NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
February 18, 2014
12:00 p.m. ET

Suzanne Theberge: OK, we should be in the main line now.

Karen Pace: OK.

Operator: (Eugene) has joined.

Karen Pace: Hi, (Eugene).

(Eugene): Yes. I guess it is maybe afternoon.

Karen Pace: Yes, almost dark here. Suzanne, I didn't see the slide in for public comments so be sure to remind me as we get closer to the end of the call.

Suzanne Theberge: Yes. (Eugene), it's Suzanne. Can you see the slides on the webinar OK?

(Eugene): I'm just getting there now.

Suzanne Theberge: OK. Great.

(Eugene): What's the – my display name, I get, I'm sorry. It is connecting. Loading slides. Yes. I see slides.

Suzanne Theberge: Great.

(Eugene): While I was multitasking I was trying to write a response to David Hopkins.

Karen Pace: (Kevin), are you on the phone, here on the webinar?

(Zehra), this is Karen. Would you put a note on the webinar chat to remind people on the expert panel that we can use the dial-in to participate in the conversation?

(Zehra): I'm just typing up right now because – I'm just typing it up right now, Karen.

Karen Pace: OK and put the dial-in number and – OK, thanks.

(Zehra): Definitely.

(Kevin): So, this is Kevin.

Karen Pace: Hi, Kevin, this is Karen Pace, (Suzanne Fong) and we're just waiting for the committee – panel members to get dialed in and on the web.

Operator, do we have people still calling in?

(Elise Adams): Hi. This is (Elise Adams), I just joined.

Female: Hi.

(Steven Lipstein): (Steven Lipstein), I'm on.

Karen Pace: Great.

We'll wait another minute, we see people are getting on the webinar and probably still dialing in and then we'll start with the roll call.

Is Dave Nerenz on the line yet?

David Nerenz: I am.

Karen Pace: Hi, Dave. OK …

Suzanne, you want to start with the roll call and we can always circle back if people aren't in yet.

Suzanne Theberge: Sure, yes. Let's – OK, I'm just going to run down the list and just let us know if you're on the phone. (Kevin Foseli)?

(Kevin Foseli): Yes.

Suzanne Theberge: David Nerenz.

David Nerenz: I'm here.

Suzanne Theberge: (Dean) (Inaudible)? (Elise Adams)?

(Elise Adams): Here.

Suzanne Theberge: (Mary Badger)? Susannah Bernheim? (Monica Barrel)? (Maribeth Callahan)? (Larry Castellano)? (Elena Chen)? (Marshall Chen)? Mark Cohen?

Mark Cohen: I'm here.

Suzanne Theberge: Norbert Goldfield? Nancy Garrett? Atul Grover?

Atul Grover: I'm here.

Suzanne Theberge: David Hopkins? Dionne Jimenez? (Steven Leipzig)?

(Steven Leipzig): Here I am.

Suzanne Theberge: (Inaudible)?

Male: Good morning.

Suzanne Theberge: Sean O'Brien?

Sean O'Brien: Yes, I'm here.

Suzanne Theberge: Pam Owens? (Nina) (Inaudible)? (Pia Gustani)? Nancy Sugg? And (Rachel Warner)?

(Rachel Warner): Yes, I'm here.

Suzanne Theberge: Great, OK. I can see there are a lot of folks on the webinar portion that haven's connected to this phone yet. The folks that are on the webinar, you do

need to dial in to the phone portion in order to speak. The streaming doesn't allow you to speak. Has anyone joined who missed their name on the roll call?

Dionne Jimenez: Hi, Suzanne, it's Dionne Jimenez.

Suzanne Theberge: OK great. Anyone else?

Karen Pace: And operator, can we get a status if there are people still in queue to get into the call?

Nancy Sugg: Hi, this is Nancy Sugg.

Suzanne Theberge: Great.

Operator: Yes, there is.

Karen Pace: OK.

Female: (Thea Sani) sent a chat on via the webinar? I think she's on hold to join the line.

Suzanne Theberge: OK, great.

Karen Pace: OK. We'll give it a few more minutes then to let people get on the line before we get started.

Suzanne Theberge: And while waiting I'll just mention the usual housekeeping stuff which that if you're streaming this and dialed in on the phone, please turn off your speakers otherwise, we'll get some pretty bad interference and also, don't put us on hold because we'll hear the whole music and if you could use your mute button when you're not speaking that makes things much easier.

So folks who are connected to the webinar, to the phone line, if you could just let us know that you're on. That would be great. We see several folks that we're expecting are connected. Operator, (Nan)?

Operator: Yes.

Suzanne Theberge: We're seeing some chat messages from (Thea Sani) that says she's on the line but she is not able to speak. Can you make sure she's on the speaker line?

Operator: Yes.

Suzanne Theberge: Thank you.

Karen Pace: If there are any of the expert panel that are unable to be heard, please press star zero to signal the operator.

(Thea Sani): This is (Thea), can you hear me now?

Suzanne Theberge: Yes.

Karen Pace: Yes.

(Thea Sani): Excellent, thank you.

Suzanne Theberge: Great.

Karen Pace: OK, well, maybe we should go ahead and get started. Has anyone else joined that we haven't heard – that we didn't hear when we went through the roll call.

(Nina Ponce): Hi, this is (Nina Ponce) from L.A.

Karen Pace: Great, hi.

(Nina Ponce): Hi.

Karen Pace: We'll do a few introductory things and then we'll maybe do another check, Suzanne, just to …

Suzanne Theberge: OK.

Karen Pace: So just wanted to talk about the purpose of the call today is to really clarify some of the key questions related to the draft recommendation on the discussion last week and then through the continued e-mail discussions, we think that it's really important for us to clarify what methods are included as

when we referred to risk adjustment or adjustment. And so we want to just make sure we're all in the same page with that today.

Does the recommendation to address for some sociodemographic factors include statistical methods such as standardization and multivariable models or are we talking about only stratifications and we'll want to talk about why or why not.

And then also can multivariable models be used to accomplish the same purpose of stratification as to how. The next one is that we want to really get some clarification on it's about the purpose of making disparities transparent and the purpose of making correct influences about quality of care for accountability purposes accomplish simultaneously. So that's just certainly two ideal goals and we just want to make clear that to have some discussion that they can both be accomplished simultaneously in a performance measure.

And then what if any differences between clinical and sociodemographic variable, why different treatment. We talked quite a bit at the in-person meeting and the preliminary recommendations that that has a principle for selecting factors apply to both. I think there's some discussion about some unique things that we need to keep in mind with sociodemographic factors.

Then as time permits, we'll review the current iteration of graph recommendation. If we don't get to them, or through all of them, we will continue that the online poll.

So just to recap from last conference call, some of the key things that came out is that one of the things that was talk about is recommendations related to sociodemographic factors should be parallel to NQF expectation with clinical practice. We heard of minority physician, question mark, about only stratifying for sociodemographic factors that we – obviously, due to conversation that we needed to further refine the definition of adjustment and there were some discussions seem to be some agreement to not include the post score operation such as peer groups for comparison. Again we're going to go through that more systematically.

There still seem to be some confusion or maybe it's me. I'm unclear about where we're at with including sociodemographic factors to statistical method and about maybe some reasons to treat sociodemographic factors differently. For example, is conceptual pathway and analysis indicate that it's captured by the clinical factors or mediated by poor care practices and I think the other thing that came out is the need to clarity without being overly prescriptive.

So before we get into the discussion I'm going to just ask if Kevin and David want to make any opening comments and then we'll get into our discussion.

Susannah Bernheim: Can you hear me? The people are having trouble getting into the line.

Karen Pace: And who is that speaking?

Susannah Bernheim: It's Susannah Bernheim. I waited about six minutes on hold but I just heard from CMS that they can't get in either.

Karen Pace: OK, thank you. Would you check with the operator?

Suzanne Theberge: I'm sorry.

Karen Pace: Would you check with the operator?

Suzanne Theberge: Yes. We got some folks who are having …

Karen Pace: OK, Suzanne, I can't hear you.

Susannah Bernheim: I'm here. Sorry. Were you talking to me? I thought you're talking to the operator.

Karen Pace: No, I was talking if Suzanne Theberge.

Susannah Bernheim: OK.

Suzanne Theberge: OK. Karen, I'm back. I was on a separate line with the operator.

Karen Pace: OK. So, do you want to do a quick check in on others who have joined us? We've heard from Susannah Bernheim.

Suzanne Theberge: OK. So the folks that I have missing at roll call are (Inaudible)? (Monica Barrel)? (Maribeth Callahan)? (Elena Chen)? (Marshall Chen)? Norbert Goldfield? Nancy Garrett?

Nancy Garrett: I'm here. Nancy Garrett is here.

Suzanne Theberge: OK, great. David Hopkins?

David Hopkins: I'm here.

Suzanne Theberge: OK, great. Pam Owens?

Pam Owens: I'm here.

Suzanne Theberge: OK. And (Sue Pott)?

(Sue Pott): I'm here.

Suzanne Theberge: OK.

(Crosstalk)

Norbert Goldfield: This is Norbert calling. I didn't realize you were calling the roll again but I was here at the beginning, am still here.

Suzanne Theberge: OK.

(Larry Kathslinger): This is (Larry Kathslinger) can you hear me?

Suzanne Theberge: Yes, we can.

(Larry Kathslinger): Great.

Suzanne Theberge: OK. All right. I think we've gotten almost everybody, Karen.

Karen Pace: OK, all right, thank you. So I'll ask if Dave Nerenz and (Kevin Foseli) would like to make any opening remark.

David Nerenz: OK, Dave here. I just wanted to do a quick thing. First of all, very heartfelt thanks to the folks who contributed to the rich e-mail exchange that we've had since our call last Monday.

I think it's been very informative to both clarify some positions I think to help us find a bit of common ground but also to identify areas where there still are some genuine disagreements.

I think people have been thoughtful. They have been clear in their expressions. They've been respectful of each other and I think the entire exchange has moved us forward. So in particular, thanks to the folks who have been active in that.

(Kevin Foseli): This is (Kevin Foseli). I would certainly echo that. I think the task today is to now get into some of the operational details in terms of the methods going forward.

Karen Pace: OK, thank you. And let me just ask from the expert panel if there is anything that I didn't mention in the opening in terms of what we wanted to do – kind of what we try as the key questions that we needed some clarification that is factored into several of your recommendations. If there is anything else that you think we need to (inaudible) address on the call today.

David Hopkins: Karen, this is David Hopkins. I'm not sure you captured the issue around overly broad character of our recommendations, a point to all outcome measures versus only those that might be reasonable.

Karen Pace: Right, and we'll touch on that, that’s a good point, OK? I think where that comes in is starting with the conceptual model but we may need to actually have that all elevated to a level of recommendation. So, anything else that we need to be sure to address today?

Well, certainly, if something occurs as we go through this, we'll, you know, want you to speak up and as always we really want to hear all of viewpoints, so feel free to speak up.

So, the first thing that we just wanted to really get clear about what is included in risk adjustment. So if we start with a broad definition of the process of controlling or accounting per patient-related factors when examining outcomes of care and this comes from the (inaudible) risk adjustment, and I'm just going to make a separate note that we're going to talk about process measure separately in the recommendations.

So for those of you who, you know, are very interested in process measures as well, the intention is not to leave that out. But for now let's talk about outcome measures.

And last week, we talked about perhaps we should limit what we referred to under adjustment as the operations that are performed during the calculation of the performance score and separate out anything that would happen after the score is calculated such as the peer groups for comparison.

But we wanted to really make sure that we were on the same page about what was included in this when we use the term risk adjustment or adjustment. So, it includes statistical methods which would include multivariable statistical models, indirect standardization I believe would also be characterized as a statistical method, and direct standardization. And certainly under the indirect standardization, it could be based on clinical categorical models as well as individual variables. So, and then we'll come to stratification whether that's in or out.

But before I move on, I just want to clear if there are any objections that risk adjustment terms should include the statistical methods.

OK, hearing none then the next question is stratification. And here with stratification, we're talking about where scores for each provider are computed for each risk strata. So essentially each providers has a performance score for each risk strata.

Was someone going to mention something?

David Nerenz: Karen, this is Dave Nerenz here. I wondered before we move on to any discussion of stratification, if we could just pass the boundary of that first

thing because I think we're all pretty clear. Everything that you just listed there is examples of risk adjustment and I would have been surprised if there had been concerned objection.

But I'm wondering if for simpler edge, we've had a couple of examples that have come up in discussion about you know, does a simple weighted average when you've got two or more demographically define group. Does that qualify as a term adjustment or earlier in our discussions we talked about calculating observers, these expected ratios where the observed is simply on adjusted raw score but the expected comes out of some stratification and demographic variables.

The actual mathematical manipulation is very simple, very straightforward but it might also be called adjustment. I just am hoping we can clarify if, you know, does any mathematical manipulation of a score no matter how simple, how straightforward, how transparent called under the label of adjustment and is that acceptable to everybody?

Norbert Goldfield: This is Norbert speaking. I think there is a complementary issue that come up and I'm not sure if it's going to come up, is the issue that (Jean Nutro) brought up which is to say that they're – and I wrote about this in response to Susannah Bernheim this morning. But (Jean) I think hits the nail on the head when he said there are not infrequently.

So I think particularly challenging for issues in this socioeconomic disparity. And situations from a clinical model point of view and then, you know, there's not a bright line between the clinical model, you know, and the socioeconomic factor. So that's a separate issue.

But we're statistically, you know, any of the techniques that are specified where statistically it doesn't work and then the models, the risk adjustment model doesn't include that.

And I think (Jean) really hit it on the head and I very much agree and this comes from our perspective as developers that there has to be very, very strong, you know, case validity also for the measure, and if it doesn't work statistically of any of the techniques that you use in, but it does – from a case

validity point of view, it should be included. I think there has got to be very strong reasons why it should not be included.

The reason why that is important as I said in the e-mail is that in fact that there have been NQF measures that have been approved that in fact did not include measures or variables or characteristics that has case validity but then works specifically.

Karen Pace: This is Karen and I think that will come up more in our discussion about clinical versus sociodemographic variables. But I think what we're trying to get at right now is just what do we mean when we say adjustment, right and David Nerenz's question is, is there any mathematical computation or is it restricted. I think we have learned to expect it (inaudible) in to what happens with the standardization or multi-variable principal models. I'm not sure that this simple weighted average is, taught some others as well about what would be considered an adjustment …

(Kimberly): This is (Kimberly). So I just have a question building on that. So in terms of smoothing, taking it one step beyond the risk adjusted rate for instance, where does smoothing set? Right now to do a reliability adjustment, smoothing tends to align closely to volume where the smaller hospitals are basically set to the mean in the larger hospitals stay where their rates typically are. With that, what happens if we smooth through this socioeconomic status of some sort?

Karen Pace: As big point, I would say that would definitely fall into statistical methods that is probably one that we need to call out here.

Male: Yes. You don't have priority call risk adjustment on this list Karen.

Karen Pace: Right.

(Larry Kathslinger): This is (Larry Kathslinger). I wonder if we're using the best process for our two hours here. You know we can get very, very micro on each thing and the micro is important, you know, extremely important. But I guess – I'm not sure that – I guess I would like to hear one, two, three, four these are our main recommendations in just in general terms. They're not sure that there is

agreement on those. That I think we would still have in our mind, this is the map, right? These are the four things, or the three things or the six things, big picture that are going to be recommended.

So for example they are ought to be or not to be risk adjustment for SES right, or it must be considered or not considered whatever. See if we understand what our main headings are recommendations. And then within each one we can talk about the extent to which we want too in detail describe on how we think things ought to be done. But otherwise I think, you know, we're going to spend a lot of time on micro level things. I know we're going to get through everything or if we're really going to have an overview of what we're talking about.

Female: I agree with (Larry). This is (Inaudible).

Karen Pace: So this is Karen and I appreciate that and we can certainly move to the recommendations. I'll just say the reason that we wanted to delve in to some of these questions is because they come up over and over when we're talking about the recommendation. So for your example, adjust for sensitive demographic factors. The question is what does everyone mean by adjust because some includes statistical models, others think that that should just be stratification.

So, if you could bear with us maybe just for a little while and we can move through these maybe quicker. Because that would be OK for just a little bit to get a sense of these and maybe we don't need to get into them, each of these questions in great detail but I think they all have some implications for the terminologies and recommendations.

So let me just – I think just – and this is the important one on the stratification. We need this – consider that as part of adjustment or just call that out as a separate method and talk about it separately so then for we could say adjustment or stratification and not they're thinking that stratification would be one part of adjustment. So, I will just ask if there are any thoughts about whether we should consider stratification as a method of risk adjustment or

whether we should – just to be clear and talk about that as a separate type of method.

And what I mean by stratification of that, we've talked about some examples of stratifications is where each provider would get four for each (inaudible) stratum of risk. They should (inaudible) there might be quintiles of risk but each provider would end up for example in this case with five scores versus one overall four. And this could be, there's also risk by some socioeconomic factor or combination. Or, you know, it could be the clinical risk as well. But the idea is that each provider is at multiple stores.

And so – that a lot of times this is actually the first step in standardization where you're looking at, you know, what's expected for each stratum. But I guess the question is technically, whether this is considered part of adjustment or this it is considered an alternative or different method.

David Hopkins: Karen, this is David Hopkins, I think we have to consider the audience that – that were going to be sharing our results with. So for general audience, I think you really need to differentiate between stratification and risk adjustment. Most technical people would not think of stratification as a risk adjustment. Not that, I don't think.

Norbert Goldfield: I would agree with that, I would agree with David there in my document …

Female: Who was that that was speaking?

Norbert Goldfield: Norbert, I would agree. Sorry, didn't say, I would agree with David.

Female: Right, I'm sorry I didn't hear.

Female: OK.

Sean O'Brien: This is Sean. I would agree with that as well because of the way you've described it where each provider gets a separate corporate stratum. He's not – you're not adjusting something and I would just clarify if you're not going to include this risk adjustment that, you know, you be sure that you're not talking about standardization based on stratification.

(Elise Adams): I would agree with it, this is (Elise), I mean I think the part of the problem is it's there – it's almost never occurred that you have like a single indicator that's, you know, sort of explains all of the differences. And so ideally say the age was the one thing that mattered, stratification and risk adjustment would be more less the same. But it's almost matching the case and it's OK that we have some stratification that might occur as part of a risk adjustment strategy. Or even a way to explain the result of a multivariable strategy in terms of the level of risk where the risk is based on for example. But I think it is important to separate the two. Because stratification in and of by itself and it seems that there'll be very rare circumstances under which that would be – that would qualify as risk adjustment.

Nancy Garrett: Yes and this – this is Nancy Garrett and I just wanted to also agree and it think Norbert in one of his e-mails pointed out an important point which is when you're doing stratification, how you adjust, how do you set up the peer group? So for being able to introduce stratification in the state level and there's two safety net hospitals like in my states you know, end up with two comparison in that group. And even with inner group, you might still have a lot of heterogeneity.

So I think risk adjustment is really a different concept and stratification should be considered – it could be a part of that risk adjustment method. But I think it's a different set of method.

(Jean): And this is (Jean), I'll just second everything, with (Ted). Risk adjustment to me means that you have, you have mathematically transformed in some way. The observed performance of the – of the healthcare provider. How stratification is a display of results based on slicing and dicing based on some risk areas that you think are important to that particular outcome. And so I think they're quite different animals.

(Steve Lipstein): Hey, (Jean), this is (Steve Lipstein), let's address the question, how do you think of comparing an actual result to a – a an expected or a predicted result. It that – is that risk adjustment or something different?

(Jean): Well I think that's with David Nerenz just mentioned the – if you just – if you did – even a very simple observed to expected ratio. That would be – that would result in a value that – that's different from the observed value. The way it's done on health home health we compare the observed value with two predicted values of – so with national reference. And an agent to the reference or expected value. And so we can move it, move the observed value either up or down. Depending on how the agency does compared to the national standard.

So I think all those are method – methods of risk adjustment where you've adjusted the observed score. Stratification as a way of displaying comparative performance based on some other metric. Towards another measure, in this case maybe SES, or age or gender or whatever.

Male: Thank you that – that's illuminating I appreciate that.

(Jean): OK.

Karen Pace: OK so I think what I'm hearing when is that we should really talk about strategies as separate method and not assume that people will consider it part of adjustment.

(David Hopkins): Karen, this is (Dave) and when you say – talk about it as a separate method, where?

Karen Pace: Well, when we get to our recommendations, it would – it will be important for us to talk about whether, you know, work – the recommendation to adjust. It's really specifically about statistical methods of whether we – we should say adjust or stratify. So that's where we need to be clear, whether that's – that the recommendation is actually saying and what would be acceptable.

(David Hopkins): So I mean it sounds – I may have misunderstood but it sounds to me that the consensus of the last few speakers at least was that if we recommended risk adjustment, we would do so recommending using statistical method. Stratification would not be considered a method of risk adjustment. And then there's a question of whether and where and how to talk about stratification.

Karen Pace: Correct so the recommendation could be, you know, statistical adjustment or stratification as part of the recommendation or it would get into more specifics about then to use either approach.

(David Hopkins): And do we have any kind of consensus in this group about which of those alternatives the group favors?

Karen Pace: Well that's one of the key questions we want to get to today. So I'm going to finish up this one last thing about terminology and then we'll move into that.

(David Hopkins): OK.

Nancy Sugg: So this is Nancy Sugg. I just wanted to be clear in my own mind. So we're not saying we can't do both risk adjust and stratification. We could potentially have a model where there is some things is put into a multivariate analysis. But then we could also give the recommendation that – that this may need further stratification for certain uses of this method. Is that – or of this measure, sorry is that correct?

Karen Pace: That's correct. So – and all of those things are possible but we definitely need to make sure we're all talking about the same thing. And that's where I think – sometimes when we just use kind of a short hand adjustment. People have had different things in mind what that means to what all of the possibilities would be.

(Larry Kathslinger): Yes, and this is (Larry) again. I think this goes back again just to highlight the ideas having or a kind of a big picture conceptual map in our minds because if we said for example, OK you need to – you should consider whether risk adjustment needs to done or not. Explain to us why it should or not, by risk adjustment we mean using statistical methods. It sounds like if we said that for example, and then it sounds like any discussion or stratification would really be a discussion of how measures ought to be used, right? And I don't think we have ever come to any consensus on whether we expect measure developers to say how the measure or be used or not.

For example you ought to show things stratified in that right? Or whether NQF is thinking about issuing as – an e-mail this morning suggested I am

forgetting you sent it, my apologies. And it's come up at other times whether NQF should make some kind of statement about, you know, the general principles for how measures perhaps ought to be used.

Karen Pace: Right. OK so the last thing that we talked about in this that from the very beginning in this big adjustment bucket, and – and I think on the last call, it was clear that people did not link to include post score and when I say include I'm just talking about including in the terminology that we want to specifically talk about these as separate methods. So we'll have statistical adjustment, stratification and then the third kind of big area that we talked about is creating peer groups for comparison.

And the example of this is similar to what Med PAC has suggested for the readmission measure. But this is just a simplified display where when you're creating peer group, you're arranging the providers into deciles base on, you know, deciles or quintiles in this particular example. So that you will have quintiles of percentage of patient added below federal poverty level. You would see what the average for all patients to have for all the providers in that group is. And then an individual provider, you know, based on their percentage of patients below federal poverty level, would fall into a particular comparative groups and then their rate would be compared within that comparison group. And so this was a different way of looking at doing something to account for different case mix of patient served.

And so the discussion last week is that we shouldn't just lump it under this term adjustment that talk about it separately. And that's all I'm getting at right now not the specific recommendation on what should be done but just how we talk to them.

David Nerenz: Right, Karen, David Nerenz here. I think I like the terminology here because what we're looking at here I think in the past we've also referred to as stratification, I think we called it organizational stratification. But because of the continuing and future confusion just over the word stratification, it probably would be simple to use this term for what we used to call organizational stratification and then reserve the term stratification to refer to

transparent reporting of two or more groups within a provider. This was a way of grouping multiple providers.

Male: This is (Inaudible). I would strongly agree with that.

Karen Pace: So that's how we'll talk about this and I think that will help us when we get to have the draft recommendation. The next question we wanted to address and maybe let's go back so this question is really about should sociodemographic factors be included in statistical adjustment and so perhaps this is another way this really is getting at (Larry's) question of, you know, if do people agree that statistical adjustment is OK for sociodemographic factors.

Some of the discussion we've had is yes, yes, they need the principles with selecting risk factors that we've talked about. And some example of those present that started care. There is a conceptual relationship. There is an empirical relationship to the outcome. So it's not just a (yes) part launch. It is really is following good standard practices for how you create risk models. We've had some discussion about no stratification we only have acceptable approach, or know that stratification or the peer group for comparison are the only acceptable approaches.

We've heard possibly if for example if clinical factors do not capture the variants related to sociodemographic factors or if not, also correlated with full care or if you're unable to stratify. And related to this is can multi-variable models be use to accomplish the same purpose stratification. So I think this is a central question that – that we have some perhaps differences of opinion and we'd like to really hear from everyone about. I think the key question is, you know, what method or methods should be recommended to address, you know, accounting for variation and case mix on sociodemographic factors.

And I mean in this – before we open this up see if (Dave) and (Kevin) want to help this discussion in some way.

(David Hopkins): (Dave) here, I just would reinforce the point of if there's essential question of upon which many other things hang, I think this is probably been one and at least going by our preliminary polls, the various types in our in person discussion. This seems to be the one that, you know, strong majority of

panels in favor of adjustment. But some very clear – clear and well articulated are reservation by central, the reservations being primarily on this point of asking disparity.

So I suspect we don't gain a lot by just restating all positions but what we might see is – is there any, specific suggestions about possible middle ground or ways in which you know, the reservations about, you know, one of these you have still answers that other might be addressed.

And I think we got a pretty clear stance about where the group is, does this group, based on poll results and in-person discussion I think people have been quite clear in stating on what they think about it. And what would be useful certainly would be if there is some new additional coming down that we could find beyond what we like currently have.

Karen Pace: OK, good – good challenge.

Atul Grover: So this is Atul, my sense on looking at this are that, you know, right now it's not being done in any of these three areas for the most part and I think those case be made for, in principles to make sense yes including the risk adjustment. And when necessary either along with having it done with a mathematical model or instead of then you can organizationally stratify among common peers. But, you know, right now none of these seem been getting done. So I've – I've sort advocate for the strongest recommendation.

Pam Owens: This is Pam Owens. It's the question really about on the table just about statistical adjustment in this particular question and stratification is a subsequent question? Because I think that this fundamentally how strongly stated that it is part of risk adjustment for understanding of what our recommendation are?

Karen Pace: Right, I think, you know, this is the question I mean and that's why we want to get first talk about terminology because, you know, adjustment when – when we were originally talking about it included all three of those things, last week we kind of talked about post for the comparison peer group. And now we've

been stratification is another message but it's – its really thing the – a lot of the people don't think of it under the term adjustment.

So, it seems that the key question is that, you know, whether people want to say any of those three methods, so for example, the recommendation could be, any of those three methods, it could be if we underline to what NQF guidance and criteria for clinical, is that it could be statistical methods or stratifications and so we could, so particularly that model, but if really – the first question is about statistical adjustment but I think it includes all three, you know.

(Larry Kathslinger): This is (Larry). You know, one thing I'm not sure has come up in our discussions at this point is, what's going to happen with the committees that review and propose measures? And one concern that I have is if we recommend, well, it can be anyone of three say, is that we'll have those committees kind of having the same discussions and fighting the same battles that we've had here and first of all, that could not be a very good use of time but secondly my impression is that we have pretty sophisticated – we have some excellent people on this panel in our group and possibly better qualified than some of the review committees would be to thrash out these questions.

So just in general terms, I guess my advice would be does folks considerations will have any value, it might be better to try it and give fairly specific guidance about, you know, it's this – it's statistical adjustments rather than say, we could use extra Y or Z unless again there's a lot of people in our group would think that Y and Z are equally valid which it doesn't seem like there are.

Suzanne Theberge: If you count, Larry, I would say to that is – the point I was trying to make in my e-mail, this is Suzanne, I'm sorry which is that, we're trying to make guidance that would apply to any measure and measures really differ and I think we also potentially trap committees and developers if we give really prescriptive guidance that just won't fit every measure.

And so, you know, I mean the original language that came to this group was that NQF should requires sociodemographic factors be considered for suggesting outcome performance measures and require information about

what factor, how they were considered analysis and rationale for adjusting and not adjusting? And the, you know, in that recommendation, had everyone support from the get go, it creates a lot of transparency about this.

I think it does make sense to consider peer groups and stratification separately but I, there's been a lot interest in having the measures, it be more transparent the decision making about whether or not statically risk adjust for SES. And we have a recommendation at everyone supported that says, this adjustment approach that allows for measures coming forward like the one that you pointed out a bunch of times, you know, blood stream infections where it doesn't make any sense.

So I think – I think that probably smart to statistical risk adjustment first and to stick with the recommendation that's a full committee has already approved.

David Nerenz: So David Nerenz. I would like to speak into (inaudible) thank you specifically for the note this morning of pointing out the we should allow yourselves from really had some common grounds earlier and the ideas of pretty importantly that measures are quite different in their underlying cause and structures and how they use and how to have clinical and sociodemographic factors feed into them and that being sort of two simple and too prescriptive and maybe not helpful.

I'm wondering just a friendly amendment to what you just said that when we worry of the initial recommendation, we should consider, I was a little concern that that was too soft in the that measure developers, could just point out that they thought about it, but they've just decided on kind of on the whim not to do it so that would be fine.

I'm wondering if maybe we could find a little more congenial common ground if the recommendations had the spirit of or of any individual outcomes performance measure that the measure could or can be adjusted for sociodemographic factors, if a full set criteria were met. And if doing so, did not have clearer adverse effects of masking disparity. I realized we're

working in this semantic nuisance space of what's the difference between should we consider, can do it, shall do it, must do it?

What I have – I think we may have made a few more folks uncomfortable by a relatively simply stated message about should as opposed to slightly different one that says can which again, would be quite different from the current position.

Nancy Garrett: And David, this is Nancy Garrett. I like your suggestion and I think, you know we had the option A and Option B. And option B, we had NQF should require that outcome performance measures should be adjusted for – to see demographic factors unless that's (inaudible) presented for not adjusting and mean to me that's stronger than the can and I do feel like we should explore that a little bit to see where people are at. I would definitely advocate for the should. I think we, we need to take this strong position here to change what we're doing.

(Steven Lipstein): Yes, this is (Steven Lipstein), just to weigh in on what Nancy said, when we think about trying to accommodate measure developers in having them come forward with new measures, you know, one of the things that attracts people to these new measure is when the measure developer can demonstrate lots of variability in the outcome and if socioeconomic risk adjustment would eliminate a lot of that variability then it makes to measure that much less useful to organizations say like CMS. I think we need to increase at that bar.

I think CMS should not use measures with variability can be explained by sociodemographic factors and so for me, if we're going to get broad acceptance of measures or CMS is going to get broad acceptance of measures, they have to raise the bar on socioeconomic adjustment.

(Larry Kathslinger): This is (Larry), I think, if we say if you can, your people will say that's nice but I don't really feel like it. I agree with, you know, I think the recommendation we've had on – discussed a lot now, you should do it unless you have a bad, you know, unless you can give good reasons for not doing it. And just maybe have some discussions on what would considered good reasons. We're not being too prescriptive.

Male: What if …

(Larry Kathslinger): But then the question is, what is it and I'm not sure we have consensus on that still because I think some people want to say it could be stratification or just the risk adjustment or a peer group comparison and if I understand properly, there are other people that are saying, no, it really the peer comparison groups or stratification are not risk adjustment and so which should have been included in that if. So it's already two questions. What's it and are should we say, should do unless you explain why not or you can do it, to me I read, you can do it if you feel like it.

(Jean): (Larry), this is (Jean), if you look at the guidance that NQF has regarding what needs to be presented when you present a new measure, if they are probably prescriptive and fairly been very detailed with regard to clinical measures and that slide that just disappeared, as opposed on there that in terms of validity and reliability and so on and so forth.

The second – the probably – possibly if at some additional items that need to be considered by measure developers, so it clearly is not a win. When a developer puts together the defense of the measure that they are presenting to NQF that they try leave no stone unturned. So if you say that we must now consider sociodemographic variables that that is – that's that and we're held for the same criteria as, you know, as NQF currently has.

So I would argue that that the statistical risk adjustment models, whether it's a categorization or hierarchical or, you know, whether you're using a logistic or (OLS), are indeed adjusting the scores of the performance of the agency. If you want to go further if someone wants to go further and say, now we would like to display the results based on some other categorizations or stratifications, so be it. But that is – there's some simply displaying results is not risk adjustment.

Pam Owens: This is Pam Owens. The other sort of thought I would like us to consider is the purpose of the measure in the first place and I talked a little bit about this. The meeting which is right now, where NQF endorsement, the measure is supposed to be a quality improvement and not just strong end public

reporting. But we know that these measures because having NQF endorsement is, you know, something that people look for and they don't realize that NQF endorsement is that in public reporting. They do in fact use a quality improvement. They do it to detect where there are disparities in care from a health planning perspective. State agencies use these measures.

If we were to risk adjustment in, you know, in this statistical model, we would be taking away their ability to look at where the disparities occur and we need to do some targeted interventions in a particular way. So, again, it comes back to what is the purpose of the measure. I know what NQF policy is but if – if it's broadened to how are they actually use and if NQF changes that piece of that really strong end, that having a very definitive statement that it should may in fact undo this piece about quality improvement and targeting intervention.

Karen Pace: So, I think Pam that launches us to this question about the purpose – can the purpose of making disparities transparent and making correct inferences about quality of care for accountability be accomplished simultaneously. So, just to play out the improvement aspects even if the measure is adjusted for sociodemographic factors, that doesn't mean that it has no value in identifying areas what people need to improve on an outcome or process in general across the patient population.

They may need to drill down to see if there are particular sub-groups of patients for which there are problems versus all of their patients that that scenario plays out with any type measure but I think this is part of the (inaudible) of the issue that we're talking about is the concern about masking. And there is, you know, in some discussion about whether, you know, different purposes for example public reporting versus pay for performance should have different methods, but perhaps the real question here is that I think (inaudible) one of the workgroups at the in person meeting is that the purpose of making disparities transparent and the purpose of quality performance measurement is that where the distinction is and can you really achieve both with one performance measure.

And so I just mentioned, the performance score alone can't identify disparity but it's only possible with the additional information on you know, the various sociodemographic factors. So I think that's, you know, the question in terms of – again, how we proceed with these recommendations.

(Kevin Foseli): This is Kevin. Just to follow up on that point. You know, if the committee were to recommend, you know, statistical model type of adjustment, are there – is there additional information that people would want in terms of transparency in order to highlight the presence or absence of disparity and so what would those be?

Susannah Bernheim: (Kevin), can I go back I'm not actually directly answering your questions. I just want to go back to what Karen said in a way that was represented in the meeting, so that – because I think the language about masking disparity is not very compelling and isn't really the most important thing, I don't – I think everyone thinks that understanding disparity is important and then there's lots of value to understand (inaudible) but I don't think that we think that that is in general the most important purpose for quality measurement.

But the other issue is incentivizing improvement. I think this is what Pam was starting to get at which is that what happens to the incentive to improve outcomes for poor patients if you set the bar at how we're doing now which in many cases is inadequate. So, I'm much less concerned about masking disparities than I am about sort of what the impact on incentivizing improvement is.

And I'm really moved by what I've seen hospital is doing right now. I mean at Yale-New Haven Hospital, they've literally set up an arrangement so that there are respite bed for homeless patients when they leave the hospital to lower readmission rates. And it's something they should have done a hundred years ago and I just don't know, I don't know what the right answer is but I worry that if we said, you know, being – having homeless patient means that we expected you to have higher rate, there would be a different set of incentives to really improve care for those on (inaudible) populations. I think that's the balance that we have to be conscious of, it's less about sort of can you see the disparities or not which is important but not as important.

(Crosstalk)

Norbert Goldfield: I would just say, the reason I disagree with that …

Female: Who's speaking?

Norbert Goldfield: Norbert Goldfield, is that the Yale example, while very touching is probably a spot in the bucket of what needs to happen. And so I think that as one that as I said in one of my emails and it may have also to you Suzanne, that I think we need – just IME, you know, there are many different ways of doing it, so I don't want to talk about it, you know, indirect medical application authority but …

Male: Be careful …

Norbert Goldfield: Say again?

Male: Be careful of the IME there, buddy.

Norbert Goldfield: Why I only point on saying is that I may claim it that has been decreasing and I think we need to stop by putting this challenge that we have with socioeconomic disparity on the playing field as much as possible today, understanding that we're hoping that by – there will be a significant effort to even take on these patients.

So, if we or you know, the other flip side Susannah is that if we give them the incentive but of course it's all at the end of day is money, and no on is supposed to talked about but it's all money is that we want to encourage providers to take on this patient and it's – that's the bottom line. We can – we can decrease incentives overtime. That can happen but I believe that there needs to be – that are specified in details so I don't want to take up so much time. There's many mechanism to that.

Susannah Bernheim: But there's (inaudible) is a perfect one because, right that's an additional pain that the hospital gets. So if we think teaching hospitals need additional support, we don't say, "OK. We expect you teaching hospitals to just have

worse outcomes." We say they need additional support I mean that, I think it's an appropriate example for mechanisms that I think will be much stronger in accomplishing this goal. Like I think that people will be thoroughly disappointed if you just put a risk adjustment variable then how little it impacts the (inaudible) providers where if you set policies that protect the patient and providers, it is a much stronger approach and like the IME.

(Steve Lipstein): So, Susannah, this is (Steve). I think you gave a good example of the respite beds for the homeless. Let's assume the respite beds cost that provider more because you have to provide the beds and you have to staff the beds and make sure it's a safe and a warm environment for the patients.

On the value-based purchasing calculations, absent sociodemographic risk adjustment, you're going to punish that provider of the respite beds for higher cost produce the same outcome as another provider who doesn't take care of homeless patients. Where do you get the incentive's rate, you keep imposing financial penalties on people who care for sociodemographically disadvantaged patients.

And so there's a perfect example of where the cost to produce the outcome which is the true definition of value it going to be higher for the provider of care for homeless patients. I think you just made a great argument for why it's really essential that the demographics and the characteristics of the patient population that are present at the time the care begins have to adjusted for. Because they do affect not just the readmission calculation and expected rate of readmissions, but they also affect the expected cost to produce the same outcome as you can produce for a patient who doesn't.

So if you're economically or demographically disadvantaged.

Susannah Bernheim: But listen. If you put SES into this risk adjustment model, it would make no difference. We looked at this for mortality which is …

(Steve Lipstein): That's not true, the variable you selected just as a third gets a sociodemographic status makes no difference but that's because the variable you selected isn't the true indicator sociodemographic status.

Susannah Bernheim: We've done it with lots of variable. It's not going to make the effect that you want.

(Steve Lipstein): We disagree on that one.

David Hopkins: You know, this is David Hopkins. I don't work in the provider world but so I'm looking from the outside so forgive me. But it seems to me, that we have some providers that are specifically resourced to care for these populations with others that are not, or less so. Same with the communities in which they worked, some communities have support services for the people who were talking but others do not. I don't see how we can just throw all these patients into a risk adjustment model, irrespective of the circumstances that the provider and the community and say that we've done something good.

We should have higher standards for those providers that have the resources.

(Larry Kathslinger): This is (Larry). I think this is very difficult because we keep moving back and forth between what the measures should be and how the measures should be used and for accountability whether that be post reporting or pay for performance. And it's not because we're stupid that we keep moving back and forth. I think between and what the measure should be and how it should be used because it's hard to separate those and I think some of us might be or against for the particular measure depending on how we thought it was going to be used.

But as a specific suggestion about the measure what if we encourage and we can talk about what's the wording for that. Measure developers to show us how are the measure would be calculated not adjusted for sociodemographic factors and adjust it for sociodemographic factors and also encourage them to show us how this particular measure in their opinion lead to – it could be displayed in stratified form right.

And so three things, you know, show us how you would calculate the adjusted measure and the unadjusted measure. Show us how do you calculate a measure adjusted for SES and show us how you think things ought to be

stratified. And that's what we mean by dealing with risk adjustment of others, that we – some of us feel that the stratification is not a self risk adjustment.

And that's it, you know, then NQF, I think just think about or whether on how to make any suggestions in general about how a measure is ought to be used. But I'm not sure if it's the responsibility of the measure developer to be telling the public reporting programs or pay for performance programs for example. Whether they ought to be incentivizing improvement or absolute scores and yada, yada, yada. So just give to the adjusted and not adjusted as you would do it tell us how you do it and why? And tell us how you stratify and why.

Male: And (Larry), just to be clear so you're suggesting that this would then be also done for public reporting? Each of these three approaches you would have the adjusted, you would have the unadjusted and perhaps a stratification?

(Larry Kathslinger): Well I think.

Male: Is that what you're suggesting?

(Larry Kathslinger): No, not exactly I mean if that gets into use, right. So I guess I'm saying if you want to give us the measure, I think, what I'm thinking is moment if you want to give us NQF a measure, we think that you should, of course how you would calculate the measure, so the review committee how it would calculate the measure and not adjusted for SES. Then show us how you would calculate it adjusted for the SES, unless, or else, give us very good reasons why you don't think that ought to be done. Or can be done, must be a better language and also because this will differ for different measures and the developer's expertise could be useful.

Show us how you would recommend that measure with be shown – performance would be shown on this measure in a stratified way. And that's it. Then either each user of the measure decides for himself how to use to it but which is – well, that's going to happen, no ones has any control over that. But when NQF does have control over is whether or not to make recommendations about how much measures are going to be use, or whether it wants to have developers of a particular measure how they think it would be used which I think would be are not very feasible on a mistake.

Male: I think – I worry that, you know, if you have publicly reported three different sets of a measure that you really do lack control over how it's going to be used, you know, and their specifications to say, in a policy the risk adjusted number should be used. I think you run into all kinds of manipulation.

I do think it is worthwhile for institutions that are being recorded on to have access to their raw data because I do agree with Susannah that yes on readmission that has given hospitals an opportunity to work on at all the improvement on things that they didn't have a great sense for all those years for that.

I am also given the raw data that they wouldn't just say adjusted away, you know, you don't have to make any changes because even though we've got 50 percent readmission rate compared to the other hospitals taking of these patients, you know, we're right on target. So, I think having access for quality improvement to the entity that's being measured is the real important piece of information.

Nancy Sugg: This is Nancy Sugg. I would agree with that. I think having multiple measures would confuse everyone and it could essentially lead having that been misused in many ways. I guess I still want to get back to these measures are about quality and we talked a lot about worrying about masking disparity, but I really worry by not adjusting in some way for socioeconomic status, we actually mask quality, and that we lead ourselves actually not be able to do something about quality of care because we want to keep this broad disparity measure looked at.

My feeling is these are quality measures, they need to be adjusted the socioeconomic standards and state want to use them to look at disparities, we need a caveat that says these measures would not be appropriate to use to look at disparity and to be very clear about that.

David Nerenz: David Nerenz here, this is a question of Susannah. We're describing – this suggested that a hospital serving a low SES population of – we currently look at an unadjusted measure showing apparently bad performance and be

motivated to improve but within adjusted measure would relax or not work as hard and I think that's one view that things would be in important driver of the concern about adjustment.

But there's another view and I think (Steve) and others has expected this way that an unadjusted measure can be dismissed by providers as long and bias and inaccurate of an adjusted measure more likely to be accepted as valid.

And then my core question is, if in the adjustment measure there still this variability with performance so that hospitals truly doing a good job with the high risk population that show up that way in the adjustment number hospitals not doing such a good job show up with those numbers, it seems to me that the incentives for improvement are still present with the adjusted measures but now you perhaps have the additional of how do we factor of greater acceptance by the providers of the adjusted numbers is actually being valid rather than being invalid, does that resonate at all?

Susannah Bernheim: Yes, Dave. Since you directed it to me, I'll respond. I think that there are plenty of cases where adjustment is going to be the right thing to do. I never would want to proceed and saying risk adjustment is the wrong thing to do. What I think the two things that I think are important is that it depends a lot on the measure and the mechanism and the committee seems kind of sort of course it is always the right thing which worries me a lot because I think there are as many risk down that task as the path of, you know, never adjust before the problematic task.

So, they're my two main concerns, are that we build recommendations that allow for some flexibility and some confidence that would transparently, the public and stakeholders and committees will be able to understand the decisions that were made and decide whether they're valid. And my second piece is a real concern that people are feeling like risk adjustment is going to solve the safety net issue and I think it's not strong enough.

And the although our job is not to create a message about implementation, I think even risk adjusted measures, you may want to stratify for payment penalties that in my experience with our measures and our best efforts to find

a wide range of SES variables, we find that the impact on who is penalized or on provider scores of adding risk adjustment is a very small impact and that there are still real risk even with risk adjustments that there will be penalties that hurt fragile providers. And so my two things, to build flexible measure – recommendations and let's make sure to make a stronger comment about implementation that highlights the concerns of committee field.

Male: Thank you.

(Crosstalk)

(Steven Lipstein): This is (Steve). I liked exactly what she said and I wanted her to know that.

Atul Grover: I like it too, this is Atul.

Female: I agree.

(Thea Sani): This is (Thea). A policy follow up to that from – in that I agree wholeheartedly and it actually goes back also to an earlier comment. We need to separate somewhat the differences in outcomes versus the policy considerations of what constitutes equitable payment.

So for example in a very simplistic world, if there was a 20 percent difference in outcome between low and high SES, then we'd want to risk adjust that, but to close that 20 percent gap could double the cost. So, I mean the fact that we've risk adjusted for that gap doesn't mean that everything is even. And there may well be reasons why you need to give certain providers extra money. That's a policy decision that separate apart from risk adjusting for outcome.

Male: So this is not done outside of the penalty program we're talking about?

(Thea Sani): I think in the commercial health insurance world, it would be extremely rare. It's certainly something that we think about in the Medicaid world, but certainly in the commercial heath insurance world, it's quite rare. And in fact I think the tendencies as we've been discussing is that the more challenge providers, less of a bargaining position to get the money.

(Crosstalk)

Male: (Inaudible) to do that, so those that have a large proportion of under served populations.

(Marshall Chen): This is (Marshall). I just got a clinic, so I missed the first hour, but I think what I heard (Larry) just said and (Susannah) said, I think that really answer the answer here that I don't think we're going to find some of a single approach that (inaudible) probably is correct.

And so to me (inaudible) makes sense to me, in other words, (Larry's) point about presenting maybe just two different actually measures. One unadjusted, and then if you adjust to SES, what would that look like?

And then we have a separate document which if NQF approach to disparity which has not been done yet. But I think that would address the issues that (Thea) was just talking about of what different uses of these measures and when it is recommended to use different thing.

I think that's probably the way to do it though very explicit and it has a flexibility so we have measures unadjusted. I think that somewhat NQF is distorting the issue about like implementations and this is a hazy area about what is the scope of NQF, (inaudible) NQF of formally addressing it.

I think we do have to, and this is maybe the way to do it which is clean, so you know, unadjusted, adjust for SES. There's a separate document which goes into the issues that we've been talking about, the policies and what's it's used for, payments and all. But (inaudible) probably the way to do it (inaudible) which I don't think we made progress in terms of a single way.

(Elise Adams): This is (Elise), that's an interesting proposition. I'm just wondering, you guys envisioned that sort of when people reported with and without if there is a sort of accompanying explanation for what they think it means, because I think this issue of you know, what are we measuring is important particularly for particular variables and I'm just wondering if the idea is to just represented

without judgment or it's presented in the end also to provide some sort of justification for why one might be more appropriate than the other.

Karen Pace: This Karen Pace. And, I think part of the practically that we need to be considering and as the context of this project and some of the initial recommendation is that NQF currently endorse just measures that are suitable for quality improvement and accountability applications, which includes public reporting, pay for performance et cetera. And so endorsing a measure that's calculated both adjusted and unadjusted, I'm not sure what we would be saying about those.

I can see that as part of the information that an analysis that provided as we talked about in terms of submitted for review and evaluation but ultimately at least in a current, you know, per diem of what NQF endorses and how those measures are currently used, you know, that may change in the future but we are, you know, these recommendations will have some life for now and need to be applied in the current framework.

Nancy Garrett: Karen, this is Nancy Garrett. I think that's a good point and, you know, because of that and the way that these measures end up being used I really said that we need to make a recommendation one way or the other and, I mean, I think recommending different ways of doing it for reporting is fine but once the measures are endorsed, they really take on a life of their own and I think we need to figure out what methodology we're going to recommend.

(Jean): This is (Jean). I have a question for all the hospital people on the call. In the world of home health, we have a reports that go to the agency known as the CASPER reports and those reports provide each agency with its observed value as well as its a risk adjusted value, as well as third value that not even going to try to discuss which also does – it's an adjusted value. Do hospitals not have access to their own data in a national reporting system?

David Nerenz: David Nerenz here. I'll just quickly answer that and I'm sure (Steve) and others can response as well and it may vary a little bit by state. But I think for the most part, hospitals have access to their own data, you know, in the Medicare side it's in hospital compare. They get reports sent to you by CMS.

Most hospitals have their own analytic capabilities as well, and they work with standard measures. I think were the fundamental concern, you know, through all these discussion is that the measures that are available can maybe clinically adjusted, they're not socially demographically adjusted.

And, you know, using readmission rate of sort of number one example, the interpretations inference made about quality in the data that are available seems to be an inaccurate because (inaudible) presume the effect of documented effect of the sociodemographic variables independent of quality of care.

(Jean): But you do have access to your observed values, you do have access to a currently adjusted for patient case mix value and so I'm not sure what – what we're asking is that – or what I was saying I think is that we should also in addition to risk adjustment for patient case mix that we are also adjusting for sociodemographics in some way shape or another.

David Nerenz: I guess that would be a …

Male: (Steve), I think that's correct. And the larger health systems have more sophisticated resources probably to do risk adjustment but at least at the state level and I'm guessing most state hospital associations have an industry data institute similar to Missouri, we're able to collect readmission rates on all of the hospitals and do – and we're able to replicate the CMS risk adjustment methodology as well as apply local data for other kinds of risk adjustments.

And therein is where I think we get in to some of the discussion about that Susannah and I were having earlier about which variables to select because the variables that maybe available at the state level maybe different than the variables that would be available to people to national measure development and risk adjustment.

Male: And (Steve), you have access in your own data if you're a patient that is admitted, readmitted to but you don't necessarily know a timely way if that patient is readmitted to another hospital and you actually know there is such of demographic factors.

Male: Well, we do if they're readmitted to other hospital within the state of Missouri. We live in a cross boarder situation as you know close to Illinois. So, I don't know that that we have real-time data about readmissions that occur across the state boundaries.

Nancy Garrett: And this is Nancy Garrett. So, in Minnesota, we have an organization called Minnesota Community Measurement which is a multi-stakeholder organization that developed and endorse with measure and this is due as come up there and so, because of some of the feedback around the importance of risk adjustments we did create a risk adjusted measure that have a very rough indicator of SES, which is health plan. So, it divides the measure, it's stratified 3 sorry, not stratified but risk adjusted by commercial Medicare and Medicaid. And so it's so very crude kind of SES risk adjustment.

So the problem is that, that risk adjusted measure is on the page 150 of that 300-page report and the unadjusted measure is what you see when you go to the website it's what is on the first pages of the report. It's what all the private health plans are using in their P for P programs.

So, I mean, if we're going to make recommendations of have two measures I think it's important that we're saying that risk adjusted measure is what needs to be used when we're talking about moving money around.

Nancy Sugg: This Nancy Sugg. Again, I would not be in favor of multiple measures. I just think it will add to confusion. I think if we feel that socioeconomic status is an important thing to and somebody taking to account when you're looking at quality measure then that is the score that needs to be out there reported use for performance whatever, you can always back it up by saying if you want to look at more raw data, this is how you can access that for people who may want to look disparities or may want to look at the data in a different way.

But I agree with Nancy. I think that what you will see is the score and it will be used in a multiple different ways and so I think we need to have the score that we feel really represents a difference in quality. And I think that we've all decided that socioeconomic standard has to be in some way part of that score

and, you know, I go back to – I agree with Susannah said earlier. I mean, I think we do need to give some latitude to designers about how to do it.

I think there has to be a very compelling reason not to do it and that has to be a very well laid out and that we may need to do some statistical adjustments, and then on top of it you still may to do stratification because we know there's not going to be a perfect way to statistically get it all the variables of socioeconomic status. So, that I really strongly feel like this is a quality measure. It needs to represent quality and it needs to be one score.

(Mary Parker): This is (Mary Parker). I think if there is conceptual reason that you can show that SES affect this outcome just like kidney disease or whatever and just, you know, and a developer shows that then it should be the risk adjusted measure that's the public sees.

And I agree though that there should be flexibility in the developers on how they go about doing that and I also agree that maybe it may need to further stratify for other purposes but we don't put up measures that are unadjusted for medical factors and I think if there is a conceptual reason that SES affect this particular outcome then it should be risk adjusted measure that shown to the public.

Susannah Bernheim: This is Susannah. I just want to show that my comments earlier were clear. What I'm arguing before it's not that all measures should do risk adjusted but that the guidance that we originally started with which is it should be considered and analysis should be transparent and developers and steering committees would have all the information before them to make a decision about a given measure.

I think the simple patient level relationship between a risk factor and an outcome is not alone strong enough to argue for risk adjustment, I think you need to do more than that. And I think it's really informative to think about reasons for example. If all of the things we were just saying now and we are replacing rates of SES, I wonder if people would feel the same way.

(Mary Parker): I agree with you, Susannah. I'm not saying that everything should be risk adjusted. But as we talked about at the meeting there's some reasons that where rates really is a clear factor, like in my area and obstetrical outcomes clearly research, clearly has an effect. And so there's a framework for why it should be included, but if there's no clear conceptual reasons then I don't think it should be included.

So I'm with you, but I think – you know, so I don't agree that not everything should be risk adjusted for SES but you have to show that that conceptual relationship doesn't exist.

Karen Pace: So this is Karen. I've decided (inaudible) advance to – yes, principles that we've listed in terms of selecting risk factors. And it started out with this list as things that are generally considered when you're building clinical risk models and look at these in conjunction with sociodemographic factors and saw that they are also useful than looking at sociodemographic factors.

I think the first one – you know, these are not necessarily that every docs has to be checked but, you know, in building risk models it's an iterative process. But I think that the clinical – conceptual relationship with the outcome of interest is one of the key ones and I think was highlighted in terms of the panel thinking about how you look at sociodemographic variables as well.

So, I don't want to lose sight that even if the recommendations says, you know, includes, you know, risk adjustment for patient factor that there has to be some conceptual and empirical underpinning of what you throw into a risk model. And, you know, we can certainly raise it more to an elevation of a recommendation as people are more comfortable, but I think the idea is that whether it's worded, do it unless, you know, conceptually or empirically not appropriate or whether it says consider it and survive the justification. It's still seems to come down to following these principles and I'll ask it if that resonates with people, if that makes sense, or is that alleviates any concerns or raises more concern.

Male: All right, I think these make sense.

Female: I think it's a good principle.

Female: I agree.

Male: Karen, on the first bullet, is it any relationship no matter how weak or is it – is there something about the strength of that relationship?

Karen Pace: Well the first one is, first it's clinical and conceptually. The third one is about an empirical relationship. And generally, measure developers will – and we ask for this in the submission form, what are those criteria? So sometimes the initial criterion that measure developers use is there are certain level of significance or certain size of the correlation and I know (Jean), and Sean, and so as (Mark) might want to address that, but again that's something that the developer works on it and we ask them to provide to us, but I'll let others comment on that.

Male: And in our work because of the sample size with several million episodes of care, typically we have P values, 1 in 10,000 and I still end up with way to many risk factors. And so we try to (clear) those down in terms of the efficacy of how much variances is in indeed being predicted.

So, yes, I mean, they did at least a P-value of 0.01 if not, you know, one in 10,000 which is what …

Karen Pace: As you said, it could be the effect size however you're measuring that. That's more relevant to when you're dealing with large data set.

Male: But that the last bullet is a challenging one if you as – Susannah seems to have indicated any number of variables that she's tested that related to sociodemographic don't seem to have a greater effect on the model or add anything significant to the model, but do have strong face validity. As (Steve) indicated, you know, you would get certain amount buy in if you include those in the model. I think that's an issue that's beyond a pure statistical argument one way or the other.

Karen Pace: OK. So, actually the slide here that's showing up on the webinar is for some reason the strike three came out. But this is, and I'll point where they are –

where they should be which was looking at some discussion last week of how does NQF handle clinical factor as a criteria and simply be looking at the sociodemographic.

So, this is the exact language from the current criterion guidance to be forwarded about risk adjustment, and then there is a note 15 that goes along with that. If we change the language to be consistent with the idea of including sociodemographic factors, we perhaps would say is based on patient factor including clinical and sociodemographic factor that influence the measured outcome.

This information (inference) would be (stricken) because that's about – but not factors related to disparities in care. So, we would take that out and actually all of notes (15) would be struck, so that we would no longer be talking about differences related to disparities in care. And actually note 15 is where – if you see here on the last – at the very end of note 15 where our current guidance basically says that it is preferable to stratify measure by races and socioeconomic status rather than to adjust out the different.

So, this is the old or the current, I should say, criteria but if we move along the lines of making the criteria more balanced, it would one way would be, you know, the strike those causes that talked about disparity and just add (inference) behind patient factors and note basically indicating that it includes both clinical and sociodemographic factors.

So, I just wanted to put that forward and I don't know if that and, you know, help some people are raises concerns for other but I just wanted to, you know, work with the idea of, you know, looking at this parallel to requirement with clinical factors.

Atul Grover: OK. This is Atul. I think the way you modified to before make sense in striking 15 would certainly mash the intent of that change.

(Kevin Foseli): Karen, I like the parallel language with clinical factors and I wonder if some of the issues are related to identification of disparities could be addressed separately in the way they we're with race, ethnicity and language through

development of disparity measure by NQF. And taking that – taking that sort of out of here and addressing it separately for separate set of a measures which then gives us out of trying to do two things at the same time here.

Karen Pace: Right. And (Kevin), could you talk a little bit more about that or give an example of what a – what a measure that you would be thinking of in terms of disparity?

(Kevin Foseli): So NQF convened a committee a year or two ago to look at disparities on the measures. And they did that – it was done for a race and ethnicity and I think I'm trying to remember now if language was included or not. But it was primarily for some reason and – ethnicity where one seen large disparities or large gaps, so there was a big, you know, review of the literature done to identify those. Priorities were given where there was great impact and large differences.

And one could conceivably do the same with socioeconomic factors and add that to that in the sense – and in a sense, take it off the plate here and move it into its own separate category where it was done in more systematic way. So you would develop disparities measure were one – so for example, let take adherence to medications where you see large disparities by SES. So you would develop an SES-based disparity and then essentially stratify by the SES of the patient in reporting that out. And, you know, and hopefully it could be linked to some sort of pay for performance for, you know, lower SES patients who had low continuity in terms of medication adherence.

Karen Pace: Right. So – but in that sense then we're saying that we should stratify if it's disparity sensitive?

(Kevin Foseli): Yes, in essence. In essence, although – what it means is that, you know, it would be dealt with in a different – in a different realm. So the same process that one use to identify disparity-sensitive measures for race would be used for disparity sensitive measure by, we'll just say SES for the moment.

And then those would be reported and stratified but it would it essentially take it off the plate here. So that, you know, number 15 wouldn't be relevant and

there would be parallel language between clinical adjustment and SES adjustment.

(Marshall Chen): (Kevin), this is (Marshall), one of the challenges with that prior disparities (annals) is that if some might be kicked the can down the road. I mean, this is basically an empirical approach of where are there existing disparities among the, you know, few hundred conditions and, you know, (25) they have the most disparity which gives you this disparity-sensitive condition.

You know, where – what you're suggesting I think is a more fundamental approach of actually stratifying (inaudible) disparities that's acting from there. I think somewhat having sort of the stratification parts to look for disparities, this is probably the fundamental solution here. But we, instead of (inaudible) by try to come up with much more complex ways to approach it. and I think that we (inaudible) separate panels to look at SES measures, potentially would keep the can down the road with that approach also, compared to what I think that you're also suggesting of just looking where are the disparity and stratify that approach.

(Kevin Foseli): I mean and (Marshall), I agree with you that it is a bit of a kicking the can down the road at the time. If is this committee decides, "OK, we're going to go with a full statistical adjustment." Then in essence, there is no way to identify where are those disparities are. And I think one would need a separate process to do that if that's what we're interested in doing.

(Marshall Chen): I guess what I like to say overall was that people are thinking about quality and disparity are separate things whereas, you know, (inaudible) equity, you know, is a fundamental part of quality. And the process – a lot – some of the questions we've taken at the committee will really (inaudible) the whole equity issue. And so – and so we have a holistic approach, I think there are different ways to do it, you know.

One was a away that (Larry) suggested about NQF just about having unadjusted, adjusted measures, and then a separate documents that (inaudible) generally, or else, you know look at well, overall how would NQF think about some disparities and quality. But somehow, the approach I mean doesn't need

to explicitly to think about the equity issue, otherwise whether of this committee or the prior ones, it really is kicking the issue down the road. And the danger is that then that disparities are overlooked.

(Kevin Foseli): Yes, I certainly share that concern. What I worry about though is that we're – is that we have three major issues on here. One is the implementation issue that, you know, we've talked about potentially separating them out. One is producing what we think is more valid inferences, and the third is actually reporting out differences by SES related factors, and when and how that should be done. I worry that by mushing all these three together that we're struggling.

(Crosstalk)

(Marshall Chen): (Inaudible) NQF, it was absolutely the narrower approach from this panel if that could be fine. But only if NQF addresses more broadly, you know, approach thinking about equity and disparities. I mean, I think for some ways, some of us would be looking at the risk part of NQF is trying to address this issue, and using this particular panel to try to address the whole thing. But until we see NQF say, "Well, we will look at this holistically, you know, this by default is the way that was being done."

David Nerenz: Dave Nerenz here. Just a quick observation that's driven by a couple of comments probably, (Marshall) most recently, but others really I have talked about this problem of, you know, trying do too many things at the same time, trying to report quality accurately and have incentives for improving quality and identify disparity all at the same time.

I'm wondering if we gain a little bit of traction and consensus by pulling these things a part a little bit in the following way. If the fundamental goal for performance measurement particularly using NQF endorsement is to measure quality, recording quality, incentives to quality at the provider level then it seems, at least in my mind that the arguments in favor adjustment are pretty strong because of that adjustment. If it meets all the criteria we've listed here, takes you in the direction of more accurate, more informative quality measurement.

I don't think that the disparities all of a sudden just inevitably vanish and are impossible to see, meaning any analyst who wants within a hospital, within a state, within an academic setting, still has access to raw data sets. Still can look at disparity and readmission rates by race, can do all of these things. I don't know that an adjustment specifically for the purpose of the quality measurement and performance supporting renders disparities invisible for all other future purposes. And I'm not sure we have to worry if the desirable goals automatically clash.

(Larry Kathslinger): This is (Larry), you know, I think though again having listened to a lot advocates for socioeconomic for disadvantage people and discuss this issue, I think if there's one number and that's it. And that number is risk adjusted score, risk adjusted for SES and clinical factors. Then – and that score is all that's out there, it's always used, I do think it gives providers that are caring for disadvantage people but not very well. To some extent, it gives them an ongoing pass to keep that in, you know, to keep actually performing worse with the risk adjustment, you know, masking it.

And I think it doesn't really do much good for the public, the idea that an analyst in a hospital somewhere could look at this, him or herself if she wanted to or in a medical group or a nursing home. To me, that seems not a very – not likely to have widespread – be a very widespread solution. So I guess – I guess I don't really see the problem with saying, "Here is how this entity provide a – here's how this entity is scored, you know, overall." And this was just scored, this is how it compares to other people.

But some entities take care of sicker patients than others, or however you want to call for patient status. And if you adjust for that, this is their score, right. So, then, I think even an individual not even necessary that, you know, literally an individual can understand that and certainly people who work for provider entities can and people who devise incentive programs can. So, that way you have just two numbers and one number gives a sense of about how disparities are affecting that institutions performance or not.

You know, how to deal with – I guess, you know, we've then surrounded the stratification issue. I think most people seem to feel that that's not a substitute

for statistical risk adjustment but it seems like although that might or might not be something one might want to use for public reporting or pay for performance, it seems to me developers to show how it would be done not as a substitute for this statistical risk adjustment but to make it available for people who want to use the measure, you know, that doesn't seem overly complex to me.

Female: Yes, this is (Inaudible). I agree with that, I mean I'm the one that put the schools example out. And I think it's not – I don't think it would be that complex. We have two different quality measures that get at, you know, the need for improvement but also get a compositional effects that unduly burden providers serving underserved patients.

Karen Pace: This is Karen, I'm sorry to cut us off that but we do need to open the lines and see if there are any public comments. And then I'll talk about next steps from how we can continue to start to move forward.

So, Operator, would you just check if there are any public comments.

Operator: At this time, if you have any question or a comment, please press star then the number one on your telephone keypad.

You have a comment from Beth Feldpush.

Beth Feldpush: Hi. This is Beth Feldpush from America's Essential Hospitals. Thank you for the work of this committee and the comments that were made today. And I just wanted to reiterate a point from the perspective of those hospitals that really serve those vulnerable and safety net patient population that our hospitals and doing work and looking at their own data have really found important lessons when they're able to – capture some of the socioeconomic factors from their patients into their analysis. And we have found that it does not actually get rid of differences but can help for hospitals highlight where their opportunities for improvement are, as well as where they area really doing successful work.

And we think we've seen that growing in the published peer reviewed literature as well, you know, just some of the studies showing incidents of

pressure ulcers in minority patients versus majority white patients looking at prenatal care. So, we think that these instances are really the area and that it's not washing out any of the results from providers, but can actually really highlight some differences and give more insight into the quality there instead.

So, I highly recommend that the committee recommends exclusively or looking, specifically at a race and socioeconomic status in the quality measurement.

Karen Pace: OK. Thank you.

Operator: And there are no further comments.

Karen Pace: OK. Whoever was starting to make a comment, go ahead while kind of switch it to where we're going to go with next steps

Any final comments on the …

Sean O'Brien: This is Sean. If you turn to the slide that has, you know, review criteria. I would just made a comment that if our recommendation was to make the edits that involves striking out the 0.15 and – but not factors related (inaudible) and care and inserting the text you inserted in orange. I thought like that would be a recommendation that I'm very comfortable with and that (10) additional recommendations that I feel some risk about that they go a little beyond what I'm comfortable with and may have uncertainty about, you know, unintended consequences of those recommendations.

If some are those could be more than form the discussion, rather recommendations, you know, that was an approach that I would consider. What I would like about this change here is basically takes the issue and instead of going out with a limb and making a statement about, you should always do this when it comes to demographic, it just takes an existing framework and so the existing framework to the right way to be thinking about the issues. And you know if you really wanted to make strong statement, a strong statement to me that review panel shouldn't be approving measures that are misleading or confounded by case mix but there's not a lot of text here about what you should and shouldn't do in terms of

sociodemographic fashion but it think changing the text will have a good impact.

Male: Yes, I think the problem is that not all of us can support it. This is overly broad and it has the potential of not revealing the disparities that we're all concerned about. And a lot of our conversation today has been about use by individual providers of data for improvement and the only way you could do that right is to actually have a stratified result. So, you're comparing your underserved patients with other underserved patients.

Sean O'Brien: That's why I think (Kevin) and others suggested, you know, that we need to address the issues with, you know, these recommendations and original recommendations that were there for a reason and we need address these good important reasons but it's not the way you address it, that it doesn't basically create measures that aren't interpretable and that they're misleading, the way they address it is they're addressed directly by, you know, doing that with suggesting I just heard reporting outcomes separately for sub-groups of patients so you can make head to head comparison of different populations and there's other way to incentivize directly improvements in respect to disparity.

So, the accompanying recommendations beyond what to do methodologically could be about different ways to address disparities and performance measurement.

Male: I think we haven't heard Susannah's argument based on a lot of analytical work. That shows that this is not the right answer.

Karen Pace: Well, I think – OK, I think, you know, this is an area where we're going to have a difference of opinion. So, we will – basically what we're going to do is we will see another poll on (inaudible) at the recommendations statements, another poll the same way that you all said the opportunity to (just) recommended changes. Staff had been working with the co-chairs on he report that kind of lays out a lot of the discussion we've had, but this – the key thing we need to continue to work on is what are the recommendations that are going to put out for draft – for public comments.

So, we're still working on revising the timeline a little bit, to give us some time that this report together with the recommendations. We will probably have a short time for the panel to look at the entire report but we're first going to concentrate on the recommendations.

And, you know, we will have a 30-day public comment period, and during that time the expert panel can also continue to review the reports and especially the, you know, discussion facts and make some suggestions for changes. If they don't – aren't able to do in that short time turn around.

After the public comment period, the panel will review all the comments and look if these whether that indicates any revisions to the recommendations and those will be sent for feedback for approval.

So, that's where we are. We are working with (Dave) and (Kevin) to put together a list and we did have a revised list in this set of slides but we will look at those again in light of the further discussion we had today and put something out to you for your review and polling to see where we're at and we will definitely – for things that don't have strong consensus, we will be indicating the minority position because that is something that we'll want to really highlight the public comment.

So, (Dave) and (Kevin), I don't know if you have any additional suggestions for us or for the committee or thoughts about going forward.

(David Hopkins): So, (Dave) here. I think this is been really useful discussion and I, you know, it's always ideal bringing a group together if you can find (true) common ground that every single person can agree to. I think we still may try to do that although I do recognize that sort of on this fundamental issue of statistical risk adjustments, there are just some differing views.

I would think perhaps as a process step, we might try a couple of tweak variations and wording on this core recommendation about adjustment, just see if there are any sort of middle positions that we can – at least that the vast advance majority of the group rally around and I want to make sure that sort of we all understand each other where we're coming from.

For example, I think I understood Susannah's discussion certainly not blanketly opposed to our adjustments but neither is she in favor of sort of blanket the other way that all measures must be adjusted all the time. And although, a nuance recommendation is perhaps a little more complex to write, we may try of one or more of those variations and just see whether we can hit an excellent comfort level on something just tiny bit different than what we've got right now.

Karen Pace: OK. Well, and (Kevin), do you want to add anything?

(Kevin Foseli): No. I would just remind people of the tight timeframe so they get their comments in quickly.

Karen Pace: Right. So, we will try to get a poll up within next day or two and we'll let you know as soon we brought that up. And as always, I wanted to thank everyone for their great participation and engagement and, you know, thoughtful considerations of the thorny issues. So, I will sign off and thank you and we'll be in touch through e-mail.

Male: Thank you, Karen.

Male: Thank you.

Female: Thank you very much.

Operation: Ladies and gentlemen, this concludes today's conference call. You may now disconnect.

END