terms of I thought the purpose of our body was to create measures, as opposed to tell people who collect data how to present it.

Maybe I'm wrong about that and you can correct me if that's the case. But if we

do do what's suggested here in determining how we should present it, there's a very famous in statistics, a thing that's called Anscombe's quartet. And it's presenting data in a graphical fashion, that has the exactly the same meaning, exactly the same standard deviation.

And if you look at the plots that come from that, they're entirely different, and so I think we have to be very cautious about the interpretation data and how people will use it in order to make whatever point they're trying to make.

So is it, in fact, our goal to tell the end users of our metrics what they should do in terms of reporting their data, that's the question.

DR. HAVRANEK: I think we have to don't we? I mean that's what the, the devil's in the details here.

CO-CHAIR ANDRULIS: NQF, it's a question of guidance for application, you

know, is it NQF's intention to provide, not only like in this case measures, but also guidance on the application of the measures, interpretation of the measures.

DR. BURSTIN: Not very much other than clearly what the directionality is of a given measure being good or bad. Some measures have thresholds some measures don't, so not particularly.

CO-CHAIR CORA-BRAMBLE: If I understood your question correctly, you want to know what are we going to do with all this, it's the so what question?

DR. MCCADE: That's it, I mean the fact is we can talk about how to present the data but once I own the data, it's my data and I can present it in any way that I want to, without having to be necessarily responsible for saying whether it's absolute or relative, it all depends on what trend I like to show, is that not what happens when I own the data?

DR. BURSTIN: That would be true for your data, to do what ever you choose to in terms of quality improvement. But if there's a standardized measure that's been selected for use for public reporting or in performance, no you would actually need to follow the standards.

That's actually part of the reason for NGF, is to try to standardize the measures across different groups into these, whatever the case may be.

DR. MCCADE: Well standardized measures is one thing, but standardizing the way the data is actually reported is potentially another thing. And so I think this gets into the secondary aspect of it, and that's what I was questioning, is to whether we, whether that's our role.

This is my first NQF panel, so I have no idea what you typically do, but it strikes me as it's going to be hard to control.

DR. BURSTIN: In general the display of the measure is also outside of purview, but I do think a lot if the same principals apply that come through the measure is often times what comes to the display of the measure.

DR. HAVRANEK: Yes, when I first, saw this, that graph this morning it really bothered me because I wasn't sure what I would do about it.

And so I've been thinking about it a lot, and I've listened to a lot discussion here and my personal conclusion is to accept your proposal, that you show the trends, show the relative difference and show the absolute differences.

And the reason I feel that way are two things primarily, the first is that it's the most ambiguous, and on the surface that seems like a bad idea.

I think the most ambiguous here is the thing that stimulates debate the most,

and out of debate is going to come first, attention, which is good for this. And the second is it's going to come, we're going to get closer to the truth if it's debated.

The second thing is that this is the most transparent way to do things, and I think that transparency is becoming a really important value, and that if people feel like they're not getting the full story we all lose something. So you know, it's tough but I think you guys are right.

DR. CLARK: I will say a couple things, one is when you have a graph like this, at least visually, unless the curves are superimposed the disparity contains to exist.

And I guess the question really goes back to the issue that Colette raised earlier and that is what is a clinically meaningful difference.

And it might be helpful, if we could provide either some guidance on that, the

fact that should be addressed, because I think that's where we're caught.

There's statistical significance, and then there's the issue of what might be clinically significant or important. And I'm just, if there's some way to provide guidance for the provider or the individual who is going to have this data, as to what to do with it, rather than just leaving it to their discretion it might be helpful.

If there's not then maybe there's something we could say to them that would be helpful and that they need help interpreting it.

DR. WEISSMAN: Can I just comment? So I think we have to consider here, there are two issues going on here, one is looking at trends over time, and seeing if things are getting better or worse, that's one issue that's what this graph speaks to.

The other which I think is important is, when is a disparity, when is a difference

not meaningful? And that's another discussion which we didn't really address in here.

And I think even as, Norman, said, you know, what is clinically significant well if we had ten Normans in here I bet we would have ten different answers.

And it's not always easy to determine what's clinically significant, you often have to, and it depends on, are you talking about a process measure, you know, so if are you talking about some sort of a difference in function.

And in that difference in function does it make a difference in, it makes a difference whether that patient is an athlete or whether it sits in front of a desk, whether that patient's going to miss work.

So what's clinically significant has different meanings to different people, but that's not to say that it's not worthwhile and, you know, you get a bunch if clinicians

in the room that's what they want to know.

So I think that's an important discussion to have, so I guess I'm making two points, one is that, don't conflate the issues, this is trend over time, versus what's meaningfully different but, and they're both almost impossible questions to answer clearly.

DR. CLARK: May I just respond briefly? But I think that makes my point, I mean, if the experts can't make these decisions why would expect the reader or the interpreter to be able to make an appropriate decision as to what to do with the data?

DR. WEISSMAN: So in other words your saying, well I think one is, I mean, I think at some point there has to be a judgement on whether or not, and it is a judgement.

Let's say, take a single point in time, on whether or not a disparity is meaningful or not, and that is, I think that

does and I think you do need experts to weigh in as to whether or not it's important or not, clinically.

What does the evidence show, where's the rate now, is the rate starting at a high level, low level and so on. So that's at one point in time you make that decision and then over time you're seeing if things got better or worse.

But then I think you would also at that, you know, at the end point, you might say is there still a clinically significant difference or is there still a meaningful difference.

And again, you'd have to have that same discussion over time with all kinds of new clinical evidence that might come into play and so on.

So I think that you won't always have an RCT, a Randomized Controlled Trial, to tell you when a difference is meaningful or not, but I think that you'll have to rely

on expert opinion and you're going to have to make that.

And that people will have to make that distinction, in the same way that we make the distinction just, forget disparities, just about any kind of, you know, what's the optimal level of quality to make, I think, that it's all dependant on some research and some clinical opinion.

CO-CHAIR ANDRULIS: Marshall.

DR. CHIN: So I don't agree with, Ed, in terms of a very explicit recommendation that, you know, in general we recommend presenting trends absolute and relative data and then an example graph like this, or slight modifications.

And it's very clear in terms of explaing well, this is what we have in mind or, you know, if people want to, you know, step one step back, say at a minimum these data are available 3:06:46 consumer report cards and have a simpler display.

Fine but, you know, this more detailed data needs to be available also. But this way I think that if you don't have this type of explicit description there's a danger of what Bill said, that people misinterpret it. You know, in the way they want to do it, or present data in a way that is maybe nefarious.

You know, Bill, was right that, well, technically our charge ends at a certain point, but in practice this is probably at least as important in terms of the actual impact, and so that if we have a chance we should go ahead and make a recommendation.

CO-CHAIR ANDRULIS: Luther or Dawn, Luther is up.

DR. JACOBS: Sorry I have a very quick question, which is, are we going to make these blanket recommendations or are they going to be different recommendations by, as you said, how and they're somewhat

different depending on the measure?

DR. BURSTIN: I think that should be whatever the group feels comfortable doing.

DR. JACOBS: I'm just wondering if it might be measure specific, how we want to actually make our recommendations. Just to give more work to people.

DR. BURSTIN: NQF does also produce for the measure to offer some guidance on measure construction. And so I guess some elements of this that you think are important you would bring that to our Consensus Standards Approval Committee to see if they want to weave that into the measure of construction guidance.

In which case, it would be something that people would consider it, each time they're developing a measure.

DR. JACOBS: Okay, thanks.

CO-CHAIR ANDRULIS: We have a motion of sorts on the floor for approval around presenting trend information where possible

and providing an example of that information visually, much like it's presented in the report. And that there be some narrative related to mitigate that kind of path of least resistance.

What I want to hear, you know, some kind of narrative that describes the information presented, in a way that describes, use that term again.

Describes any trends, disparities or other findings of note related to race ethnicity and disparities. Some kind of narrative that would accompany the information rather than simply say okay here's the chart, good luck, God bless, and we're out of here.

Does that, I mean, in it's own clumsy way does that kind of capture kind of the three points?

MR. EPSTEIN: I agree, I think this is almost looks like an ethics caveat in the instructions, it's really a moral ethical

issue that we really need to come to terms with. And I think it makes sense, we've done this in the past, and this is a perfect time do it here.

CO-CHAIR ANDRULIS: Colette.

DR. EDWARDS: And the only thing that I would add is Ernie's idea, as part of the explanation, that if you have the results going in opposite directions you need to really think long and hard about what your next step's going to be. That's the big red flag going off.

CO-CHAIR ANDRULIS: We'll ask NQF to quote those words. "Thinking long and hard." Do you have sufficient direction on this, or do you --

MS. NISHIMI: Yes, I think we could review the transcript and come up with something that we would float by you again. And obviously MGH is free to, is free to alter their manuscript as they wish, but I think we have the sense of where the group

would like NQF's position to be.

CO-CHAIR CORA-BRAMBLE: But you know I think the message though is that the lack of clarity is because it is unclear. I understand that you would have to review the transcript, I'm not sure you're going to find the wisdom or the specificity that you're looking for. I'm just saying the consensus of the group, as I hear it, is that it is not clear.

MS. NISHIMI: But we would be able to convey in a concise manner the lack of clarity.

CO-CHAIR CORA-BRAMBLE: The lack of clarity? Yes, feel free.

CO-CHAIR ANDRULIS: Okay. We are on to Bullet 3, paired comparison versus summary statistics. Should a pair wise comparison using historically bench group as the reference point be checked to see if a positive finding from the summary reflects superior care received by the disadvantaged

group.

DR. HAVRANEK: Could you give us a concrete example to illustrate that?

PARTICIPANT: Joel?

DR. WEISSMAN: Yes. And this is, well I'll give you an example of this one and then I'll make a comment also. In using the BGV, Between Groups Variance, summary statistic, which is the difference between variance among many groups.

So the rationale for using a summary statistic is when you have many groups and you're trying to come up with a single number that summarizes whether there are disparities among that group.

And it turns out in, and a concrete example is in Massachusetts when we looked at this among hospitals, when we looked within a hospital certain hospitals the minority patients actually had better quality of care than the historically advantaged, meaning white non-Hispanic population.

And as a result that BVG statistic showed a disparity and in an incentive program they were not eligible for the incentive payment. So to me that's a little bit problematic.

And we pointed that out in a recent article that we mention here. So that's the specific recommendation that we're referring to when you have a summary statistic that summarizes the experience of many groups. It lacks directionality and you ought to do some pair wise comparisons.

I can also tell you that I'm not a statistician, maybe Sean is more familiar, there are about ten or 12 of these different summary statistics. They get very complicated. I had went to a presentation at Academy Health where Sam Harper was talking about using chaos theory to describe these things.

It gets very ornate, very intricate very quickly. And the biggest problem with

some of these summary statistics is they're not very transparent and you don't really know what's going on.

And so use with care, but again, I wish I had a better recommendation. Because if you've got three, four, five, ten groups that you want to track over time, you know, that can be pretty tricky.

This simple graph that I showed can get pretty crazy. So I don't know, Ernie, does HRQ use a summary statistic?

DR. MOY: No we don't, but Sam Harper was funded by the Federal Government. And we've encountered a lot of issues with it. We get this push to do it but on the other hand there's a lot issues dependant on weighting or non-weighting.

And so if you don't weight it, right, so each group is treated the same then you get very, very small groups having as much weight or importance as very, very large groups.

And it's then a function of how many groups you pick. Because the more groups you pick the more possibilities for having some kind of influence on the summary statistic.

On the other hand if you weight it by population size then disparities or small groups don't make any difference at all, because they're so small. So you get one of those possible side-effects.

DR. WEISSMAN: So Ernie's would probably give a much better representation of what some of the issues are and I think that's a different and also very important issue, the weighting issue.

CO-CHAIR ANDRULIS: Joel, in your example did you or anybody else go into that hospital and look at pair comparisons?

DR. WEISSMAN: Yes.

CO-CHAIR ANDRULIS: And what happened when you --

DR. WEISSMAN: Yes, that's when we found that the African-Americans were

receiving a better quality of care than the whites. So the summary statistics showed that there was a disparity.

But then we looked, we had white, non-Hispanic, black, Hispanic and I think Asian. And we looked at the individual hospitals and some of those hospitals the minority populations were receiving better quality of care.

CO-CHAIR CORA-BRAMBLE: Joel, may I ask what measures you were looking at?

DR. WEISSMAN: Those were the HQA, the Hospital Quality Alliance, core measures.

CO-CHAIR CORA-BRAMBLE: Okay.

MS. WU: I don't know if I completely understand this. But as far as I understand of this, it seems like there are a lot of cons to using summary statistics. And the only reason to use it is because you're aggregating small populations to make some conclusion?

DR. WEISSMAN: I would say there

are, close, the overriding reason to use a summary statistic is simplicity. Because if you can imagine making all these pair wise comparisons, you know, once against the other and then over time and many years.

Just imagine in your head, it gets pretty complicated pretty quickly. And if you want, like Sean said, if you want to come up with a single number this is sort of a nice way to come up with a single number.

But there's the directionality issue and then there's the other one that we didn't actually address in the recommendation which is the weighting issue, which is as Ernie described, and I'll just summarize again just very quickly.

So you have maybe four or five different groups and one summary statistic just says compare each group to the best measure or to the reference group and take the average of those differences.

Well that means that even like the

small group, people from Togo count just as much as the white non-Hispanics, if you do it that way. Or you can weight by the size of the population.

You know there are those sorts of issues that come into play as well. So they have a lot of drawbacks. But I'll tell you when an insurance company or health plan wants to present their data on many different measures it gets very complicated very quickly and so it's a real tension.

MS. WU: But just to be clear, your recommendation is to not use summary statistics and use the paired comparison?

DR. WEISSMAN: We weren't that explicit. We said you may want to use a summary statistic, but if you do you better do two things. One is check to make sure what the direction of the disparities are.

And if you believe that the historically advantaged population should be the reference group and they're the ones that

are receiving the worst quality care then you have to think about that before you report that a disparity exists.

And secondly, you have to be explicit about the values, which is coming up, on whether you think on whether you should be weighting the different population groups. And you thought the absolute relative issue was easy.

MS. WU: But it sounds like, just for the sake of speeding the conversation along, it feels like generally we should use compared comparisons and when you are using summary statistics it should be in very specific cases with many guidelines.

DR. O'BRIEN: One difference between reporting the paired comparisons compared to reporting a summary measure is that the paired comparisons are something that you can interpret.

A ratio, a risk ration or a difference, you know what that means.

There's probably no summary measure that you're going to come up out there that the average person will know what to do with. So the end result is a number.

And therefore I think they're only useful, if you're talking about summary measures, you're really talking about comparing this number that's calculated on this unit, maybe it's a population or a geographic area, and comparing it to another unit.

And maybe the reason for doing it is data reduction. There's just too many comparisons and it's cluttery, it's too much for anyone to process. You condense it into a single number. But that process of condensing is filled with information loss, subjective decisions or just really arbitrary decisions, so there's no single right way to do it.

And I certainly don't have any favorite summary measurement. So I certainly

agree with everything Joel has said. And I think that NQF, I believe, has a position paper, another framework for evaluating composite measures.

And I think you can kind of think of these summary measures as a type of composite measure and use a lot of what was in that prior work for this.

And I think that one of the recommendations from the composite measure committee was if you're reporting composite measures also separately report the individual components that went into the composite measure and that seems like exactly what's being recommended here.

CO-CHAIR ANDRULIS: It seems to me also there are a couple of issues. It's not just one issue. I mean take your example, Joel, of that hospital. I mean, I was looking at the overall disparities but there are different ways, black/white disparities.

I mean there are populations of

sufficient end in many places that you can do these kind of comparisons against the advantaged group on a chart. But for others, the smaller groups, then you run into small ends and you run into a variation.

And then you've got, you know, there are layers of complexity here in terms of how you approach the use of summary statistics versus the disaggregation to a point. Where there's potential benefit of looking between group comparisons.

So I think one of the charges for us is to consider when you're going to use the summary versus when to use paired and that there's no clear, again there may not be clarity on when to apply them.

You may want to apply them all, to the extent you can within large groups. And then when you get the smaller groups they may have to go into summary statistics of some sort. Or use, you referred to the methods that you talked about before, where you

aggregate, roll up. Ellen? Elizabeth.

DR. JACOBS: I had a question about often these sorts of summary statistics are used by NQF? Is it a common thing? I mean maybe this isn't used that often, it's kind of a moot point.

DR. BURSTIN: We actually do, increasingly, have had a lot of composite measures, but to me that's really more where the direction has been rather that a summary statistic, per se.

DR. JACOBS: So I wonder if given the concerns people have about it and it's not normally used whether it's something we need to continue to discuss?

DR. CHIN: Although what may happen is that if we end up coming up with a long list of disparity measures then it may become more of an issue in terms of simplicity.

CO-CHAIR ANDRULIS: Colette.

CO-CHAIR CORA-BRAMBLE: One comment. The one thing I would argue from a

clinician's perspective is the more complex we make this I feel the less likely that people are going to use it in a very tangible way.

So although some of it is almost an academic exercise. I want to make sure we keep it relevant to those, to the end user. And the end users they're not going to be statisticians, that's just not going to be the case.

CO-CHAIR ANDRULIS: So. Motion for lack of clarity?

(Laughter)

(Off microphone discussion.)

MS. NISHIMI: I didn't hear a silver bullet so let's take a break.

CO-CHAIR ANDRULIS: Okay. Ten minute break.

(WHEREUPON, the meeting went of the record at 3:25 p.m. and went back on the record at 3:45 p.m.)

CO-CHAIR ANDRULIS: Okay. We've

beaten a few horses so let's go on to the next one. Normative judgements about disparities measures. We're in Section 4.f on Page 33. Section 4.f, is that it?

PARTICIPANT: It's actually 4.e.

CO-CHAIR ANDRULIS: That's right. 4.e right above 4.f, conveniently. And the question is, what can be recommended to minimize normative judgements in the selection of disparity sensitive measures? Can objective criteria be identified in this regard?

And what we were discussing up here among NQF staff and the co-chairs is that I think what we're going to try to do here is instead of getting closure around recommendations per se, that the focus should really be on discussion of these points and trying to give staff direction for, then formulating, at least drafting a direction that we can then consider later, if I've captured that correctly.

MS. NISHIMI: Right, if something is clearly bright line and we can move on with a recommendation then I think we should go for it. But the last couple we've kind of gone round and round and round and beat the horse really dead. And I only want to beat things sort of dead so we can move on that way.

CO-CHAIR ANDRULIS: Yes and we've had a reconsideration. We're probably going to aim to close at 5:00, so it's been a long day. So we'll get as far as we can by about 5 o'clock. Okay. So Section 4.f, you have any opening comments, Joel, on 4.f?

DR. WEISSMAN: Not really. Although I'm wondering if the question should be to minimize normative judgements or if NQF ought to make the normative judgement in its recommendation. That would be the only comment I'd have.

And I would just, I guess as an example, I think an example helps. Talking about this Healthy People 2010 Report. And

the normative judgement that came into play there was around the summary index, which had a bunch of different groups.

And they decided that they would compare each subgroup to a reference group and that each one would count equally. And that was the normative judgement they made.

They made a judgement saying each group had equal importance, no matter how small the group was.

And you can make a different judgement saying you should weight those contributions by the population size, that's an example of a normative judgement. The other one has to do with should the least healthy group make the most progress?

There are a number of different things and the statistic that you chose can influence the kind of progress that you show.

And so I think this, you know, we were just doing our job in terms of raising these issues so that you'd at least consider

them.

CO-CHAIR ANDRULIS: This is actually an interesting variable and a measure that we looked at before. Where it may again require further drilling down on the measure and their reference point for normative may be up for consideration as well.

Because in the did looking at changes in the 100 largest cities and suburbs over time we picked low birth weight as a variable to examine. And we found that the trend at the point, this is back in the 90s to the 2000 Census Data, that there had actually been an improvement in some of the minority populations while there had been an increase in the majority of the advantaged population, so called.

And it wasn't until you looked further into the information about low birth weight that you realized that it had to do with multiple births that were occurring among women who were holding off on having

children later in life.

And that it was happening in places statistically that you hadn't thought of, like the suburbs that you were seeing this up tick.

So it's just kind of a number reference point that the issues around normative and perspective on numbers may need to be taken into consideration at another level as you try to identify those reference points for these populations. Comments on normative judgement? Sean.

DR. O'BRIEN: I think it's just important just to recognize that they are implicit in the way we measure things and the way we weight things. But I think this is posed as a problem and I'm not sure it's a problem.

CO-CHAIR ANDRULIS: Elizabeth.

DR. JACOBS: I just want to say I agree with Joel. I mean I think we should make recommendations about what's normative.

I mean that's what measuring disparities is all about, right, don't harm in making normative judgements. I mean, what's good?

CO-CHAIR ANDRULIS: So on the table, questions about normative judgements. Luther.

DR. CLARK: I was just going to say I agree and maybe just a statement that a reference should be made to minimize normative judgements.

DR. JACOBS: To what?

DR. CLARK: Excuse me?

DR. JACOBS: Reference to what, sorry.

DR. CLARK: The bias that is apparent in --

DR. JACOBS: I'm sorry.

CO-CHAIR ANDRULIS: So a reference to minimize normative judgements. Then what?

DR. CLARK: Inherent bias.

CO-CHAIR ANDRULIS: Inherent bias. Other tents?

MS. NISHIMI: Anyone have any other comments on this issue?

CO-CHAIR ANDRULIS: Ellen.

MS. WU: I don't understand it. Sorry, it might be too late in the day for me. Or effects of sugar. Can you define normative judgement?

DR. WEISSMAN: Yes. Value judgement. What do you think is important?

MS. WU: It's just so -- so then from what I understood Liz was saying, it's what the standard should be, like what the improvement should try to achieve?

CO-CHAIR ANDRULIS: Joel, mic.

DR. WEISSMAN: You could think of a number of different things. That's a good question, what do you want to achieve? Do you want to bring minority populations up to the white population or do you want everybody to improve? That's one value judgement, that's a value choice.

Do you think that all groups should

be equal or are you more concerned about the more populous minority populations, and focus on them.

I don't know, people help me out. I think if you just continue to ask that question, what do you want to achieve, that's a value judgement, normative judgement.

Probably should use values. And it turns out that the measure that you select, or the statistic that you select, can often, some people think reflects the value judgements of the people who select that statistic.

The CDC believes that all groups should contribute equally, no matter their size. And so in Healthy People 2010, they report a summary statistic where each group contributes equally. That's a value judgement. I'm just raising the issue.

CO-CHAIR CORA-BRAMBLE: Can we agree that it just needs to be mentioned, and you know, just move on?

(Off microphone discussion.)

CO-CHAIR ANDRULIS: Ellen, you have a look of puzzlement still.

MS. WU: Well I guess that's fine and it feels like there's a lot of these questions that we're theoretically talking about now which is going to happen really in the future when we look at the measures and the data, that, and I know we're trying to set some parameters or guidelines.

But that for particularly for this instance it feels like when there's a red flag in particular it maybe should get vetted somewhere.

If there's a decision to be made one over the other value in what statistic or what goal we're trying to achieve. It feels like there should be some process for input.

MS. NISHIMI: I think what my sense here is though we can't identify a single rule set.

MS. WU: No, right. For me I guess

it's more about process. That when this issue comes up there's a process for input. That there's a tangible example on the table. Does anybody understand?

CO-CHAIR CORA-BRAMBLE: I'm just trying to make it relevant, but not more complex than it needs to be. And I think in this case merely metioning it should be sufficient.

However, if there is a situation comes up then we should discuss and vette in more detail we certainly can do that.

But right now, at this juncture, I really think that just including it is sufficient. And I just want to make sure the group is in agreement with that.

DR. WEISSMAN: Yes, I just think one more comment. I think it's important as you're making these recommendations to just kind of keep this in back of your head and think of how it's going to play in Peoria.

You know, I always tell my kids we

live in the People's Republic of Cambridge, it's kind of like this room, you know, we're pretty well a liberal group.

But imagine if there was some sort of incentive program that encouraged the reduction of disparities. In order to do that some health plan actually made the quality of care for white non-Hispanics worse.

You know I think in certain parts of the country that wouldn't play so well. And in thinking about how -- And it's not just about measures. I mean we talk a lot about measure selection, but we're going beyond measure selection here I think.

We're talking about reporting of differences in measures and that's where the rubber hits the road. And people are talking about it.

This is a very charged area and so I just think that just maybe I agree with Robin, we just say it's an issue and move on.

CO-CHAIR ANDRULIS: Okay. It's an issue an we're moving on. Okay. We're going to skip down to the second last bullet to risk adjustment and stratification and that's Section 4.i.i., that's on Page 39 and 40 and 41.

And with the recommendation that performance reports stratified by race/ethnicity should not be risk adjusted by SES or other contributory factors. Instead should be further stratified at the date of permit.

Question to us, as you can read for the stratification, are race, ethnicity, primary language be performed when there is sufficient data to do so, should NQF review its policy of risk adjustment, vis a vis, inclusion of race/ethnicity?

MS. YOUDELMAN: Where are you reading that?

CO-CHAIR ANDRULIS: It's in the list on the agenda. It's on Page 4 of the Agenda.

I think -- Elizabeth.

DR. JACOBS: I have a question. Isn't this redundant with what we talked about before? Because we said we're going to pull out these measures on which we're going to ask people to stratify on race/ethnicity, right? I guess I need clarification on the question. Because didn't we already discuss that?

CO-CHAIR ANDRULIS: I guess I was wondering about this too. It seemed like we had a discussion saying as much as you can you want to do that. Want to stratify by race, ethnicity and language. So is this a kind of revisiting or it has another angle?

DR. BURSTIN: I think this has another angle, mainly in terms of risk adjustments for outcomes.

So one of the issues that comes up repeatedly is outcome measures come to us especially society's other developers, they have done their risk adjustment models, and

lo and behold the race/ethnicity is significant, as is gender as maybe SES.

They put them in their models, they bring them to us and we have traditionally said no race, ethnicity or language in our models. That we want to be able to stratify by those results. We don't want them buried in the risk adjustment model.

This comes up, I was just downtown at a CMS meeting, I was just asked this question 15 minutes ago. So this is an issue that comes up constantly when you're using measures ultimately to pay for performance.

So we, I think, feel like we're on solid ground but we thought it was an important issue because it keeps coming up and the stakes have gotten higher and higher, especially on readmission measures for example with a penalty coming in 2014.

There continue to be concerns people have raised that some hospitals in very poor communities, for example, may not do well.

But again it's more philosophical, but we thought it was important to bring it to this group.

DR. WEISSMAN: And just as a further level of clarity. To simplify, there are two levels of the risk adjustment question.

So first of all it should be clear, now we're talking about an overall measure of quality and should you risk adjust for race and ethnicity when there's an assumption that certain minority patients have poor quality of care. Are you going to be disadvantaging the provider.

The second level of that is when you're identifying disparities should you then risk adjust for socioeconomic status and other social determinates. That's a separate question. But you could adjust for all of those at the same time.

You know, Sean could probably tell you, you could throw everything into the model at once but there are different

conceptual issues for each of those two questions. Good luck.

CO-CHAIR ANDRULIS: Mara.

MS. YOUDELMAN: I'm confused this time. Can you just explain what NQF's current policy is? I've missed that, I'm sorry I don't quite get it.

DR. BURSTIN: I probably need more decongestant. So the current policy is that NQF precludes the inclusion, that doesn't allow the inclusion of race, ethnicity and gender in risk models. That indicates to developers if they measure that those results should be stratified rather than risk adjusted for those variables.

Exactly for the reason of being concerned of masking disparities. But because it keeps coming up we thought it'd be worth raising again.

And recently the conversation has shifted a bit to say, okay, not race and ethnicity and language, let's SES. Let's

talk income, let's talk geography, let's talk some of those other issues, which are other considerations.

MS. YOUDELMAN: I guess I'm confused about you're saying it's not masked, it's stratified by. And maybe that's because I just don't --

DR. BURSTIN: If it's in the model you can't stratify on it. You've adjusted those differences.

MS. YOUDELMAN: Okay. Got it.

DR. WEISSMAN: Helen, maybe ought to give an example of how it would be used. Stratified versus risk adjusted.

MS. YOUDELMAN: So what you're telling me is -- If you tell me I can't stratify by it later, I'm fine. I agree.

DR. BURSTIN: What does any risk model you can't stratify by it? So basically like some of the STS measures for example, cardiac surgery measures, have traditionally in the past, I don't think they do anymore,

included, race and ethnicity.

And so once it's in the model you can't afterwards, post hoc, show how we'd stratify it. Sean, you've thought a lot about this, do you want to --

DR. O'BRIEN: Yes, I'll try. I think my personal position is it really depends on the particular purpose of what you're trying to do with the measure and there's not a one-size answer, it's really context dependent.

Traditionally when you think about risk adjustment conceptually the question you're asking is what the outcomes of this particular unit looks like if the case mix was not the actual case mix that they treated but some other case mix.

And standardizing that case mix, and conceptually is to take factors that are present at the point of where the care episode begins and then you might generalize that to really think, really present before

the time of accountability beings, really, maybe.

And try to, don't adjust for factors that were within the control of the healthcare provider or things that you might want accountability for. Because then you would adjust away he differences that you're trying to measure.

But I think it gets, I can imagine exceptions to the rule with scenarios where, in gender for example, it is a reasonable question to ask, you really think there's something about women, or there's differences that aren't explained by any quality difference.

Anything that providers are doing in cardiac surgery sewing around the aorta that the size of the anatomy can really make a difference in terms of how long the operation takes and things like --

So there may be just inherent differences. And I'm probably saying things

that are wrong and there's probably people that would disagree on that.

But I think there is a perspective that says well, if you're trying to estimate this standardized difference then, sure, you could have gender in the model.

And that you'd still want to address these gender disparity issues. But maybe that's a separate topic. And I think, for me, the risk is we have normative goals, policy goals and separate just kind of being able to define and estimate some quantity.

If your goal was to estimate some quantity you want to define that quantity and estimate it the right way and you don't want to leave something out of your analysis that makes you not estimate what you think you're estimating. So that's where I'll stop.

CO-CHAIR ANDRULIS: Edward.

DR. HAVRANEK: I guess I would sort of echo a lot of what you said. So you may think that you're on solid ground.

But I think the ground is not as solid as you think it is because of the biologic and genetic issues that are inherent, certainly in gender and to some extent in race and ethnicity.

So the issue with cardiac surgery for instance is that women have higher mortality than men. Is that a disparity or is it related, as Sean was saying, on the fact that surgery takes longer, is technically more difficult because of body size?

I would maintain that it's the latter, that there is a real biologic difference there and by failing to adjust for gender you have created a disparity perhaps, where none exists.

You know I can make similar arguments with regard to race in congestive heart failure, in atrial fibrillation. Markedly different in African Americans.

And so the prognosis would

essentially be better so you would expect on a genetic or biologic basis there to be better outcomes in African Americans.

So equal outcomes is actually a disparity in that case. Failing to adjust gives you a false impression of reality. And then the socioeconomic, failing to adjust for socioeconomic position.

I think you've probably heard a lot of push-back that for some measures things like 30 day readmission, failing to adjust for socioeconomic position, puts safety net hospitals, hospitals that disproportionately care for the disadvantaged, puts them at a distinct disadvantage with regard to pay for performance and public reporting and all that stuff.

Because the determinates are social and are things that hospitals don't have control over.

And so I think that it might be reasonable to continue the policy as you've

had it but there needs to be come caveat or some statement that goes along with that that said we are aware of the problems that this approach is creating.

CO-CHAIR ANDRULIS: Okay. Marshall and then Elizabeth. Elizabeth then Marshall.

DR. JACOBS: I'm going to disagree. Respectfully disagree with that position. I used to work at a disparity institution, Cook County Hospital in Chicago, and really we should design our care to serve our patients who happen to be quite disadvantaged.

I think it's important to note I really appreciated in the report that there was a discussion about really not adjusting for these things because then you don't see the racial disparity.

I mean unfortunately these things co-occur, race and lower socioeconomic status in this country and higher socioeconomic status. And it's our job as healthcare providers and institutions to actually

address them.

And if we take away those differences by adjusting for things it doesn't give us an opportunity to recognize what are the determinates of what happened and then to address them.

And I would also argue that, and Bill's really the expert on this, but this issue of genes verus what happens in the environment, I mean it's not clear that people are necessarily going to be more disadvantaged biologically, that that distresses in their lives because they happen to be a different race and puts at a disadvantage biologically.

So I think it's really hard to sort that out but really it's our job, even if it were genetic we would still have to do a better job of taking care of them because they're at higher risk and we should know that.

DR. HAVRANEK: Can I respond to

that? I think maybe you misunderstood where I was coming from here. I think that the issue with, and you know I agree that systems should be designed to take care of the patients that they are responsible for taking care of.

However, the current systems are such that, you know, with pay for performance and things like that, that if you penalize people whose care is more expensive to provide to the patients that they are responsible for, you've done them a great disservice. You've hit them twice essentially.

And I agree that if you're at Cook County in Chicago that taking care of a group of patients who has trouble getting to appointments because they can't afford cab fare, et cetera, et cetera, et cetera, providing them with that stuff is going to provide them with better care.

But it's more expensive and it's

more resource intense. And the current system potentially takes resources away rather than adds resources in.

DR. CHIN: Yes, I think that actually what Ed and Liz and what you all wrote in the paper are all consistent. I think it's actually one of the elegant things about paper in that it shows us though that we can't just stop at doing the descriptive work with the disparity measurement.

But because of the implications downstream we have to include all of this implementation issue in the recommendations.

In other words we're saying don't stratify but don't include race/ethnicity within the risk adjustment models becuaes we want to stratify so we can see if there were differences.

Ed's bringing up the point that well it could lead to problematic issues like with this hospital compare program to the readmissions where you're penalizing the Cook

County's of the world.

And that was like another part of the paper where it's saying well you know we need to build into the implementation safeguards to prevent the rich getting richer problem.

So whether there are additional QI resources for the Cook County's or the idea about like risk adjusted reimbursement. So in other words once we start down this pathway we have to do it all in terms of, I think we do have to do it all, in terms of going through these different scenarios for implementation.

Otherwise you put down one thing which helps disparities but that can hurt unless they're all addressed. So I think we are on a pathway that I think we need to do, of being comprehensive, just like the paper was comprehensive.

CO-CHAIR ANDRULIS: Other comments? Elizabeth. Colette. I'm sorry, Colette, you

were in the queue.

DR. JACOBS: Actually I think Marshall covered it. I was just going to say that we do, on Page 8, really lays out a lot of lovely options I think and it's not immediately clear to me.

I'm just trying to think it through. If you embed it first, if you kind of do it after the fact, is it always going to be the case that you won't pick something up.

But the bigger statement really is related to what Marshall said in that we definitely need to pay attention to this one because it is going to be tied to dollars. And that's only going to be more and more and more true.

So we just have to safeguard against whichever choice we make so that the people who are taking care of the patients with the issue, get the resources that they need to do that.

CO-CHAIR ANDRULIS: Elizabeth, back

to you.

DR. JACOBS: I just want to say at the risk of Dennis asking earlier if I'm always the troublemaker and I said yes, I have a reputation in my family, and I may have it here.

But I just want to actually put forth that, I mean this is a really tough word to use but this is how racism gets institutionalized.

If we have a system that penalizes people, or organizations, for taking care of people who happen to be disadvantaged and then what we do is then, so we can't measure it, we can't measure that they're disadvantaged, we can't measure the ways in which they experience disparities because it would penalize these organizations.

That's the way this all stays quiet. And so I loved, I thought the ideas that you put forth, Joel, in your report were really creative about actually paying people more

to take care of poor people instead of disadvantaging them.

And in fact maybe paying them even more if you lower the risk of that institution. Could you imagine if there were a certain risk at Cook County and then you pay them more to take care of theme and then you also give them some incentive to bring it down.

I mean there's different ways you could do it. I think there's other models to figure out how we can. I just don't want the current models, which I think unfortunately ask us not to ask these hard questions and keep us from asking these hard questions or measuring these things. Not to let them determine what we do.

CO-CHAIR ANDRULIS: But I think what that does though, your comments also are kind of a bookend to, or work well with Marshall's comment, in the sense that there is an obligation then to make sure that this is

explicated in some way, shape or form.

Because if it were done, as Denice has said, kind of in this academic abstract of well, let's present these data in this way and let's not stratify on these indicators. Then those points could very well get lost in the discussion.

Therefore it kind of comes back to your point, Marshall, that if you're going to start down this path that you've got to follow through. You've got to also acknowledge these other elements. You know, Joel's good presentation notwithstanding on this point.

I think we may need to explicate on these other indicators to make sure that we're not institutionalizing racism further, but at the same time we're not dooming these providers to insufficient dollars and insufficient support to actually carry out these more complex social determinates and particularly related challenges.

MS. CUELLAR: I have a comment and then if Colette could bring Page 8 up I made a note here. I just want to reiterate, I believe someone had brought it up earlier.

One of the things I feel is missing in here, again, is we're talking about the provider from the standpoint of the practitioner and the patient.

But I think if some standards are going to be met by health systems we really have to think in the global standpoint, something that affects system-wide.

I'm thinking of the person who takes care of the patient 24/7, the nurses the respiratory therapists and it has to come from above.

So we really have to think more globally, much like joint commission does or CMS does, that we really have to set some standards that are more or less implied institutional wide.

And I think a lot of onus has been

put on the practitioner and the patient. But I really think to really drill down to where it's a win-win for everyone it has to hit everyone in that organization.

And the only way I think we can do that is to hit, that's what I feel is missing here is some systemwide, because it's the health systems that are ultimately going to start the continuum of care that goes on to the outpatient basis. And throughout this I felt some of those measures were missing.

CO-CHAIR ANDRULIS: And in some ways that point kind of reflects back to our discussion. And since what comes to mind for me also is, in the case of safety net organizations, one thing is primary care or emergency care. The other is specialty care, getting referrals.

And if you're going to penalize some of the safety net institutions for not having access to specialty care that has to be done, it shouldn't be done and it has to be taken

into consideration in the context of the systemic viewpoint rather than just simply the practitioner or the --

CO-CHAIR CORA-BRAMBLE: I do have a comment. This issue of risk adjustment, it is very nuanced and I do agree with your recommendations, Joel, in the paper, which you suggested in the terms of risk adjustment of payment to the providers which you had mentioned, I totally agree with that.

My concern is that if we say we don't want to risk adjust, but yet this group, this body, has no impact on the policies as it relates to paying providers, we're basically saying don't do this but we're penalized as providers on the other end if we take care of large groups of minority patients.

And there is no differential in terms of our payment, so we're saying don't do this but we really have no authority over what providers are going to get. So I'm

trying to put a very realistic framework.

You know in my group here in D.C. we are the largest provider of primary care services for underserved kids. So most of 90 plus percent of our kids are Medicaid enrolled children and 90 percent are minorities.

So the risk adjustment issue is a very real question for us. Particularly when we're not getting any added payments for all the extra effort that it takes to raise the bar for these kids.

So I'm just trying to be very realistic about what it means to say don't risk adjust, but on the other hand we don't have any say so in terms of how much we can pay for.

DR. WEISSMAN: Yes, I mean I guess my answer, and I think I'm sort of aligned with Liz's comments, but with real appreciation for what's going on. And I think perhaps one of the values, one of the

values of what this group can do is to really highlight this issue. And, I mean, people are talking about it, but I don't think they're giving it enough attention.

And, you know there is some lip service, I think mentioned here there was, you know, a letter from the AHA or something, but this, you know, I mean people have to change policies to address this.

That they have to realize that providers with disadvantaged populations have a tougher time and have poorer outcomes. And we don't want to institutionalize those outcomes and excuse them, right?

We want everybody, we want equitable care for everybody, but on the other hand, it's a resource issue and, you know, don't get me started, I think if anything the country's going in the wrong direction. I mean, you know, the gap between Medicaid and commercial insurance is getting bigger. That's exactly the wrong direction.

So I think to the extent that this group can really come out strongly about what the, you know, why risk adjustment is fair and yet dangerous.

CO-CHAIR CORA-BRAMBLE: I think it's important, I mean, that clear example is when we're caring for language minority kids, limiting this proficient you know populations that, yes we care for them.

But the issue is, in many places, there's no reimbursement for interpretive services. So yes, care for them, don't risk adjust, but incur the expense, you know, that is our reality.

CO-CHAIR ANDRULIS: Mary, then Ernie.

DR. MARYLAND: So if we accomplish nothing more than what the conversation has been in the last five minutes, we've done a really good job. Now if we could figure out how to make the policy follow our discussion.

And I think, the other piece is that

we have to be aware that, the more we fail to talk about the tough stuff, the more that we allow it to perpetuate.

As an example in many areas of Chicago, and I also trained at Cook County, we no longer have a trauma system, because we've allowed people to withdraw from it because it's expensive.

And so Lourdes and I, at the beginning of this meeting, were talking about the outcome that related to Congresswoman Gifford's being so wonderful. Because of her immediate access and sustained access to excellent care.

And so if there's a way for us to impact the availability, accessability and mandate to require excellent care, we have more than done our jobs.

DR. MOY: Yes I would be hesitant to put out a message that don't risk adjust period, but I do think that the way it's written is that instead of risk adjustment

you ought to stratify.

So stratify and then you can compare, apples with apples, you can compare each box, black with blacks or low income blacks with low income blacks. And then it's a fair comparison. I think that's a good message, to stratify instead of risk adjustment.

DR. WEISSMAN: Although I'd like to point out that that also has the danger of institutionalizing poor care. I mean, in spite of our recommendation you know people shouldn't think that that solves the problem, right?

Because imagine, I mean, I think you have to take the next step and understand how this could be used. And so if it's used for incentive payments, stratification in some ways does nothing different because you say oh, okay your poor black populations are treated just as well, or poorly, as everybody else is, so you're okay.

But at least it's more transparent and so there are, that's where the other policy recommendations that I made come into play. That we have to recognize, we have to be fair to the providers, but recognize that we need different policies to eliminate those disparities.

CO-CHAIR CORA-BRAMBLE: I would say that in terms of the entire report, whatever we submit, this is probably one of the most critical areas. And that it needs to be very nuanced, it needs to demonstrate to the public that we understand the issues, and that we've looked at it from every perspective.

CO-CHAIR ANDRULIS: And that they're tied, which you can't do one without the other.

I think we're going to Edward, and then Marshall and then to Sean and Grace.

DR. HAVRANEK: Just wanted to maybe clarify or propose that I think what I've

been trying to say is that, not that you should change the policy, in other words, I don't think that risk adjustment for age, gender, socioeconomic status, I don't think it's a bad thing.

And I think a policy should continue, I'm just asking that the policy come with an asterisk. That says that there are times where failing to adjust for these things actually is having the opposite effect to what we intend.

And I also agree with, Joel, that stratifying doesn't solve the problem with risk adjustment. So that's why I'm asking for an asterisk.

CO-CHAIR CORA-BRAMBLE: And I would add that it should be more than an asterisk, I think it's an important enough issue that it needs to be in the body of whatever, you know, not in a footnote, not as, you know, as a sort of sidebar thing, but imbedded in whatever in whatever it is we're doing.

DR. HAVRANEK: In bold, or italics?

CO-CHAIR CORA-BRAMBLE: Yes, there you go.

DR. BURSTIN: Actually to add to that, NQF does a sub criteria and specifically unintended consequences. So maybe it's important to actually tie these two together, stay grounded to the criteria as logical approach, which I like.

DR. CHIN: Lourdes point about teams and care coordination, just reminded me that, one thing that we should explicitly do, you know, in the next half day or so is to include, to some discussion about the measures that are going to be the ones that aren't on the radar screen now, but are important given the current organizational forums on the pike.

ACO's, medical homes, bundle payment, so things like the care coordination multi disciplinary teams and so this whole variety of different measures that, I know

that the wider point, which I was trying to think about, but for disparities, maybe they were particularly relevant in terms of some institute care coordination. Continuum of care, team best care, things that you mentioned, Lourdes.

CO-CHAIR ANDRULIS: That's also, it's an interesting point in the Affordable Care Act because, that the community health centers are funded but the speciality care, other links, regarding the benefit about upping funding for community health centers is not described in explicate, in great details and some medical homes and medical care but there's not that, there's that missing link.

Sean and then Grace, and Mara. Sean, go ahead.

DR. O'BRIEN: So I think my main concern about the risk adjustment position isn't that I don't agree with it 90 something percent of the time. It's that I, when you

put in a policy we don't know who's, we haven't really had time to think through, and innumerate all the different possible implications, follow it to its logical conclusions.

Is there any scenario in which it may lead to a kind of counter intuitive result, have the opposite effect of what we're trying to accomplish.

And my suspicion is, yes, if we had enough time to think about it there would be enough counter examples, problems that we might want to revisit.

I think one way I think about quality measurement is one thing measurement does is it, you can think about it from the perspective of what type of behavior and change are you trying to effect.

And that from the perspective of incentivizing and changing there is a real rational for not adjusting, not trying to institutionalize, racism.

But another thing you're doing when you're doing quality measurement is you're trying to measure something. And there may be research applications, there may be internal quality improvement, where you look at a result and you're trying to attach a particular interpretation with it.

You're trying to see how your results have changed over time. And if you look at your results, and it could be a situation where an institution has really put an intervention in place, to improve their quality on some disparity group.

And within that group, that population quality has really, really improved but the composition, the population composition of that institution changed over time.

And so now it looks like they are doing worse and that would just be kind of a mistake, it's kind of the data fooled them because they didn't analyze it in a way that

accounted for the change in their population concentration over time.

So I think an alternative approach instead of saying let's tackle this issue up front, at the phase of risk adjustment, is to go to other creative, out of the, you know, out of the box, or ways of addressing these issues that I think that probably everyone in the room agrees should be addressed and are important.

But just try to figure out are there other ways of doing it that don't meet the challenges of interpretation when you're using these measures to estimate quantities that may have a lot of different purposes.

MS. TING: Thanks, and I think that too, Denice's point, it's just very important, that we make it very clear, that we've been thoughtful about it and that we lay out the various implications.

Because you know in rolling out this kind of data to physician groups, from a

health plan, was actually provided dashboards where we didn't risk adjust per se, but we showed their performance compared to that of their regions, and how their, the physicians panel compared to the panels or the patient populations in their regions.

And we found, you know, that there are high performing groups was health disparities and there are low performing groups with no disparities and when we vetting it with the physician leadership of various groups for feedback, what the actual response we got back, was this is very valuably information.

It helps us think through what we can implement, but because the claims data may be flawed or the way we bill may differ from group to group, before you think of compensating that space on this type of stuff, you know, we really need to resolve some of these data feed issues and these adjustment issues.

So I think that that's a point that we will need to make, is that we can't even think about these different payment models until some of these issues get resolved.

But I can say that even without risk adjustments the physician groups seem to really appreciate this as a quality improvement tool. So we shouldn't let the risk adjustment or not discussion stop us.

MS. YOUDELMAN: My first is an apology that I'm going probably have to run before the conversation ends today to make it to daycare in time.

But the question that I have is on this stratification should be performed when there is sufficient data to do so. Who defines sufficient? Does NQF, does someone get to decide? How does that come into play?

Because that's, I think a big if, depending on how it's defined or how it's, and how much leeway there is.

CO-CHAIR ANDRULIS: Joel did you

have an idea behind that or are you punting that?

DR. WEISSMAN: I'm not punting it entirely, my foot's getting tired from punting a lot of these issues.

You might think of the, there are statistical tests you can do to determine what would be a statistically stable number.

For example in the premiere demonstration, you know, the rule of thumb is 30. You know 30, units of quality, and so there's probably is some minimum level at which you would feel comfortable stratifying the data by. And other than then you would want to roll up into larger categories.

DR. BURSTIN: Not on that point, but actually going back to Sean's comment, about are there other creative ways? I also just wonder if there is an option to consider balancing measures here?

Where, you know, if you're really concerned about the change of the case mix of

a population is there an opportunity to couple an outcome measure where there's not risk adjustment for these factors, with an examination of the population or something like that.

CO-CHAIR CORA-BRAMBLE: Give us an example.

DR. BURSTIN: People have talked about this, for example, when concerns have been about certain providers not taking on the highest risk patients, in addition to looking at the outcomes of care for those highest risk conditions, you also look at the case mix overall, to see if as, you know, are they funneling down to the lowest risk patients to get their better outcomes.

So it's kind of just a balance on case mix, just, I don't know, Joel had any thinking about that. It was just sort of struck me when Sean said, think creatively here, about other options.

DR. WEISSMAN: We've certainly

thought about risk selection as a way to improve you quality scores, and I think that's why taking in Marshall's comments, and others, is that we're sort of often thinking about just one type of policy, without trying to think of the bigger picture.

And that it's why in other, that I think that if you just try one policy lever then it's always going to be able to be gamed.

And that the best approach is to use a number of different policy levers to get at the different outcomes that you want to achieve. So I mean that's more of a policy question than a measurement question but that's how I would look at things.

CO-CHAIR ANDRULIS: Okay it's settled again, right?

MS. NISHIMI: Can I just ask Helen, is this getting the conversation where you need it to be in terms of the NQF policy?

DR. BURSTIN: Yes.

CO-CHAIR ANDRULIS: I think we're going to, we feel that there's a certain brain drain here. So, it's been a very intensive day, and I think it's time to perhaps move on to other things, so I think we're going to call it here and pick it up probably with this minor point of consideration of socioeconomic and other demographic variables tomorrow.

And that's one of the reasons I think we're going to call it now. And what time are we convening tomorrow, what's your preferred time? Eight?

DR. BURSTIN: 6:00 to 7:30 for breakfast, 8:00 for starting.

CO-CHAIR ANDRULIS: 7:30, breakfast, 8:00 starting, Starting at 8:00, breakfast at 7:30, so bon appetit.

(Whereupon, this meeting concluded at 4:35 p.m.)

D

H

Q

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Healthcare Disparities

Before: NQF

Date: 07-11-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Court Reporter