NATIONAL QUALITY FORUM

+ + + + +

IMPROVING DIAGNOSTIC QUALITY AND SAFETY IN-PERSON MEETING

+ + + + +

TUESDAY
JANUARY 10, 2017

+ + + + +

The Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Missy Danforth and Mark Graber, Co-Chairs, presiding.

PRESENT:

MISSY DANFORTH, Vice President, Hospital Ratings, The Leapfrog Group, Co-Chair
MARK GRABER, MD, FACP, President, Society to Improve Diagnosis in Medicine, RTI International, Co-Chair
JENNIFER CAMPISANO, JD, Attorney and Patient Advocate, Booby and the Beast Blog
MICHAEL DUNNE, PhD, Vice President, Research and Development North America, bioMerieux, Inc.
DAVID GRENACHE, PhD, Professor of Pathology/ Laboratory Medical Director, University of Utah
HELEN HASKELL, MA, President, Mothers Against Medical Error
CARLOS HIGUERA-RUEDA, MD, Vice Chair of Quality and Patient Safety, Orthopaedic and Rheumatologic Institute; Assistant Professor of Surgery, Cleveland Clinic
MARILYN HRAVNAK, RN, PhD, ACNP-BC, FCCM, FAAN, Professor of Nursing, University of Pittsburgh

MIRA IRONS, MD, Senior Vice President, Academic Affairs, American Board of Medical Specialties
NICHOLAS KUZMA, MD, Attending Physician, Section of Hospital Medicine; Assistant Professor, St. Christopher's Hospital for Children
PRASHANT MAHAJAN, MD, MPH, MBA, Vice-Chair, Department of Emergency Medicine, Section Chief, Pediatric Emergency Medicine, University of Michigan
KATHRYN MCDONALD, PhD, Senior Scholar and Executive Director, Center for Health Policy and Center for Primary Care and Outcomes Research
LAVINIA MIDDLETON, MD, Deputy Chief Medical Officer and Professor, Department of Pathology, The University of Texas MD Anderson Cancer Center
DAVID E. NEWMAN-TOKER, MD, PhD, Professor of Neurology; Director, Armstrong Institute Center for Diagnostic Excellence, Johns Hopkins University School of Medicine
MARTHA RADFORD, MD, MA, Chief Quality Officer, NYU Langone Medical Center
DAVID SEIDENWURM, MD, Quality & Safety Director, Sutter Health
THOMAS SEQUIST, MD, Chief Quality and Safety Officer, Partners Healthcare System (via telephone)
HARDEEP SINGH, MD, MPH, Physician Researcher, Veterans Affairs Center of Innovation and Baylor College of Medicine

NQF STAFF:

JOHN BERNOT, MD, Senior Director

HELEN BURSTIN, MD, MPH, Chief Scientific Officer

TRACY LUSTIG, DPM, MPH, Senior Director

ANDREW LYZENGA, MPP, Senior Director

VANESSA MOY, MPH, Project Analyst

CHRISTY SKIPPER, MS, Project Manager

ALSO PRESENT:

PAUL EPNER, MBA, MEd, Society to Improve Diagnosis in Medicine

KERM HENRIKSEN, PhD, Agency for Healthcare Research and Quality

DAVID HUNT, MD, Department of Health and Human Services

JEFFREY JOPLING, MD, Gordon and Betty Moore Foundation

CONTENTS

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

9:03 a.m.

MR. LYZENGA: All right, everybody. I think we are going to get started. Thanks to everyone for coming. We are really pleased that you could join our committee and participate in this meeting. I think we've got a good meeting planned out here in terms of agenda. We should have some good discussion. I will hand it over to our co-chairs to say a few remarks and welcome. And I think maybe Helen wants to say a few opening remarks as well.

So, go ahead, guys.

CO-CHAIR DANFORTH: Good morning, everyone. I'm Missy Danforth. I'm the Vice President for Hospital Ratings at Leapfrog and so excited to be here with all of you today. I worked with some of you in the past on this particular topic and other related topics, but very excited about the work that this committee is going to be doing over the coming year. So, thank you.

CO-CHAIR GRABER: Good morning, everyone. My name is Mark Graber. It is a pleasure to be here. This is a landmark day for me and I think everybody interested in doing something about diagnostic error. We have been talking about diagnostic error now for several years, but we have a weakest link, and our weakest link is measurement. Everybody says, well, how can we start to measure it? And it is amazing to me we have been practicing medicine for 2,000 years, and there has never been this kind of discussion, discussion at this level about, shouldn't we measure diagnosis? And can't we measure? And how should we measure? And how can we improve through measurement?

So this is an incredibly important day. I think what we're doing is really important work. And I thank you all for your interest in participating, and a huge thanks to Helen and the NQF and the NQF staff for hosting this event and to David for -- thank you David, for getting this whole thing going. So this is

an amazing event, and we really look forward to working with you.

DR. BURSTIN: I'd just add my thank you as well. Helen Burstin. I am the Chief Scientific Officer here. It has been a wonderful -- this is one of the most enjoyable committees to empanel because it was truly an amazing array of you who applied. So thank you all for joining us.

Do you want me to just go ahead and do the script for disclosures as long as we are at it? Okay.

What we are going to do next is we are going to go around the room and ask each of you to introduce yourself. Each of you has already completed a disclosures of interest form that you shared with NQF. So what we would like you to do as we go around the room. In addition to introducing yourself, say where you are from. Please indicate if you have got any disclosures. We recognize we picked you because of your expertise. Expertise does not equal bias or

conflict in any way. So, you don't need to share your full CVs, or we won't ever get any work done. Because I have seen them. They are great. They will take a very long time.

So just give us the highlights, and let us know if you have anything in particular, a financial interest, anything that what you think would be an important consideration for your other committee members to know as we have this discussion over the next couple of days.

And with that, why don't we begin with the chairs? And then we will walk around the room.

CO-CHAIR GRABER: So Mark Graber. I am an internist. I live in Massachusetts and California. I have no financial conflicts to disclose. My day job is through a group called RTI International, and I am the unpaid volunteer President of the Society to Improve Diagnosis.

CO-CHAIR DANFORTH: Missy Danforth, again, Vice President for Hospital Ratings at the Leapfrog Group. Nothing to disclose.

MEMBER CAMPISANO: Jen Campisano. I am a patient advocate, and I don't have any disclosures.

MEMBER SINGH: Hi, I am Hardeep Singh. I'm a patient safety researcher and a general internist at Baylor College of Medicine and the Houston VA Center for Innovation. I have received funding from AHRQ, the Department of Veterans Affairs, and Office of the National Coordinator for the work that I may be discussing in the committee, as well as I have provided expertise to several healthcare organizations for giving advice on ambulatory safety.

MEMBER HIGUERA RUEDA: I'm Carlos Higuera. I am an orthopedic surgeon at the Cleveland Clinic and their Vice Chair of Patient Quality and Safety. And I have a particular interest in research. I have a significant amount of conflicts because I receive funding from industry, and I am a consultant for Pfizer and CD Diagnostics.

I have a particular interest in

infection diagnosis, and I have been involved in their development of new technologies for diagnosis of infection.

MEMBER SEIDENWURM: Hi. My name is David Seidenwurm. I am a neuroradiologist at Sutter Health in Sacramento, California. I have been interested in performance measurement and appropriateness for quite a long time, and quality and safety is part of my day job.

My principal conflicts of interest are that I am a radiologist, and I also do a substantial amount of medical-legal work, much of which concerns actual medical error, in my opinion, or accusations of it that are not actual instances of medical error, in my opinion.

MEMBER HRAVNAK: I'm Marilyn Hravnak. I am a professor at the School of Nursing at the University of Pittsburgh. I understand I may be the only nurse on the group. So I am standing in for Florence Nightingale, I guess.

In terms of disclosures, I'm PI on two R01 grants from NIH. Probably the most

interesting relative to the work of this group is that we are looking at ways to enable nurses to better detect patient deterioration and instability at the bedside in order to prevent failure to rescue.

MEMBER NEWMAN-TOKER: David Newman-Toker, Johns Hopkins, Director of the Center for Diagnostic Excellence there. And I have disclosures related to federal grant support for the work we do related to diagnostic errors and improving diagnosis from the NIH, AHRQ, and private foundations, as well as some research equipment has been loaned to us by device manufacturers to help us with our stroke diagnosis research work. I have no other financial interest in those companies.

And I am an unpaid Board member of the Society to Improve Diagnosis in Medicine. And I guess my academic conflict of interest is I care about that.

MEMBER MIDDLETON: Good morning. I'm Lavinia Middleton. I'm a pathologist with sub-

specialty expertise in breast diseases. I am coming from Houston, Texas, where I am the Deputy Chief Medical Officer of the Medical Affairs for the Anderson Cancer Center. My area of expertise is in diagnostic errors and also hardwiring improvement and measures and accountability in annual performance evaluations of our faculty.

MEMBER MAHAJAN: Good morning. My name is Prashant Mahajan. I am a pediatric emergency physician and the Vice Chair of Emergency Medicine at University of Michigan.

My interest in diagnostic errors has been simulated when I first attended the DEM Conference five or six years ago. And my conflicts of interest are I am funded by the AHRQ for improving diagnosis in the pediatric emergency medicine realm.

DR. HENRIKSEN: Good morning. I'm Kerm Henriksen with the Agency for Healthcare Research and Quality. I'm trained as an experimental psychologist. Most of my working life, I have worked in the area of human factors

and systems engineering. I'm glad to be here.

DR. HUNT: Good morning. I'm David Hunt. I am a general surgeon, and I am Medical Director for Patient Safety at the Office of the National Coordinator. I am incredibly happy that this meeting has actually started. I am thrilled to see each and every one of you.

MEMBER KUZMA: Good morning. I am Nick Kuzma. I am a pediatric hospitalist in Philadelphia. I don't have any conflicts of interest to disclose, and I am equally excited to be here. So, welcome, everybody.

MEMBER HASKELL: I'm Helen Haskell. I am a patient advocate with Mothers Against Medical Error and Consumers Advancing Patient Safety. And I am also co-chair of the Patient Engagement Committee of the Society to Improve Diagnosis in Medicine.

MEMBER RADFORD: Good morning. I'm Martha Radford. I am Chief Quality Officer at NYU Langone Medical Center in New York. Other than my day job, I have no conflicts of interest,

which makes me probably living below the Manhattan poverty line.

I just want to mention also that I am an active participant in the AMA-sponsored Physicians' Consortium for Performance Improvement, who asked me to nominate myself, I guess, on their behalf. So I am representing them.

MEMBER GRENACHE: I'm David Grenache. I am a professor of pathology at the University of Utah and a medical director at ARUP Laboratories. It is a large national referral lab that is owned by the University of Utah.

I guess that's it. I am very happy to be here and looking forward to the next few days of work we have to do.

Oh, not conflicts of interest.

MEMBER DUNNE: Hi. I am Mike Dunne. My training, I am a country clinical microbiologist. I am a professor of pathology and immunology at Washington University School of Medicine, a professor of pediatrics at Duke

University.

My interests are human microbiome, next-gen sequencing, and metagenomic diagnostics. And my conflict of interest is that I am currently Vice President of DMRU, which is an infectious diseases diagnostic company.

MEMBER IRONS: Good morning. I'm Mira Irons. I am Senior Vice President of Academic Affairs at the American Board of Medical Specialties. I am also a pediatrician and medical geneticist.

Before coming to ABMS, I was the Clinical Chief of Genetics and the Training Program Director for the Harvard Genetics Programs at Boston Children's Hospital.

My interest in this area is both as a clinician -- over the years I was involved in genetic diagnosis of children with rare diseases and how neurotesting can actually helped with that. I have lived experienced in how that neurotesting can also lead to misdiagnosis with people who don't know how to use that.

On the ABMS side, my interest is how we can use certification, continuing certification to help in this area to reach 800,000 physicians.

MEMBER MCDONALD: Hi. I'm Kathy McDonald, and I am at Stanford University as a health services researcher, focusing on safety and quality in measurement.

In terms of conflicts of interest, I am also a non-paid volunteer at SIDEM on the Patient Engagement Committee, co-chairing that. I have funding from AHRQ and the Moore Foundation, and both of those are related measurement and quality and safety. So that is where I come from.

DR. BURSTIN: Great. And I know we have got Tom Sequist, who is sadly in Boston with a fever, with us on the phone. Tom, can you do your introduction and disclosures?

MEMBER SEQUIST: Hello, everyone. I'm sorry I'm not there with you. My name is Tom Sequist. I'm the Chief Quality and Safety

Officer for Partners HealthCare in Boston. And I don't believe I have any conflicts but I am really excited to be part of this. Thanks.

DR. BURSTIN: Great. Thank you much. And we have two guests joining us today, if you would like to introduce yourselves.

MR. EPNER: I'm so glad I don't have to do disclosures. I'm Paul Epner. I am the paid Executive Vice President of the Society to Improve Diagnosis in Medicine and the Chair of the Coalition to Improve Diagnosis, which is a 32-organization coalition focused on this problem, and NQF is, hopefully, today signing the paper to be -- actually, we have 30 now. So hopefully you will be our 31st. And Gordon and Betty Moore are going to be our 32nd. So, we are up to 32 organizations.

DR. JOPLING: Good morning. My name is Jeff Jopling. I am a fellow at the Gordon and Betty Moore Foundation and a general surgery resident at Stanford.

I have no conflicts of interest

because I am here for an alignment of interest. The Foundation is interested in supporting the achievement of diagnostic excellence. And in general, the Foundation is extremely passionate about measurement and then, specifically, in this issue, doubly passionate as well so is thankful to be here.

DR. BURSTIN: Thank you so much for joining us.

So one quick comment. Thank you for those disclosures. Obviously, we have assembled an amazing cast for this discussion.

A couple of you have indicated on here representing X. You are actually not. You are representing yourselves. You may have been nominated by somebody else, but you sit as an individual. You don't necessarily bring an organizational perspective. You can bring a perspective, but you are not speaking on behalf of anyone else.

And lastly, as you heard those comments and disclosures, it is also an

opportunity if you have any questions of anybody who has given their disclosures, this would be an opportunity -- I will leave an open opportunity for anyone to ask questions. But really, through the course of the next two days, if you have any concerns that someone is potentially being biased or trying to -- in this instance, it's different. It is not about endorsing or approving measures. It is really more of a conceptual piece. So it is less of an issue, but please feel free to come forward to us or the chairs. It is always easier to kind of deal with those concerns in real-time, rather than after the fact.

So I will just stop there. If anybody has any questions of each other based on disclosures, otherwise, I will turn it back over to the chairs.

Oh, actually, I also want to have the staff introduce themselves if nobody has any other questions.

Great. All right, well let's do staff intros, if we could. Go ahead.

MR. LYZENGA: Hi. I'm Andrew Lyzenga. I'm a senior director here at NQF. I have been here since about 2009. I have worked on a number of our consensus development projects and MAP projects and a couple of these types of framework projects as well. I am very excited to be working on this.

MS. SKIPPER: Good morning, everyone. My name is Christy Skipper. I am the Project Manager on this project. I have been with NQF for almost a year now. I can't believe how quickly time has flown. But in addition to this project, I have also done a couple of measure-endorsement projects related neurological and surgical topic areas. And I am currently working on an infectious disease project.

But I welcome you all here, and I look forward to the next two days.

MS. MOY: Hello. Good morning, everyone. My name is Vanessa Moy. I am a Project Analyst here at NQF, and I have been here for about four months, and I look forward to

hearing your feedback. And I am very excited to learn more about the NASEM framework and just hear everyone's feedback. Welcome.

DR. BERNOT: Hi. I'm John Bernot, also relatively new to the NQF. I have been here for about four months myself. I am a family physician, and I am really interested in this project from both a clinician perspective, as well as a patient perspective.

And I just want to thank you all again for making the trip here to Washington in the cold.

DR. LUSTIG: Hi. I'm Tracy Lustig. I am also a senior director here. I had a first career as a podiatrist. I, after that, was at the Institute of Medicine for 12 years. I was there when the Report on Diagnostic Accuracy was happening, but I was not working on it. And then I came here about nine months ago. So new but that new.

DR. BURSTIN: Thanks. I will turn it back over to Christy.

MS. SKIPPER: Okay, next slide. All right, and we have kind of already covered this. Keep going.

All right, I will start with the project objectives. So we have been asked to convene a committee to develop a measurement framework to help identify and prioritize measures of diagnostic accuracy, diagnostic error. And we are trying to identify and conceptualize structures, processes, and outcomes related to this topic. And a measurement framework is a tool that will help us do that. The framework will help us organize our thinking and have a shared understanding of the conceptual structural for doing our work over the next few months.

As you all know, the National Academies of Science, Engineering, and Medicine did a lot of expansive work for thinking about diagnosis and diagnostic accuracy, and it was created with a broad acceptance and consensus among the healthcare stakeholders. But we do

have an opportunity to refine this framework just a bit for our purposes, but we do want to retain the overall framework. Next slide.

So this slide restates a bit of what I just said, clarifying that we will be identifying a measurement framework that can help us identify measures or concepts, identify any significant gaps, and set priorities for measurement around diagnostic accuracy.

In this project, we will not be developing a new conceptual framework, nor will we develop any measures or endorse measures.

So over the next two days, we will hear more about the NASEM framework, and we will hear from other professionals in the field this morning about other models or concepts related to the topic. On Day 2, we will do a deeper discussion of some of the measures that we have already identified and then also break you all into groups for some group work to help us identify any measures, measure concepts, and domains. Next slide.

Just a couple of ground rules. If you would like to be acknowledged, we ask that you turn your tent card into the vertical position, and one of the co-chairs will call on you to speak. And please always use your microphone and lean in like I am doing now so that you can be heard. The meeting is being recorded. And also note that only three microphones can be on at a time. So once you are done speaking, click the speak button so that the red button turns off.

And then also throughout the day, just please openly share and respect differing points of view. There are no wrong answers. We appreciate every viewpoint that is represented around this table, and we definitely want to hear from you, but we ask that you also avoid dominating the discussion, and please allow your colleagues to be heard.

So I will stop right there. Are there any questions or anything that any members of the team would like to add? If not, we will jump right in to -- I'll stop right there.

Are there any questions? Okay.

MR. LYZENGA: And just to sort of set the stage here, we have asked a few of your colleagues on the committee to, and they have kindly agreed to, present a bit of work that they have done or been involved in just to sort of reinforce the background and context within which we are doing this work and sort of get everybody up to date to the extent that we aren't.

So thanks to our colleagues here who have agreed to say a bit.

Do you want to get started, Kathy?

MEMBER MCDONALD: Hi, everybody. Yes, I was asked to speak a little bit about the framework that was developed during the NASEM/IOM/NAM work by the Diagnostic Accuracy/Diagnostic Error/Improving Diagnosis Committee. We have been called all those different things.

And as you will remember from the phone call that we all had together, we were shown the framework in the state that is up here

on the screens in front of us, where failures are at the top. But this framework was built sort of in pieces by the committee, and what I wanted to do is to acquaint you with those pieces, in case you haven't read every word and page and looked at every figure in the tome that was that report.

So we will start with this, and we will end up coming back to it. So let's take it apart. Go ahead.

The important pieces are right here. The definition of diagnostic error, and again, on our phone call, we talked about that some, but I will show it to you again. And then the diagnostic process.

So these are the components of the conceptual model. The work system and factors that influence the process. And finally, the outcomes of the diagnostic process.

So conceptually, these four bullets are the parts that we want to drill into and understand if we are going to think conceptually about improving diagnosis and diagnostic error.

The next one.

Again, this was the definition that you saw as we had the committee call. It is patient-centered. It draws from other definitions, but it was a new definition at the time. It gives a certain primacy to the second part, to communicate that explanation of what is known at the time to be as accurate and timely as possible an explanation of the patient's health problem. So it has these two components and it is an Or. It would be a diagnostic error not to communicate that explanation. It would be a diagnostic error not to provide an accurate enough and timely enough explanation of the patient's health problem, given what is known at the time.

I mentioned on that phone call that it is worth reading the material around this definition because no short definition can capture the nuance of at least what was discussed by the committee. And I know what we started to discuss on that last phone call. But I just

wanted to point out that the more unique aspects of this definition are the idea of communication and the notion that there is sort of a process and outcome element to this within the idea of accurate and timely, the process being a little bit more on the timely side, that there is a process, and accurate being again constrained by what is known at the time or what could be known at the time. Next.

So the second big component of the conceptual thinking and framework is the idea of a diagnostic process. Again, building on a lot of other work. The process in these committees is to look at the evidence that exists and work from that.

The committee spent a lot of time developing this particular picture, and there is a lot of words behind this picture, but what I want to emphasize in terms of understanding this part of the conceptual thinking is that at the very bottom, in the smallest of print, but with a long line, it says time. And time is very

important, obviously, in diagnosis. So, in some ways, that probably should be larger and easier to see. But I want you all to know that we shouldn't forget that that's a big part of how we need to think about diagnosis based on what the committee was talking about at that stage, that that mattered. So even though it is small at the bottom, it is very much part of the conceptual thinking.

Another part of this diagram that I want to highlight is, before you even get to -- it is easy to sort of look at that big circle in the middle, which is the idea that there is kind of a flow of information and integration of that information and coming up with a working diagnosis, and it can kind of keep circling around. So that middle circle, you can sort of focus in on that and get a lot of attention, and clinically, that is where a lot of the action is. It has got the diagnostic testing. It has got the physical exam, the history. It has got everything that is in that kind of bull's eye of

clinical work, but it doesn't say that it is clinical work by a team. So I am going to get to that.

It also can attract your attention and make it harder to see that coming into that circle is the patient experiencing their health problem or being concerned that they need to be checked for something and the patient engaging in the healthcare system.

And then on the other side, you have this idea of the communication to the patient. You have some treatment, and you have outcomes. The treatment can feed back, of course, into the process. So if you try to sort of dissect each piece of this, you will see that it follows a view that would be combining the view of every member of the team, all the participants who would be involved in any sort of diagnostic occurrence.

Let's see the next piece of the conceptual framework, which is the idea that this process occurs within a work system. And that

work system is, again, this is another visual way of seeing it. The entire external environment could be affecting that process.

Right in the middle of these four circles is the diagnostic team members, so the actual people who are involved in the diagnostic work and the diagnostic process.

Around those team members, supporting them, enabling them, are organizations. There is the physical environment. There is actual tasks that are being done, and there is tools and technologies that can be involved. So this is a fairly stylized picture, but it is the idea that the diagnostic process isn't occurring all by itself.

So the next piece, let's drill into the work system. So right in the middle was that diagnostic team. And we have two different visuals in the report about the diagnostic team, and I think that is -- I'm not sure, actually, as it was being put together, what the exact sort of decision-making on it was about in terms of why

there were two visuals. So, I am kind of backing it out from my own thinking and what I recall in terms of the committee conversations.

I think this diagram is there because of the idea of the centrality of the patient and family members. So by putting together a diagram that has sort of the circles within circles, within circles, it helps us remember that critical aspect of the conceptual work reflecting what the committee felt was important.

The patients and family members then are sort of surrounded by diagnosticians. And again, the word diagnosticians was the idea that anybody involved in diagnosing is part of a diagnostic team. And then around that are other healthcare professionals who are supporting that diagnostic process. They are not making the diagnosis, but they are supporting it.

So this is one way of thinking about the diagnostic team. And for measurement I have to say I think this is pretty important because it is easy to sort of zero in on the particular

conditions that don't get diagnosed correctly, as opposed to how the diagnosis actually gets made and who is involved in it.

And then the next slide shows the other picture of the diagnostic team, and it is the team members. So you know it is focusing on the actual main sort of reservoirs of people who tend to be involved in diagnostic work.

The patient-primary care partnership again was emphasized because that can often be the starting point or the kind of quarterbacking or coordinating or facilitating hub. And then other groups are often quite involved in diagnosis and play major roles. And those are kind of bucketed again, radiology, pathology, specialists, and other healthcare professionals. Next.

Okay so now we get to see it starting to be put back together. So you see on the left side the diagnostic process sort of journey. The patient is entering that lovely circle, getting the communication, the treatment, the feedback

that can happen there within the work system, and all of that producing the outcomes.

So now the outcomes are blown up on this. So that is the last important conceptual piece. The conceptual model does not ignore outcomes at all. And the outcomes are pictured here as those that would relate to either accurate timely diagnosis -- so if we are trying to improve diagnosis, we care about accurate timely diagnosis and we care about the failures in that, the diagnostic errors and the near misses. And those are what then produce the patient outcomes and systems outcomes and produce the opportunity for learning and feeding back into the system.

So this is the conceptual framework that was arrived at. We could have stopped there, but because measurement was also prioritized as being very important, that is where the next picture that we started with comes from.

Now you add this idea of where are

failures potentially happening, because we want to understand the causes for diagnostic failures. And so to put them on this diagram, across the diagram, allows sort of the beginnings of thinking about how to do good measurement work in this area. And the committee couldn't go very far on that, which is why it is excellent that we have this opportunity here.

The last thing I just want to point out and acquaint you with, in terms of that foundational work, in terms of thinking about measurement of this area, are just a couple highlights from the chapter on measurement. And I will show you excerpts, if you advance.

So basically the committee had a dual focus on improving the diagnostic process. So that would be the improving diagnosis part, the diagnostic accuracy, the diagnostic quality part, and reducing diagnostic error.

So diagnostic errors were characterized in such a way that you would want to think about which aspects of the diagnostic

process are susceptible to failures and what would be the contributing factors to these failures. It can be factors within the diagnostic process. It can be factors within the team. It can be factors related to the work system and so forth.

So, the committee was using its conceptual model and input from other frameworks, which we will get to hear about soon, to give a context for measurement of both the causes and the risks of diagnostic error. That was the focus. Measurement can be the focus. It can focus on both the diagnostic process steps, the work system components, or both of them in order to identify causes and risks of diagnostic error.

There are two tables in the report. And what I have done here -- I would encourage folks to look at those two tables, but what I have done here is I have just pulled out the header, you know the columns -- the column headers in one example. So, this table, Table 3-2 provided methods for detecting failures across

the diagnostic process and looked at where in the diagnostic process the failure might occur, trying to say what you are looking for in terms of a failure.

So for example, a failure to engage in the healthcare system or in the diagnostic process. Not having the patient be able to engage in the healthcare system or the diagnostic process could be a failure point.

Then the nature of the failure. That could be, say, a delay in patient presenting or a patient unable to access care.

And then the methods for detecting that type of failure. You could analyze emergency department, urgent care, or other high-risk cohorts. You could do surveys to determine why and what could be done differently. So these are just ideas.

And then the other table was methods of estimating incidence of diagnostic error. And this gets at the problem of being able to know what the estimate of how frequently diagnostic

error occurs. And, again, the column headings break down how you would want to think about measuring where diagnostic error is occurring and how frequently. So the data source, the key features of the data source. Methods for selecting cases for review would be the denominator, and methods for determining if the error occurred would be the numerator with some examples here.

So that is a little less related to when we think about quality measure, but lots of the problems that have been encountered in terms of trying to determine the incidence of diagnostic error, and the data sources that have been used, and the limits of some of those data sources, and the limits of the measure you get from those data sources, a lot can be learned by that starting point of looking at that table.

So that concludes what I wanted to share to get you acquainted with a little bit more background on what the committee was thinking about as they developed those resources.

DR. BURSTIN: That was incredibly helpful. Thank you. It is actually a great -- that big book into like ten slides. That was awesome.

I have a question for you, though, because this is broader than diagnostic errors, per se. From your experience of being on the panel, how do you reflect on how the diagnostic errors framework reflects to a framework that would be more about diagnostic quality? There may be aspects of this that could be positively interpreted rather than always the error side.

Any thoughts on that?

MEMBER MCDONALD: Yes. There was a lot of discussion about needing to basically have the sea lift all boats and that a complete focus on diagnostic error would miss opportunities, and yet not focusing on diagnostic errors misses a lot of learning opportunities.

Ultimately, I think the framework and the thinking should support both directions. The report was, ultimately, titled Improving

Diagnosis, which reflected the fact that the ideas to try to understand where diagnostic error could happen were coming from a place of trying to think about how diagnosis really works and where there could be problems in that. So where are the opportunities for diagnosis to work well? It is the same conceptual -- it is hard to come up with a different way of conceptualizing this space. So yes, I think the panel was giving us frameworks and thinking that should be useful on both sides of that same coin.

I would be happy to take other questions or curiosities.

CO-CHAIR GRABER: I would just like to make a few comments, Kathy, if I could.

MEMBER MCDONALD: He was there.

CO-CHAIR GRABER: We were so happy to see this framework developed by the National Academy because it gives healthcare organizations a way to approach this. They all know how to tackle process improvement work. And for the National Academy to say this is a process, and it

has got steps, that is really huge. And it will serve as the basis for what we are trying to accomplish as well. We need to think about what are the measurement gaps and the measurement concepts that could relate to each one of those steps in the process.

There are three shortcomings in this framework that I think you should all appreciate. One is it doesn't incorporate the patients' perspective at all, and Helen will be talking more about this. So this is great from a physician and a healthcare organization perspective, but it doesn't get to what patients want. Patients want to feel like they are heard and that they are valued and that they are well cared for. And I think those are also things that, ideally, we should be able to measure and improve upon.

Secondly, diagnosis is both a noun and a verb. So this framework is wonderful for the verb, the process of coming up with the diagnosis. But diagnosis is also a label, and

this framework doesn't speak to that quite so much. So we need to keep in mind that label failures, and David will be talking more about that and Hardeep, are also an important concept that we need to tackle.

And a third aspect that has bothered me for a while, this is a question I was asked by the guy who runs the Palo Alto Medical Clinic. And he wondered how he could tell which of his doctors were good doctors and which were not so good at diagnosis. And his question was what observable behaviors are there of a physician that would allow me to tell that they are going to be doing a good job with diagnosis or they aren't. And just keep that thought in mind as we are kind of percolating through these measurement things. Because at the end of the day I would love to be able to answer his question, okay, here is the three things we should be able to do to answer that.

Thank you.

MS. SKIPPER: Okay, next up we will be

hearing from Hardeep.

MEMBER SINGH: So thank you for this opportunity to present to you.

You know, Mark said we have been doing this for about 2,000 years or longer, and one of the questions we have always asked in our research, why now? Why has there been such little progress in measurement? And one of the problems is that diagnosis and diagnostic errors lies at the intersection of several disciplines, and all of us know what it takes to get several disciplines to talk to each other in medicine. So, if you think about this, it is human factors, cognitive science, implementation science, sociology, social work, behavioral science, art of medicine. I mean you could go on and on. So it has really been a challenge.

We are still debating what diagnosis is. We have changed definitions of diagnosis of hypertension, diabetes, sepsis. We keep doing this. And for diagnostic error, it is even harder where there is confusion about all of

these concepts of what the diagnosis means. Mark brought out label failures. There is probably tons of other things as well.

There is often also confusion, and I want to highlight this because it is going to be important for this panel, with processes of screening and treatment. There is also confusion with quality and safety. Where are the boundaries between quality and safety? So I might say this is a safety problem, whereas somebody else might say this is a quality problem. What does that mean?

So in our work, we have taken some of these things into account about uncertainty and the fact that this is now black and white. And in fact, I want to show you these are real comments by front line doctors when the Institute of Medicine report came out.

And if you look at the comments, and I am going to read each one of them, it shows you about how much of uncertainty and sort of grayness there is in terms of diagnosis. And I

especially like the bottom comment, where it says many of the complications introduced by both medical, legal, and quality improvement efforts come from treating diagnosis as a black and white situation. So this is really important for us to sort of think about as we go forward in our work.

So safety begins with measurement. I think we all would agree that is why we are here. We are a believers of things should be measured and improved. It is hard to improve if you can't measure. But also it is harder to measure if you don't define the problem that you are trying to measure. So, I think some of the work that Kathy talked about, coming up with a definition by the IOM was very helpful.

What we have done in our work is we've used a very operational definition of diagnostic error, which really brings around the concept of missed opportunities in care. So we do a lot of retrospective analysis using some of the methods that Kathy was alluding to.

Looking at the case, we look for

unequivocal evidence that something different could have been done. That means that there was a clear missed opportunity to make a correct and a timely diagnosis. So we really go into details of that, and we use several methods that we can talk about later.

We also framed the missed opportunity within the context of an evolving diagnostic process. So just because you come in with cold-like symptoms to your doctor three days after presentation, and the doc says well this is viral infection, but ten days later now you have a sinusitis with facial pain and fever, and you get an antibiotic. Well, that doesn't mean there was a diagnostic error the first time you went to the doctor. So we often take into account evolving situations like that in our work.

And the third concept that we try to sort of think about is this opportunity could be missed by any members of the healthcare team, including the patient. So, this could be a physician or the nurse, or even the system.

There is several things a system can do in order to miss a diagnosis.

So thinking about what is important to patients: patients don't want to be harmed. So I think for this committee it is important to realize that, yes, we are always going to be looking for everything to be done correct and timely, but, really, it is the Area B that we should be focusing on, which is harm from delayed or wrong treatment and harm from delayed or wrong test, which our patients do not want that in terms of what we are talking about here.

We like to focus on the green area, where there are clear missed opportunities. There are several situations, there are rare conditions, there are several situations it is not possible to make diagnosis in a timely or a correct fashion, and we should not be including those situations in some of the work that we are going to be thinking about. It is not possible.

Now, in 35-40 years it might be possible to make diagnosis in every condition. I

saw that movie in which they put people on the scanner, you know Passengers, and the diagnosis comes right out. And there are so many conditions. Wonderful. I think, Mark, we'll get to it but not in our lifetime.

So we need to be thinking about what we can actually practically do today.

So what are the foundations for rigorous measurement? Unfortunately, there is not a lot of good, valid, and reliable data sources right now, which really limits us. Whatever we think about must reflect real-world practice. We need to think about not just what is in the doctor's head but what is beyond. So what Kathy mentioned about the team processes.

Mark, to your question, the diagnostic performance is really just individual and system performance. So we need to sort of take that both individual and system-centric views.

And we need to be thinking shared accountability beyond just the clinician. So whatever we think about, we just can't think

about just measures that would be physician-focused. We need to be thinking about systems. Systems need to set up and step up to the plate and be measured as well.

So this was sort of our thinking over the years, which led to the Safer Dx Framework that I am going to present, give you a little more details on, essentially was I believe instrumental. I wasn't there. It was instrumental to the IOM committee's thinking on what kind of conceptual framework they should use. And just a couple of highlights I am just going to mention, since all of you know what structured process outcomes are.

We have used the socio-technical model where we have eight dimensions we think about. David is very familiar with that work because it is used in the health IT circles a lot, where we think this a socio-technical system, which is a complex and adaptive system. So we need to think about technology as well as non-technological dimensions.

In terms of processes, we think beyond just a single provider visit.

And in terms of outcome, this is the hard part because you are not thinking about missed, delayed, and wrong diagnosis only, but you are also thinking of overdiagnosis, and you just can't get away from that situation, even though we might try hard and say, well, that is not a diagnostic error. It is because we often struggle between should I order the test or should I not order the test. Will I not order the test and miss the diagnosis? And if I order the test, I will be overtesting. So we can't just take that into -- out of the equation.

So we think about all of that. So that is what the Safer Dx Framework looks like. We call it a measurement framework, but it is really an improvement framework. I talked to you a little bit about the socio-technical work system, but if you look at the process dimensions, there is really five things we really focus on in our work. And almost all the

diagnostic error work we are doing focuses on one of those five dimensions or actually its interactive sort of five dimensions.

Well one is when you talk to the patients. So patient-provider encounter. You make initial diagnostic assessment. When you order tests, the tests get integrated, or a radiologist might read the test, or a laboratory and a pathologist. And then the tests come back and need to be followed up. So, there is abnormal tests that will get lost to follow-ups. That is the third dimension.

The fourth dimension is to have a referral. So we send referrals to sub-specialist and often, they get lost in the system. So that is another dimension.

And the fifth dimension, which is the central dimension, is the patient, which is right in the middle interacting with the other four dimensions. So in almost all of our work and measurement, we consider these five dimensions very useful in the work we do.

All of you would be probably familiar with this. Measurement has to be reliable, valid. It could be retrospective as well as prospective. Right now we are still in the retrospective phase. We are still learning from what happened in the past. But the goal of this measurement, I would highlight two things in the blue box over there. One is organizational learning, which we are all familiar with, but the other goal of measurement is better measurement tools and definitions.

So, I think whatever we do in terms of measurement is going to stimulate better measurement tools. Because we are still early, we still have a long ways to go, which should lead to safer diagnosis.

Now, a couple of other things. This feedback, which we learned from -- you know improvement feedback that we learned from measurement, should lead to changes in policy and practice but also should be feedback back to the clinicians as well as the systems for improvement

as well. So, that is the highlight of the Safer Dx Framework.

What have we learned after a decade of doing research in the area? Common conditions being missed. So we are talking about infections, cancers, cardiovascular conditions, and some others. It is the common stuff that gets missed, despite the presence of useful information that could have led to the correct diagnosis. So it could be red flags for cancer, for instance, or clearly something different could have been done, and there are plenty of those situations that we can focus on.

We have thought that a lot of the problems are within the patient-provider encounter dimension, which is where the history and the physical and all that initial diagnostic assessment takes place, but a second very important dimension that we realized from our work is follow-up of abnormalities or follow-up of test results, for instance. So these two dimensions feature very prominently in our work,

including patients.

Poor calibration is key. And by calibration I mean we are always constantly struggling between under- and overdiagnosis. This is really a problem for clinicians. Should I order the test? Should I not order the test? And we cannot forget that.

So I think it is time ripe for correct retrospective measurements. Signals from routinely available administrative data is quite weak for the clinical signals you really need to make an assessment, whether this is a diagnostic problem or not. We need to think about stronger signals. So Kathy mentioned high-risk cohorts. We have looked at several of these cancer patients. At average, about a third of cancer patients, no matter what cancer you look at, have had missed opportunities in their before they get diagnosed. So, colon cancer, lung cancer, hepatocellular cancer, no matter what cancer we look at, it is about a third that had missed opportunities. So, that is a high-risk cohort.

Test results get missed. The number is anywhere from 10 to 36 percent, depending on what study you look at. Plenty of test results get missed.

We have looked at some triggers. That means these return visits, and Kathy even showed you a slide with that as well, where patient would come back to the same system after being seen, let's say in an emergency room environment or a primary care environment, where something was missed in the first visit, and then they got unexpectedly hospitalized. So that is another trigger for us. We call these triggers.

We have built triggers for visits, as well as missed results. So after an abnormal chest X-ray, you expect to see, for instance, a CAT scan ordered or maybe a visit to a pulmonary physician. But if you don't see that, we have built algorithms that use electronic health record data to identify patients who may be falling through the cracks of the healthcare system.

And of course reports from providers or patients. I put that there. Patient reporting is very, very early. We can talk about that later. Even provider reporting is very early. I know Kathy has done work in that area, but it is hard to get providers to report anything about diagnostic error. So I would say it is early.

Well, once you identify a record to review, how do you determine what is a diagnostic error? When you look at the record for a case, how do you determine what is a diagnostic error? So we have built an instrument which is essentially to think about medical record review process and to determine whether this is a diagnostic error or not. But the same concept could be applied to get more objectivity to determine whether this is a diagnostic accuracy error, safety problem -- we are still figuring out what definitions to use, by the way -- and to figure out what is wrong. So that is why I think we need a structured approach to think about how

to do this.

So Helen mentioned why not think of quality, diagnostic quality versus focusing on just error. And that is what we have done very recently. There is a paper that I would highly recommend that you look at, where we just sort of framed the problem as a safety problem and then proposed a structure process and outcome candidate measures and measure concepts -- not even measures. I would say they are just measurement concept for people to look at. And I am actually glad to see that some of them made it to the framework that you all sent around. So thank you.

So, this is some of the questions that we asked in the paper. Here are the six questions that must be answered for thinking about measurement related to diagnostic accuracy. And so Mark mentioned one of them, which was observable behavior. So, I am not going to repeat that but appropriate time intervals. Can we agree upon what is an appropriate time

interval to diagnose lung cancer, for instance? Let's raise your hand if you agree to let's say 30 days or 60 days. We could have this exercise all day long. So we need to think about what is a standard for diagnosis of a particular condition.

How do we measure competency of clinical reasoning in a real-world practice setting? We talked about team behavior.

What about system properties? How do you know whether VA or Kaiser or any other healthcare system represented in this room has a good diagnostic performance? How do you do that? How do we leverage information technology? We are all collecting IT like data out to the wazoo. We are using electronic health records. We are collecting data how to use information technology to figure this out.

And then how do we leverage patient experiences to help us improve diagnostic safety?

So what I kind of have thought about this is we use what I would say is actionable

measurement. That means we are using measurement for quality improvement purposes and learning and, of course, for research. And this translates into feedback at the system level and maybe at the individual level. We are not really ready for public reporting performance measurement or penalties with diagnostic error measurement right now or even diagnostic safety or quality measurement just because I don't think we have the robustness in the science right now.

We really need to engage providers, organizations, patients. Patients are, of course, are getting more engaged but organizations are far from engaged in any process related to diagnosis. This is not one of the priorities.

Providers are already burned out by measures of all of the kinds and how do we think about having some measures that could measure their diagnostic performance.

We need to generate evidence and think about harm, like I have discussed, and safety but

also what is diagnostic reliability? How do we measure uncertainty in this process? So we need to think about all that.

This is my last slide. Let's only measure if it is actionable for safety. And I would emphasize safety and think about sort of the preventable harm.

We can think about sharp-end outcome measures but really the basic science in that area is very far behind.

I think we are getting ready for blunt-end measurement concepts related to sort of system-level performance. In the VA recently we had a policy change for communication of test results and when is an appropriate time to make sure that the patients have received their test results. We came up with a standard, a national standard. It was by consensus. We came up with an actual standard and now we have a measurement system in place that is looking at the entire VA to figure out how many people are meeting the measure for communication of test results to

patients. So, we came up with a time line: seven days for actionable, 14 days for non-actionable. And so most systems don't have this type of standard.

We need to think about measurement burden and unintended consequences of measurement but also I would say I think we need much more of a measured, cautious approach to inform the measures we are going to be thinking about in the next few months.

Thank you. And many of the papers I mentioned are on my research profile, if anybody wants to look at those. And I want to thank my funders.

I'm happy to take any questions.

MR. LYZENGA: Yes, questions or comments or thoughts?

MEMBER CAMPISANO: So I have a couple of questions. You mentioned that there were a few valid and reliable data sources. What would count as a data source or what are your data sources?

MEMBER SINGH: So we look at let's say medical records a lot. And so once you, let's say develop, once you get to developing an algorithm to figure out what medical records to look at, so let's say at any system, Palo Alto Medical Foundation -- let's pick on them and not the VA this time -- would have let's say 100,000 visits over a period of a week. Which one do we look at?

Now once you select -- let's assume we build algorithms to select the hundred that we want to look at. Then, we have got to look at the case to figure out which. How do you determine whether this is a diagnostic accuracy problem?

So I think in terms of data, there is documentation problems. And as many of us know, the chart load and copy and paste is making documentation worse than it ever was before.

So I think any source we look at, we just don't have the robustness of whether the source is valid.

MEMBER CAMPISANO: And is that because doctors don't want to provide them because of fear or and then you said also organizations aren't as engaged as they could be and that patients are starting to be more engaged. What types of organizations and what do you see as the reasoning for not being engaged?

MEMBER SINGH: So I think there is two different problems. One is the sort of the engagement problem. Let's measure diagnostic safety. So I think that is an organizational engagement problem where you could put together a system of using algorithms and triggers to identify some high-risk cohorts or start looking at all the patients who were diagnosed in your system to figure out how many were late.

So there are things organizations can do but I think the problem of data is a separate one. We sort of have documentation issues. We sort of just don't document very well.

I was reviewing a discharge summary the other day and it had seven signatures that

were copied and pasted into one discharge summary. And so it was everybody else's signature and the person who signed it signature on the bottom. So, just an example of what we are doing to our progress notes, which is sort of the standard of care. That is how the court gets to know what is wrong in a medical record, what is wrong in the care of a patient.

MEMBER CAMPISANO: Thank you.

MEMBER IRONS: I am sitting here sort of processing and thinking about the word time. And you know we talk about that lower bar of time that the process occurs but in your framework where is -- is time -- how is time that is used to talk with consultants, where does that fall?

Because what I worry about is as medicine has gone from really a collaborative process to a more sort of see the patient, get them out, do this process, we have lost that interaction. You know you can't find radiologists anymore to have a radiology conference because they are too busy reading

films. You don't have time to call the lab to talk about how they interpreted the test.

And where is that accounted for in the framework and how do you even measure that? I just wanted to know how you think about that.

MEMBER SINGH: Yes, so you know the framework is that of a measurement framework. Once you identify the problem, you then understand why we are having the problem. So what you are talking about is a very strong contributory factor that means we have looked at sort of primary care records and there were so many errors that were so common. And one of the things was people don't have time to talk to each other anymore and I think this is a very, very strong contributory factor and I totally agree we have time-pressured visits. We don't have time -- we are basically spending time with the EHR. So there was a study where you were spending more time with the EHR than you were with the patient. But I think that is just one of the problems.

Even in vignette studies, when

physicians have plenty of time to diagnose patients correctly, we are still not getting it right. So now we have examples of several vignette studies that have shown even when you give them time, still it is not happening. So I think time is important in the real-world setting and we should definitely be thinking about this. But I would say we are having sort of more problems than just the time because this is where the competency, the thinking, the cognitive process, and the information seeking and sort of the calibration concepts that I was talking about earlier are also important. And they are all contributory factors.

MEMBER IRONS: Agreed. I agree. But just as a follow-up to that, I wonder whether, in terms of competence, you know as testing becomes more difficult and we worry about that doctors don't know what we don't know.

MEMBER SINGH: Yes.

MEMBER IRONS: And should there be some sort of a discussion that is automatically

part of an interpretation of a result? A physician just can't go to the computer and look at the MRI and know whether the radiologist's interpretation made sense in light of the clinical context or a genetic testing result may not be what it seems on paper. Is that -- are we at that point yet or is there a way to measure that within your system, within your framework?

MEMBER SINGH: No --

MEMBER IRONS: How people interpret the results that they are getting from laboratories and radiologist --

MEMBER SINGH: I mean you can look for that documentation but we often don't find it. I mean we routinely miss documentation of even clearly labeled abnormal results. So, we have, for instance, looked at imaging studies that clearly have abnormal imaging labeled next to it and we are sort of still missing that.

I think what you are proposing is one step even further than that, which is looking at it, coming and saying well, this doesn't make any

sense; let me talk to the radiologist. But we don't talk to the radiologist anymore. We don't have radiology rounds. We don't talk to the radiologists anymore and I think the same is true for the lab community as well.

CO-CHAIR DANFORTH: David and then Martha.

MEMBER SEIDENWURM: So from the radiologist's perspective, I can't agree with you more. I mean this has really changed. And I think an analogy from all of our lives perhaps for the non-clinicians, remember when we used to pick up the phone and talk for 15 seconds and figure out where we were going to meet someone. Now we spend four days texting and we still can't figure out where we are meeting and when because the communication is atomized and lacks the richness of interpersonal communication.

So I think that you really are on to something. Now how precisely we measure that, I don't know. How exactly we do it, I don't know. But we definitely have lost something and there

is a subtlety of almost a Bayesian dimension that we lose because in 30 seconds on the phone, the referring physician will say no, he doesn't have a fever, this isn't an infection or oh, you know they mentioned that they were waking up at night sweating, maybe it is TB instead of a tumor. And that is the type of thing that comes out in 15 seconds on the phone that won't come out in 15 months necessarily until some disaster happens through this atomized process.

So, if we can figure out a way of capturing some of that, it happens, actually, in our practice best in the disciplines where we have specific clinical conferences, tumor boards are a big one, pediatric neuroradiology is another one. In my practice, we have a spine radiology one, all the other areas have them.

But we do need to capture some of that richness -- recapture I should say some of that richness, I agree.

MEMBER SINGH: And I think you are right about the time, bringing it back to time.

I think we, a lot of times, are not doing it because we are so focused on just meeting the documentation requirements and getting that patient and going from one to the other.

We could pick up the phone and talk to the radiologist or walk down and look for them, if we have the time.

MEMBER SEIDENWURM: Nobody has the time.

MEMBER SINGH: Yes, nobody has the time. It does come back to time a lot.

MEMBER SEIDENWURM: Nobody has the time.

MEMBER MCDONALD: Quickly, from the framework perspective before. So some of this would be kind of that idea of the team work, of course, and some is the working diagnosis. So there is not a place in current systems of data collection to kind of capture that evolving working diagnosis or the reworking and reworking. So that would be a piece of this, too. If it were there, then you could look at it and you

could see how we are doing.

CO-CHAIR DANFORTH: Martha.

MEMBER RADFORD: Just first a comment about the last few discussion points. Well, to me it is quite obvious why this is the case, which is cognitive work is not reimbursed as highly as the other work that the people are spending their time on. That is just a comment.

I do have a question for Dr. Singh, which is could you just highlight for us the differences between your framework and the NASEM framework?

MEMBER SINGH: Kathy, I was hoping you would do that. No, just kidding.

MEMBER RADFORD: Or both of you. I don't care. I mean they seem quite compatible to me.

MEMBER SINGH: Yes, they are.

MEMBER RADFORD: If there are any glaring incompatibilities, I would appreciate knowing from you.

MEMBER SINGH: No, I don't think so.

CO-CHAIR GRABER: Hardeep's framework served as the basis for the NAM framework very clearly.

MEMBER SINGH: But I was just in the appendix, though, which is okay.

But what I wanted to highlight is I think some of the feedback part, I'm not sure how much of the feedback part was incorporated because ours is -- we are going to just present it back. Can you take it back just a few?

So, a couple of the things I want to highlight is I think Safer Dx is a lot more -- firstly, we came up with these five dimensions, which you know were instrumental in sort of guiding our work. The socio-technical dimensions, on the outside, are very similar. We use 8-dimensional Socio-Technical Model. You use the Systems Engineering for Patient Safety Wisconsin Model, which is where the technical model originated from. So we think that is quite compatible. The process dimensions you could sort of just align quite well I think.

The measurement part and the blue circle in the middle is clearly a bit different because this is what I call the actionable measurement type of framework, where you build on reliable valid measures but though that measurement has to lead to those four things, it is collective mindfulness at the institutional level, organizational learning which I think everybody sort of knows about. Collective mindfulness is basically just comes from aviation and other human factors literature.

Calibration is something that we think which is really, really key and that is alignment between the accuracy of my diagnosis and the confidence of my diagnosis. So we have shown that physicians are very poorly calibrated in the sense that even when the diagnosis is wrong, they are very confident that the diagnosis is right. So we have shown that in a vignette study. So we think calibration is important and the better measurement.

So I think the measurement focus is a

bit heavier in Safer Dx.

MEMBER MCDONALD: Yes and I mean I concur with that, too. This is definitely measurement-focused and a little bit more diagnostic error-focused actually.

MEMBER SINGH: Yes, error-focused, yes.

MEMBER MCDONALD: So to Helen's earlier question, I would say that the framework from the committee gives you -- it is not just that one picture. It is those four components. It is the definition. It is the process. It is the work system and it is the outcomes. So, they are complementary and, obviously, that framework tried to build on this but then this gets into more of the devil of the details, if you are trying to think about measurement, which this is where you have to get pretty detail-oriented for measurement.

MEMBER RADFORD: Right. I think that part of our call here is to focus on the measurement aspect. So I thank you very much for

those comments.

MEMBER SINGH: And I would just add that I think the error focus I think point needs to be sort of also highlighted. A lot of what I told you would be very different in terms of language from some of the discussions. And I looked at what you all sent, which is very quality-focused and I am going oh, my God because once you start talking about quality, it is like this can of worms that just opens up and you can go on and on about diabetes quality. Are we doing the hemoglobin A1cs and the retinopathy screening and all that. And we want to stay away -- in our work, we just stayed from all of that.

CO-CHAIR DANFORTH: David has had his card up and then Prashant.

MEMBER NEWMAN-TOKER: I just want to take us back one comment to something that Kathy said and reflect on something that Hardeep said earlier.

So this issue of the working diagnosis and the sort of time issue, the sort of evolution

of the diagnosis from sort of a tentative I am not really sure but I think it might be something in this sphere or a differential diagnosis all the way eventually sort of combing down through a working diagnosis that is one that you take action on from a therapeutic standpoint but you may still be monitoring and then eventually sort of closing the book on sort of a final diagnosis, I think that not only have we not captured that but we are actually almost deliberately getting rid of it. And this gets back to Hardeep's point about data quality and data sources.

So one of the key problems that is actually quite simple in some sense is that we don't even keep track of what the chief complaint is. The symptoms the patients come to see providers with when we are going through the symptomatic diagnostic process, that first front-end piece gets replaced with some ICD-9 coded diagnosis as soon as somebody takes some kind of first billing action, even if they remain completely uncertain about that diagnosis. And

that loss of information is huge at many levels, including the fact that, obviously for all the rest of the providers, it anchors them to a less than certain diagnosis at a premature stage, as a practical matter.

But from a research standpoint and a measurement standpoint, it actually makes it really hard for us to actually go back and figure out whether whatever diagnostic errors might have occurred actually related to that original complaint because it has sort of vanished from the administrative and billing records. So that may be something that we want to at least think about as part of this issue of measurement framework.

CO-CHAIR DANFORTH: Prashant you had a question for Hardeep?

MEMBER MAHAJAN: Actually not a question but I just wanted to throw out this. At least the difference that I saw from the conceptual framework, you know the Safer Dx framework and IOM was based on the definition

because the second component of the definition, which is the communication of the patient's health problems. And what I am struggling with, and I recognize that we are not going to come up with measures but we are looking at conceptual framework for measurement, is not only is this communication a big aspect, which is contextualized where the healthcare setting is, but this is also this aspect of health literacy from the patient's perspective.

And somehow it needs to be baked in or probably a tacit recognition that some of these may not be ready for measurement, even though we have a full definition, some aspects are not possible to be measured at this time.

CO-CHAIR DANFORTH: Kerm.

DR. HENRIKSEN: Just looking at the two frameworks that we have seen thus far and then thinking about the IOM definition, the second aspect of it, the communication with the patient aspect, I haven't seen anything in the existing frameworks presented that are really

sensitive to the communication process. And I know it is sort of embedded in there someplace but it is not very prominent. And so are there certain structures and processes and outcomes in a Donabedian sense that are just relevant to communication, clear sound communications?

And I know there is repeat back and rubrics and things like that that can be invoked but for much of the one-half of the definition of the National Academy's report doesn't really show up that clearly in these frameworks.

MEMBER SINGH: So Kerm, I am just going to -- so there is a couple of places in the process dimensions that communication is measured, if you do, let's say, medical record reviews to look for sort of diagnostic issues. One is in the patient-provider interaction. So the history and the physical and what you do next is based heavily on communication and listening to the patient. So I think that is where it is reflected.

The second area I would say it is

reflected is follow-up of test results. Clearly, communication of test results is in the framework and is measurable and it is, in fact, a very important area to be measured and that also comes from the IOM. That is sort of compatible, I would say, with the IOM.

CO-CHAIR DANFORTH: We are going to take one more question from David and then maybe we will transition actually to have David Newman-Toker introduce a third framework. So, we will have lots of time to talk about and compare all three.

MEMBER SEIDENWURM: So I would like to go back a little bit to the issue of time, you know that sort of tiny little font and that sort of hard to see line at the bottom.

You know time is really one of the crucial things in diagnostic accuracy. And many of the systematic errors that we see are because people are hurried and because time is expensive in the operating room or the emergency room is what I meant to say. So I wonder if there is a

way that we can get at calibrating that and measuring the appropriateness of the use of time. And that sounds vague because I think it is but I wonder if people have any thoughts about how to incorporate that into our measurement framework because sometimes we do things too quickly and sometimes we do things too slowly.

MEMBER SINGH: Well I can just quickly speak. If you are an internist and if you see 50 patients a day, that is probably a measure of poor safety right there. So I totally hear you and I think we could think about some of these measures. I'll stop at that.

DR. HUNT: And time is an element in some quality measures that we are seeing. Look at colonoscopy. They have actually started to say how long is it taking you to come out with the scope. And if it is too quick, then we know that there is maybe some issue.

DR. BURSTIN: And just quickly as somebody who precepted residents yesterday, some of their logic of looking at this in the broader

context of quality is that you begin to look at how time is spent overall. So the amount of time my residents and I spent, for example, on documentation for billing, if you think about the broader context, was overwhelmingly disproportionately the time we spent as opposed to perhaps thinking this appropriateness of time concept is difficult to measure but intriguing in the broader context of which measures and tasks are not adding value and need to move out to allow time to do what is considered more valuable. It is a really interesting question, David.

MEMBER SINGH: And I really like your point because that is a system-related measure that we could be proposing as a measure concept that basically if you are billing and everything else suggests that you are seeing x-many patients a day, that is an area of risk that could be explored.

MEMBER NEWMAN-TOKER: And I think in this sort of in the Donabedian structure process

outcome kind of framework that I think we are headed for, in some sense that is a structure measure. It is really a measure of capacity. So institutional capacity, if you have one physician to see 50 patients in a day, you know 50 patients for that one physician, but if you have got five, then you have got ten per physician. So, I think that may be a very interesting place for us to explore.

CO-CHAIR DANFORTH: I think one thing to think, though, remembering a couple of Hardeep's earlier comments, is that time as a stand-alone measure may actually be inadequate. I mean you gave a great example of physicians that had more time and still got the diagnosis wrong. So there is actually some evidence that time as a stand-alone measure may be inadequate. So time potentially combined with other factors because competence, you know cognitive ability, bias which we haven't talked about, physician bias which we haven't talked about today, I mean there are other things to think about about how

that time is being used.

MEMBER SINGH: And I think, whatever measure concepts we propose, we are going to need to think about that these -- it is going to come as a menu because there is not just going to be a few things that will satisfy it. This is so complicated and complex, we are going to have to propose a menu that measures the entire process or the structure or whatever else you want to measure.

MEMBER MCDONALD: I know you are going to go but just really quickly from that IOM perspective, the thoughts around measurement at the time were that the measurement systems we have for other types of measurement activity might not be all we should think about. So I know you said we are going to move towards structure process outcome but the thinking was partly because of that process and partly because of that time perspective, that there might be other ways we need to think about measurement that we haven't thought about measurement so far

in this domain. So, I would just like to keep that out there as the committee sort of aspirational thought was measurement might need to be a little different in this area.

MEMBER NEWMAN-TOKER: So thanks for the opportunity to present today. I am not going to present to you a third sort of competing model. Can you guys hear me okay?

I'm not going to present a third competing model. This is really, it is orthogonal to the discussion we were having before. And I just want to introduce what I am about to say briefly by explaining where it came from.

So at some point before the National Academy of Medicine report came out, several of us were having ongoing internal conversations in the sort of diagnostic error medicine community about how to define a diagnostic error and what terms are we going to use and so on and so forth. So Mark and I, and Hardeep, and Gordy Schiff, and Paul Epner started having some phone calls to

discuss this issue of what is a diagnostic error and how are we going to define it and what are we going to call a diagnostic error, a diagnosis error, a diagnostic error. And after about four or five meetings it became abundantly clear that we could not agree on how we were going to use these terms.

But at the same time, it also was clear that in concept, the ideas we were talking about were actually very harmonious. And there was kind of an underlying structure to that idea of what we -- how we were thinking about diagnostic errors. We just wanted to call -- different people were pointing to different pieces of the Venn diagram and saying no, that is a diagnostic error. And I want to show you the Venn diagram sort of in its richness and give you a little bit of an introduction to that as a framework for our thinking about what we mean when we say we are going to measure diagnostic errors.

So my disclosures, as I mentioned

earlier, research funding, research support, and my membership and Board membership in SIDEM.

So everybody knows the National Academy of Medicine definition. The key thing for this definition from the sort of broader perspective is that this was framed I think in a very patient-centric way around the idea of if you didn't get the right diagnosis in time, it was a diagnostic error. So that means you have to have gotten an accurate and timely diagnosis and that information has to make it to the patient. And I think that is very sensible. But one of the important things about this definition is it doesn't mandate that we be talking about there was a specific process failure or such and such was preventable. It just says look, if the patient didn't end up with the right diagnosis, that was a diagnostic error. And I think we should maintain that sort of structure of thinking as we go forward into measurement because I think it also helps us a little bit with this dichotomy between quality and safety or

whether we are framing things as error, or positive or negative. In some sense, because it doesn't mandate that there have been a process defect in order to be in this umbrella of diagnostic error, it allows us the freedom to say okay, where are we getting -- how are we going to measure whether we are getting diagnosis right and not necessarily have to have, in the same conversation, whether we know exactly where it went wrong. We can actually just focus on the piece of the rightness of the diagnosis.

So here is the model. And what I did was I tried to use as inoffensive terms as I could so that I didn't inflame people's sentiments as we constructed this model. And there are two concepts that form the core of what I am about to tell you. One is the idea of diagnostic process failures and the other what I call diagnosis label failures. And I specifically differentiated between diagnostic and diagnosis because Gordy was very insistent that the diagnosis is the label people, the

diagnostic is the process part. And it actually took me about a year to figure out that Gordy was differentiating between diagnosis error and diagnostic error when he spoke but I maintain that differentiation here.

And this model is basically the same as what other people had published before, Gordy Schiff, Laura Swann, and others that these two things are sort of separable entities. That is to say, just to make it concrete, a 50 year old man comes to the emergency department with chest pain and they don't get an electrocardiogram and they are sent home with a diagnosis of reflux and it turns out it was a heart attack. So two things happened there. The one case there was a process failure that the EKG didn't happen. The second was there was label failure in that it was called reflux and the patient turned out to have something else. And the confluence of the two, in this case, may have come together but they could also come separately. So the patient could have been given a correct diagnosis of reflux and

still didn't get the electrocardiogram, in which case there was a process failure without a label failure. And you might think of that as kind of a near miss, in some sense. And then there is the sort of flip side, which is you could have ordered the EKG and done all the right things, such as we know, the best that we know, and still ended up with the wrong diagnosis because it was a subtle presentation or whatever it was. Mark had called those before no fault misdiagnoses.

And what is in the middle is kind of this core of this sort of patient safety question of this sort of preventable diagnostic error, this idea that we really did something substandard that didn't meet the known expected standard of care and, in the process, we had ended up with the wrong diagnosis label on the patient.

And in some sense, just to kind of clarify where the National Academy of Medicine definition fits there, really the National Academy of Medicine definition is most closely

mapped to that blue circle. It is saying I got the wrong diagnosis and it is not actually calling for, specifically, with the exception of kind of saying that it has to be timely, which sort of allows for the fact that it could be wrong transiently, if you didn't get the right diagnosis label, then there was a diagnostic error.

The process failures fall into a separate category and what I want to try to convey is that I think that that is a really appropriate thing for us to do, which is to disentangle process and outcome because I think if we try to force measurement on both dimensions in the same measures, we are going to get ourselves caught in kind of a messy thicket. I think we actually have to think about them as separate types of measurements.

Here is what I sort of added to this model or the first of a couple of pieces I added to this model that I think are important to remember as we think forward to measurement,

which is that not every process failure is a substandard care act. Sometimes it's the standard of care but it is not terribly -- it is not operating at the highest level that we could be. So we have seen this a lot where the literature says that it takes about 17 years for a new innovation to make it from a scientific fact, such as it is, to clinical practice. That evidence-practice gap is huge and it is actually probably responsible for a big part of the places where we fall down and don't necessarily get diagnosis right.

That suboptimal, yet still standard care is kind of a whole other circle. Here I have drawn it as an oval around the original sort of process failure dimension and I have called these sort of reducible diagnostic errors in this section here where there is an intersection between the blue circle and this new sort of pink circle, pink ellipse around the edge.

These are not things where we could point to and say look, you have failed the

standard of care but where we could still envision that well, look, if there is a new technique or technology that has been developed at the Houston VA, or here, or there and other people are operating at this higher level, but that hasn't disseminated, that is an opportunity to improve diagnostic quality without necessarily thinking of it in squarely the standard of care was breached sort of way of framing it.

And finally, there are the sort of rest of the National Academy of Medicine-defined diagnostic errors, which are what I called sort of unavoidable diagnostic errors, meaning we got the label wrong but we did everything at the highest imaginable quality for today and the science just wasn't there to support it. And obviously, those are opportunities for discovery. And I think this group needs to at least wrestle with briefly how that fits into the measurement framework around the issue of diagnostic quality. Some people might say well, if it turned out there was a large public health burden of

problems in diagnosis that were unavoidable in current technology, that might be something that we had to point our attention to in terms of greater research funding. So for instance, if all the screening for breast cancer or something else is being done perfectly in clinical practice but it is just our tests aren't good enough to find the cancers early enough to get the right treatments, then if we measure how often that is happening we can get a sense for how much we should invest as a society in improving the quality of that piece of the diagnostic process as well.

Obviously, any of these things can be associated with harm or not harm but I think, ultimately, from a public health perspective, I think that the space that we should focus on, and this is very similar to what Hardeep suggested in his slide with the harm, is this idea of I think patients care most about harm. They care less about process problems and all the things that sort of they get the wrong diagnosis for ten

minutes but it doesn't lead to a problem for them. I don't think that matters as much to patients.

I think, at the end of the day, this sort of preventable and reducible misdiagnosis-related harm space is really the most important space for us to focus our attention, in terms of what do we want to measure that is going to help us solve this problem. And I think secondarily, the rest of that inferior ellipse there from all the harms that are associated with overtesting, overdiagnosis, et cetera, where we are missing on the false positive side, rather than the false negative side and we are leading to harm to patients through the diagnostic process itself.

So, what are the measurement implications? Well, I think it is important that these are separable entities, that is, the diagnostic error in the National Academy of Medicine definition are the blue circle, the diagnosis label failures, as I called them, doesn't, per se, require a process failure.

Although it is clear that we have to measure process defects in order to improve our diagnosis labels, I think we should maintain some mental separation between those two concepts and really treat them as different circles.

The second is that, obviously, these process defects alone are sort of near misses, and that may be important, especially when they cause harm. And this idea of sort of suboptimal processes as being sort of one concentric ring around the idea of a failed process may help us push more in the direction of quality that it is not just a problem of we definitely did it wrong and failed in an absolute sense, but that we weren't doing it as well as we could have been and we could do better.

Obviously, overdiagnosis and overtesting may harm patients and harm is a key parameter to measure. And let me just show you one concrete example of how I think some of these principles could be applied in sort of a measurement concept.

So one of the things that Hardeep alluded to before that I think is right is that if you look at administrative data, the signals are weak. So if you look just at three 72-hour revisits or whatever, most of them are not diagnostic errors. But I think actually if you take a symptom disease framework, which we have described in the literature before and focused your attention on specific initial presenting problems that are potentially related to initial returning diagnoses, you actually can use administrative data to find very strong signals like the ones I am showing you here and you can do it without necessarily diving deeply into the process of whether there was a process failure that led to this outcome.

So let me just explain what I am showing you here. I am showing you two research studies from administrative data that look back and look forward at the same problem, which is the idea that people come to the emergency department; they are told they have benign

dizziness from ear problems and it turns out they have a stroke. And then they come back and they get readmitted to the hospital with a stroke.

These are essentially patients who are being harmed by diagnostic errors. We don't know whether those errors are preventable. We don't know whether those errors are associated with specific process defects from this information, alone, although we have ample data that indicates we know exactly what is wrong in this particular scenario because we have done a lot of work in this space. But as a measurement, this measure relies on the analytic piece and it gets us to this issue of time.

So what you can see in these graphs -- and I will just walk you through each of them and then I will take whatever questions we have -- in the left-hand graph, we have taken 180,000 stroke admissions to the hospital and looked back to see when people were seen and released with a benign diagnosis from the emergency department in the days prior to that, in the 30 days prior.

And the red bars that are sort of highlighted by the red sort of exponential-looking curve there, are patients with dizziness and headache. So you can see that there is sort of an overrepresentation in the week prior to this stroke admission of these patients who were told that they just have migraine or they just have benign positional vertigo, et cetera.

And then we have a comparison population here of abdominal and back pain patients beneath that and you can see that they are mostly flat, sort of randomly distributed. So there are some people who come to the emergency room and they have strained their back and then a month later they get a stroke that happens, just random occurrence.

But the temporal association is strong here. And likewise, in the other one looking forward from discharges for supposedly benign dizziness, you can see the rate of return in the top curve for stroke and the rate of return for heart attack on the bottom. You can see the

stroke curve has this peak in the first 60 days that flattens out to a base rate and the heart attack is just sort of a flat base rate from the start. And what that tells us here is that there is a biologically plausible relationship we know between transient ischemic attack and minor stroke and the risk of major stroke afterwards that is played out over the highest in the first few days and plays out over the course of 90 days and it matches exactly this profile event.

What this tells us, this sort of red hatched area here is kind of the space where misdiagnoses are happening. It is very hard to imagine that this actually reflects anything else. And the value of this kind of a measurement approach is that it is kind of agnostic to where it is the process defect did or didn't happen. It just points us to a problem that patients care about; that is, the outcome of interest. It is that they didn't want to come back with a major stroke after a minor stroke. That was what they actually wanted prevented

through prompt treatment at the index event. And I think that idea of being able to separably measure outcomes and process is an important one that we should sort of maintain as we think about our measurement framework as it is developed. So, that is all I have to say.

CO-CHAIR DANFORTH: We will go David and then Mira.

MEMBER SEIDENWURM: I thought this was great. Thank you so much.

I think a big point that runs through, in my mind, in the theme of what you are talking about is the degree of required certainty. And I think that the example of headache and dizziness and stroke is -- I am a neuroradiologist. That is part of where I live my life. And the problem is, of course, the overwhelming majority of disease and headaches don't have strokes and the degree of certainty that is required and it also goes back to the theme of time because the ER has to move them along quickly.

And so in pursuit of this patched

area, we wind up overtesting. We wind up harming people probably through overtesting. And yet, at the same time, we don't necessarily have the tools to find in real time those dizzies that will become strokes because we don't always have the tests that would show us that because the vascular structure that becomes occluded, for example, might not be readily apparent. The source of the clot might not be readily apparent through imaging or whatever. So, there are metrics and there are guidelines that attempt to minimize this but, in spite of that, we fail.

So, I am concerned if we define this as diagnostic error we will lose credibility with the medical community because, conceptually, it is a diagnostic error. But we have to really be careful about that if we want to bring our constituency along with us and have the credibility to improve the practice if we define as error a situation in which everything was, and I will put this in scare quotes, done right.

MEMBER NEWMAN-TOKER: So a couple of

points there that I would like to unpack. I think the last point you made is really important for this group to be thinking about. And I think this issue of whether we are scaring people away with the idea of the word error. And this is really, I want to make sure that we understand that this is a terminology problem and not a concept problem. So for better or for worse, the National Academy of Medicine defined the diagnosis label failures that I showed you, the blue circle thing, as a diagnostic error. And when we say diagnostic error, most physicians immediately think blame and process problem, like I did something wrong. And if everything was done right, then how could it be a diagnostic error. And that is a terminology fight that we should not have in designing the measurement framework, in my opinion. We should acknowledge that it is important and we should acknowledge that we may want to focus on the positive side but I am not sure that we should debate whether it should or shouldn't be called a diagnostic

error because I think the National Academy of Medicine settled that for better or for worse. There are arguments on both sides.

That is one point that I think is really important that we have to maintain. I think the separate question is the one that you mentioned about well are these things preventable. And in this particular case, it turns out that they probably are. We have shown, for instance, not to get too specific about it but we have shown that bedside examinations of eye movements are more accurate than MRIs in detecting these strokes in advance. And the problem is, just at the moment, that that expertise hasn't disseminated and we are working in a clinical trial that I alluded to before to disseminate that expertise using devices and so on and so forth.

But the fact that we may or may not know whether these individual events were preventable or not is, to me, less the point. When we think about the diagnostic quality

framework, the question is if you use this kind of an approach to say okay, look, I am going to look at the top ten diseases and I am going to have a dashboard that says these are the places where I see these potential harms that are potentially reducible there, I am now going to investigate both why those harms are occurring and whether there is something I can do to prevent them and then actually use this kind of a measurement to, essentially, if you could imagine putting your finger on top of that hump and squishing it down to the baseline, that if you could actually do that in practice, you would have done something meaningful. That is, you would have presumably reduced the harms from missed stroke in this particular case.

So I think that that idea of preventability, we should not be scared away from measuring outcomes simply because we don't know in any given case either what the process defect was or necessarily the exact solution to the problem, although in this particular case I think

we do.

CO-CHAIR DANFORTH: We are going to go Mira, Martha, David, and then Nicholas.

MEMBER IRONS: This is just a question about the graphs there. When you look at the patients that we are seeing in that prodromal period, were the signs, symptoms, and historical information that was presented at the time to those physicians consistent with evolving stroke or just consistent with dizziness and headache?

I mean I guess what I am trying to get at is is this a population that we are going to attribute to misdiagnosis or whatever we are calling it, delayed diagnosis, a diagnosis that you can actually subdivide into those people who really had symptoms and signs of evolving stroke or did it just not evolve far enough to get to that point? And that is just a --

MEMBER NEWMAN-TOKER: So if I can paraphrase, you are asking could it have been prevented, had someone been there who was aware of what to do.

So, we don't know that from these particular studies because these studies were in 180,000 patients. They weren't done with chart reviews and so on and so forth. They were actually done from datasets where you can't do chart reviews. But the question you asked has been answered repeatedly. That is, we have many other studies. We have done chart reviews on this specific issue and they all tell us the same thing. We are doing the wrong examination at the beside; when we do it at all, we are interpreting it incorrectly 80 percent of the time; and we are ordering all the wrong tests on these patients. So we know exactly what the problem is.

So I can't tell you for these individuals in these graphs but I can tell you that we have seen it over and over again in multiple studies over the course of the last decade.

MEMBER IRONS: So it is a heterogeneous population, these people that are coming in.

MEMBER NEWMAN-TOKER: Well, it is heterogeneous in the sense that there are probably different bedside features of patients who have stroke versus don't. We have actually shown that you can prospectively differentiate between the two. But most of these people are coming in with relatively monosymptomatic dizziness and being sent home. If they come in with dizziness and they are paralyzed on one side, those patients don't get sent home.

So these are the people who are the subtler ones, what you might call atypical presentations. And as was mentioned by David earlier, only three percent of those dizzy patients have strokes. So this is a small population but that is 75,000 people a year in the United States. So, the question of whether it is a small population is really linked to the issue of whether we should -- the total number of patients is how we should be determining whether something is worth studying in terms of the public health burden and our ability to do better

is what should determine whether we should go after fixing a process problem, not the exact prevalence of the incidence of the disease in this patient population.

MEMBER RADFORD: Thank for your those comments, really all three presentations. And we are sort of honing in here on measurement as our focus for this group, which I believe is where we need to be.

And I completely agree with you, this is a labeling issue. I would label these measures as possible misdiagnoses because really what you are doing is you are enriching your case review population for cases that might show you an issue. Speaking from a hospital CQO perspective here, I am always looking for ways to improve the effectiveness of my very robust case review process so that I can find more stuff to fix. And I would love to have these kinds of measures, essentially, available to me to help fix things that would delay these kinds of diagnoses.

I think also your focus on defining cohorts where you can enrich these findings that are basically enriched for possible problems is also a productive way for us to spend our time. Cancer has been mentioned. Stroke is mentioned here. Chest pain is another one where I sort of live and breathe. So those kinds of things I think we probably need to take into consideration but I agree with you about the labeling and needing to do a very deliberate due diligence on that.

MEMBER NEWMAN-TOKER: Yes, just on the last point, I think pretty much if you look at the literature on this from various diverse types of measures, if you group things into three big buckets of missed cancer, missed infection, and missed vascular events, including stroke, heart attack, pulmonary embolism kind of stuff, that a huge, probably about a third of all diagnostic errors and maybe more than half or two-thirds of the harms from diagnostic errors fit into those three big buckets and I do think that there is

something to be said for taking a problem-specific or disease-specific focus for identifying kind of the big killers sort of areas and saying that the measurement framework needs to account for that specificity as well.

MEMBER RADFORD: And your techniques of looking both forward and backward are very helpful as well. Thank you.

CO-CHAIR DANFORTH: David.

DR. HUNT: Yes, I think we are going to, time and time again, mourn the loss of the use of a differential diagnosis in our documentation. From a specialty that has a suitable diagnosis of an acute surgical abdomen, which is no real diagnosis but it gives you enough to say that we need to go in.

And I think that one of the questions that we may come across when others look and critique our work product is that we may be working at the bleeding edge of Occam's razor, if you will. That is to say patients can have more than one condition. Your example, they actually

could have had reflux and an MI. So we have to have -- I think we should have some provision for -- and I know this is a horrible framing of this term but I can't think of any better way to do it -- a p-value for our measures. That is a measure of the accuracy of our measures, if you would, to provide the possibility that it was not an error, that it was -- well, that it was something other than that.

MEMBER NEWMAN-TOKER: Yes, I think that the beauty of this kind of measurement, at least on a large scale for a health system or maybe a hospital is that those gray bars on the second graph on the right hand side are 95 percent confidence intervals. So those are your p-values. And you are absolutely right. You can have both. You can have reflux and an MI and they can be causally unrelated. But then if that is true, this pattern shouldn't exist.

In other words, this pattern only exists where there is this initial spike and then a return to a base rate because we missed

something. It doesn't exist if it is just a random association. If it is a random association, it should look like the heart attack one. So, it should basically be I was told that I was dizzy because I had some ear problem and then I came back with a heart attack -- random association. That dashed line of MIs after dizziness are clearly flat, if you count the 95 percent confidence intervals. And what that tells us is that we are not missing heart attacks in dizzy patients but we are missing strokes. And I do think that having confidence in those measures, statistically speaking, is actually really important and one of the nice things you can do with large data sets.

DR. BURSTIN: And maybe just quickly David could define Occam's razor so we are all -- not everybody lives in that world and it is a helpful concept as we talk about diagnosis.

DR. HUNT: The idea of parsimony. That is to say that the explanation for a set of events or phenomena should go to the least, the

smallest set of possible reasons. That there is parsimony. I'm not sure if I explained it that way.

MEMBER NEWMAN-TOKER: The competing principle is Hickam's Dictum --

DR. BURSTIN: Yes.

MEMBER NEWMAN-TOKER: -- which is the patient can have as many diseases as he darn well pleases.

DR. HUNT: Kerm said it better than I. It is the simplest explanation for things.

CO-CHAIR DANFORTH: We are going to take the three cards that are up. Then we are going to take a break so that we can try to get back on schedule and go to Helen's presentation.

So we are going to go Nicholas, Hardeep, Mike, break, Helen, in that order.

MEMBER KUZMA: All right. I had a thought or question, depending on how this comes out. So I think Hardeep did a nice job showing that medical records are really a great source for identifying when these errors are happening

and you showed how administrative data can possibly identify that. And I am just wondering are we asking patients when these errors are happening and is that kind of an untapped area that we could be using to kind of figure out when these errors are happening, in addition to the other things that you presented?

MEMBER NEWMAN-TOKER: Yes, I mean Mark has suggested this a long time ago and I think it is a very interesting, totally untapped potential source of data to actually just ask the patient whether their diagnosis has changed. And I do think at some level, eventually, that may be a filter to kind of capture a broad spectrum of diagnostic errors.

The trick is that you have all the same potential validity concerns and reliability concerns, depending on the patient's health literacy or who they did or didn't happen to see. In other words, if you call somebody in 30 days and say well have you had a stroke, that answer may be clear. But if you say was your diagnosis

correct of whatever it was that they had, the yes or noes may not be as accurate as you want them to be.

But I do think it is a totally untapped source for data.

MEMBER KUZMA: And it is not like the data we have now is perfect anyway.

MEMBER SINGH: I'll start with and just quickly respond to you. So I want to clarify medical records still are one of the best sources of truth of what we know but it has to be often complemented with other sources, such as talking to physicians and patients or other care team members.

Patients still reported as huge area, and AHRQ has done work in this area very recently and has a report -- we actually have -- I have a mentee who has applied for a K Award in exactly what you just mentioned, but she has applied to Agency for Healthcare Research and Quality, which we hope will be funded.

And so you know I think again it begs

the question: how are we going to advance the basic science? So I think this is a very untapped area. There is clearly concerns about validity of the information or what patients are telling them. Do they actually mean what we are thinking that they mean. And so it needs to be researched.

David, to your question, quickly, on differential diagnosis. So actually there is sort of a measurement concept that we have proposed in the Journal of Patient Safety paper that I talked about about differential diagnosis. So definitely that is sort of a shout out for that paper that you should read. Really important.

But I want to revisit your p-value concern. I really think we do need to think about -- I'm not sure whether we are going to come up with the measure -- what we have done is we have developed these concepts and then measured the predictive value. So the positive predictive value of whatever we are trying to

measure.

So for instance, if a patient comes in to the primary care clinic and then got admitted unexpectedly in ten days, we look at the medical record at the first visit to see was there a missed opportunity. So the term that we used in our work is not error as much but it is missed opportunities. Was there a missed opportunity in the first visit? And so the predictive value of that would be about 21 percent is what we found.

We have done some recent work in test results. So if there is abnormal test results as flagged by the computer that does not have a follow-up action, such as FOBT, which is occult blood in the rectum -- occult blood in the stool which is not followed up by let's say a colonoscopy. So we build algorithms to identify follow-up actions on abnormal tests and make the algorithms smarter and smarter to figure out whether the test results were surely missed or not. And the predictive value of that, after you look at thousands and thousands of medical

records and identify just a small cohort, what we call sort of picking out the needles in the haystack, when you look at that cohort, 60 percent predictive value right now we are getting on medical record review. So we confirmed that with medical records and 40 percent of the time when it is not a missed result, it is usually when the patient has either refused the follow-up test or they have gone to another system and gotten a colonoscopy for instance.

MEMBER NEWMAN-TOKER: So let me just respond briefly to the issue of the chart reviews. I think Hardeep is to be commended for the tremendous amount of work that he has done in this space and there is a sense in which the chart review feels like somehow the gold standard because it has got such high face validity to clinicians. What was the story in the chart?

We have seen with this particular problem, with dizziness and stroke, that fewer than ten percent of the relevant facts are documented in the charts. And when they are

documented, 80 percent of the time they are wrong, physiologically incompatible physical findings with the diagnosis rendered by the clinician.

The problem with charts is that it is highly likely that they are biased towards underrepresenting the key facts that actually allow you to determine whether or not a diagnostic error occurred. It is precisely because of the failure to record that information that the diagnostic error was made in the first place. And as a result, the chart review is great for identifying what amounts to gross negligence and I think Hardeep has done that in a lot of his studies, has shown that there is a lot of gross negligence out there and missed opportunities. But it is not a great gold standard and we shouldn't think of it that way. We should think of it as the best available means to get the information we really want, which is what were the patient's initial presenting symptoms and signs. And we may think of in the

future about ways we can get that information more reliably than the way that we are getting it now, which is reviewing documentation that we know stinks.

CO-CHAIR DANFORTH: Mike.

MEMBER DUNNE: I think I forgot what I was going to say.

I am naive to this process so bear with me. And I am not a clinician. I am a clinical scientist. But having read a few case studies in the New England Journal, you know that the initial diagnosis and the final diagnosis are always the same. No one has missed a diagnosis in a New England Journal case study.

That said, so you have the initial or differential diagnosis that is created after the patient interaction. That gets honed down to a working diagnosis. Then you have whatever happens to the patient either good or bad and you have a conclusion or a consensus final diagnosis.

The final diagnosis could have been part of the differential diagnosis but not

carried forward. So you get a point.

On the other hand, you can reconcile the final diagnosis with a working diagnosis to see if there are improvements in the process that could have been made that would have carried through from the differential diagnosis all the way to the end so it looks more like a New England Journal case study.

Not only that but you can measure the difference between the working and the final diagnosis, yes/no, almost like in horseshoes. And I don't know how useful ICD-10 codes are in helping you to determine what that was but that is another source of data that you can look at.

So that is a relatively simplistic point of view from a clinical scientist standpoint but I keep on hearing this throughout the discourse and it is the only thing that I can really hang on to.

MEMBER NEWMAN-TOKER: So just to follow-up on that last point, it is absolutely critical that we try to be able to do that. Like

it would be great to be able to measure the discrepancy between initial diagnosis and final diagnosis or working diagnosis and final diagnosis. And there may be things that we should advocate for in terms of the way the system records data to actually be able to do that but right now, those data usually don't exist. And so those are perfect data that would be useful for this purpose and, in fact, we have to kind of back our way into assumptions about it.

So for instance, we say well, if the person was given an ICD-9 coded diagnosis of an inner ear problem, then probably the initial diagnosis -- the initial symptoms were dizziness and so on and so forth.

MEMBER DUNNE: And we always throw software and solutions like that. But when I was at Barnes Jewish, there were 27 software interfaces between laboratory information system and the various clinician groups within the hospital -- 27 individual software programs

deciphering information from laboratory to give the results to clinicians and nursing staff.

CO-CHAIR DANFORTH: We are going to have more time over the next two days to talk about some of the things we are getting into right now with these questions like measure concepts, gaps, data sources.

For now, we are going to take a quick break, maybe ten minutes if we can. When we come back, we are going to get to Helen.

I did see that two people had their cards up. So, if the questions were specific to David, grab him at the break. If not, when we move on to measure concepts, make sure you bring them back up.

(Whereupon, the above-entitled matter went off the record at 11:07 a.m. and resumed at 11:24 a.m.)

CO-CHAIR DANFORTH: So Tom, I hear that you are still hanging in there with us. So If you have a question, please let us know.

MEMBER SEQUIST: Yes.

CO-CHAIR DANFORTH: And I am going to actually turn it over to Helen Haskell now, who is going to share with us the patient perspective on diagnostic error.

MEMBER HASKELL: Thank you for asking me to speak. This is going to be a lot less scholarly, to say the least, than everything that has preceded me.

So I want to talk a little bit first about why I am here. So I came into patient safety and diagnostic error through the death of my young son, Louis, who died in a hospital of a perforated ulcer caused by NSAID pain medications following elective surgery. He lingered 30 hours before he died with no one doing anything for him.

It was considered an egregious circumstance when it happened and yet I am still hearing of similar, sometimes nearly identical cases now nearly 16 years later. And one of the things that sort of an earlier conversation has made me think I should add was that I was really

the only one who sort of put it together. I was the only one recognized that he was going into shock, or at least the only one who said -- I didn't say it but no one said it to me either. I don't think anyone else recognized it. And after the fact, I was the only one who really did a meaningful root cause analysis.

The people who were involved did not want to see it or did not see it as something that had an explanation. They saw it as just one of those things.

So I should also add that I work with patients who have been harmed by their medical care, not necessarily error. At this point I have talked to literally hundreds if not thousands patients and diagnostic error plays a role in a large proportion of those cases. So just as diagnosis is central to patient's experience of the healthcare system, diagnostic error is central to their experience of harm. So this has been a real source of frustration to patients that errors of judgment and other

factors leading to diagnostic harm, diagnostic inaccuracy, have received so little attention until recent years in the safety and quality movement.

So the most important thing that patients say overwhelmingly is that they are not listened to and that their input isn't valued.

So even the language of the IOM report, and I was glad that you sort of acknowledged that, Mark, that some of that has also filtered into our diagnostic accuracy framework, but even the IOM report misses this point or misses the significance of this point when it talks about things like communicating the diagnosis to the patient. You know the whole concept of a working diagnosis in a conversation that includes the patient's input every step of the way is a really critical one. If this isn't invited, the situation is ripe for misinterpretation.

So in our own case, what happened to my son, Louis, was called a misdiagnosis but it

wasn't really any one person who misdiagnosed it because they were all listening to each other so that what apparently happened was that nurse's station chatter sort of became enshrined as a diagnosis. So there was never really an independent assessment. They were not listening to the patient.

So the culprit wasn't an individual; it was a mindset of superiority and of trying to save time, which of course they didn't do in the end. And it speaks to care coordination, which is probably the biggest problem that patients face, not just between specialists, although that is huge, but also within practices. For example, people not talking to each other, just layers of people you have to get through before you get to the person who is actually making a decision.

It speaks to culture of safety, patient-centered care, patient engagement, and collaboration -- not just a team, but a team with the patient at the center of it.

So we have measures for a lot of these

things and I think we need to range further afield to find more and be sure that we include more of the ones that we have on our list.

So in no particular order, here are some of the things that I consider important, as a patient:

Encouraging structures and processes that allow for constant patient input. So you are talking the medical records. You are talking about patient reporting but it needs to be ongoing throughout the process.

So there are things like user-friendly patient portals; bedside rounding in hospitals; open notes; the ability to annotate your own records; the use of technique like teach back and checking back; giving patients a record of their visit, which McDonald's will give you a free meal if they don't, but no providers do it unless you ask. And they are not always happy if you do ask.

And we need, as has been discussed, to find a way to encourage feedback to doctors about

the results of their own diagnoses or diagnostic errors so they can learn from experience. I'm not sure how we capture this but we need to.

We also need patient involvement. I think it is critical in the improvement process. So patient reporting of experiences often captures things that are not otherwise documented. As we know, patients report really different things from their providers and they usually consider their condition more serious and it often is actually more in line with their physiological symptoms.

It is an entirely different focus. Patients value their global well-being more than disease-specific components. So we need to capture that focus because that, I think, is what we need to be aiming for in terms of healing.

So patient reporting of symptoms, patient reporting of their experience within the medical milieu, which captures problems that often aren't otherwise documented.

I am a huge proponent of patient

involvement in root cause analyses and other quality efforts. Patients, in most places, are not even interviewed if there is an adverse event, so that all of the sort of information that I, for example, had in our son's case was completely lost to them because people who realized that they had not done what they should have done were not going to be forthcoming after the fact. And while it was going on, they didn't understand what was going on.

So that is pretty much what I have to say. I would want to add to some of the other cautions that people have expressed about using standard quality measures in this context. I do think this is a new area. Diagnosis has its own set of issues. The communication, the cognition, the structure that we have been talking about this morning, I think it is a new area and it is a new opportunity to sort of develop new things and I am concerned that it could lose its impact if we get bogged down and sort of going back over the same ground that we have had before.

So that is about all I have to say.

MEMBER CAMPISANO: I just wanted say thank you for your comments. And I think that you captured something really, really important here. Obviously, I am biased as a patient advocate and a former cancer patient, but I think that the patient perspective is really important here, and the communication between provider and patient is something that has been touched on a little bit today, but I think you emphasized how critically important it really is and not just -- I guess I just wanted to say thank you for touching on that. It is something that I don't think is talked about quite enough.

MEMBER HASKELL: Thank you. I think it is not -- I think the importance of it is really not realized at all in the structure of medicine, the ways -- patient communication is just sort of discounted and really the most vital information is there.

And if you filter it through, so if you filter it through the provider as a history

and physical, which I think are very important and that was among the things I didn't say that I think we really need to encourage more of emphasis on history and physical, more communication with the patient, at the same time, even the most skilled provider, just by the act of communication and relaying to a third party, the third party being the chart, there is going to be error. So you have to have the patient input to say no, this is an error. This is what I really meant because none of us can communicate that well. It just doesn't happen.

MEMBER CAMPISANO: Yes, and I know that -- well two things. One, I don't remember who said it, but in my experience and at a meeting that I have been at before, a woman said that patients are oftentimes the most knowledgeable about their particular disease and what is going on with them. And I think that they are often not given the credit because they might not have the medical background or know all of the correct terminology, but they are often

very knowledgeable about their particular disease.

And the other thing is that -- I lost my train of thought but I just think that yes, the communication -- oh, the time factor that we have touched on. I understand there are time constraints. As a mom and a cancer patient, I feel like -- I am a wife and a lawyer. There are time constraints in everybody's lives. I get that.

But one of the very best meetings I ever had with a physician was a physician who sat down with me, an oncologist who sat down with me and talked to me for 45 minutes about what my diet looked like. What did my exercise look like? What my charts had looked like and talked to me about the results that he saw and how he thought they were fairly abnormal for what diagnosis I was being told. And I realize not every doctor is going to do that with every patient. It is not possible. But I think just to feel like I was being heard was invaluable

really.

CO-CHAIR DANFORTH: David.

MEMBER NEWMAN-TOKER: So maybe I could ask both of you to weigh in on this issue of engaging patients -- that came up before -- and engaging patients in measuring diagnostic error.

So it has been bandied about that we could call patients up at 30 days or 90 days or whatever and say did you get the right diagnosis. Presumably, if you did that as a cold call, you would probably get a pretty negative response from patients in the sense of what do you mean I might have the wrong diagnosis and wait a minute, I don't get it.

But to what extent do you think we could legitimately engage patients in that dialogue from the get-go, expressing the uncertainty and then actually leveraging them to either contact us or to respond to our contacts to them to actually give us the feedback on when and whether the diagnosis was right or wrong. Do you see that as an avenue?

MEMBER HASKELL: I think patients vitally interested in their own healthcare. I think you can engage them easily from the beginning. You know I would say you engage them while they are in the doctor's office or in the hospital. That is what we have done on some projects I am associated with, exit interviews.

But patients are very interested. I mean nobody takes cold calls. I mean you obviously haven't been trying to do any polling recently but we can't get people to answer the phone anyway.

You know but in terms of -- and the patient portal is also a huge, if you had something that people could write in, then you would be able to have an ongoing conversation in their medical record, which is where it belongs, I think.

MEMBER CAMPISANO: Yes, I agree and I think some pharmaceutical companies are developing patient engagement techniques specific to certain drugs and I think that doctors'

offices could follow that lead to some extent.

You know my nurses would text with me and just say hey, I have got your lab results or how are you feeling today. And I think I had a pretty exceptional oncology office in that my nurses were willing to do that. But I think just keep in mind that patients are people.

MEMBER HASKELL: And what about enhanced registries? They are also doing that sort of thing.

CO-CHAIR DANFORTH: What I was going to say, Helen, is I had a conversation recently with some folks writing in the Pediatric NSQIP Registry. And they do contact patients after the visit to ask if they had unexpected return to the emergency room, an unexpected return to the operating room.

We have been hearing from some ambulatory surgical centers that they are contacting patients to ask about unexpected complications that occurred after discharge and I think there are opportunities to prepare a

patient for a call or to let the patient know they are going to have the opportunity to provide additional information about their care.

And something that we have learned at Leapfrog since we started assigning safety grades to hospitals is that patients are dying to tell you about their stories sometimes. I mean we have received so many patient stories, frankly, we weren't ready to receive. We have had to partner with ProPublica who has a patient story project, to help sort of deal with the number of patient stories that we received.

So I think given the opportunity, patients do want to share their experience, at least from what we have seen at our organization.

And I see Marilyn.

MEMBER HRAVNAK: I had a question about communication. So do you feel that it was not only communication but the navigation of that communication? In other words, not only speaking but speaking to the right set of ears that could do something with the information? I mean how

much of an interplay do you feel there was there?

MEMBER HASKELL: I am wondering if this is a setup question because that was my big focus at the time. We actually passed legislation in South Carolina about provider identification and navigation in hospitals because we felt really trapped in the hierarchy, in the teaching hierarchy. We didn't know. No one tells you how the hospital bureaucracy works and it is very complicated. So when you get someone who is not seeing something, you really have a hard time getting beyond them.

So yes, that is very important and I think provider identification and rapid response teams, patient access to rapid response teams.

And incidentally, I did a close analysis of rapid response calls to see what was causing that because one of the things that -- I think on the one hand when people think diagnosis they think outpatient, and I know that the statistics says that most of it occurs outpatient but when it occurs inpatient, it is much more

critical. The patients are helpless. They can't leave and go to another provider, even if they are not terribly sick. They are really stuck there and they have to have some kind of a safety valve. So, navigation is critical.

And we even got, at one point, we were advising people to call 911 because it at least gets some attention.

MEMBER HRAVNAK: Right and it is that whole idea of escalating care, too, and how care is escalated. And I think your point that in the inpatient setting, it is not frequently a primary diagnosis that we are making but it is the diagnosis of a complication --

MEMBER HASKELL: Yes.

MEMBER HRAVNAK: -- which can be as important, if not more important sometimes than the primary diagnosis in that case.

Thank you.

MEMBER HIGUERA RUEDA: Thank you. Thank you for those comments and certainly communication is key in the measurement of

quality.

I just wanted to bring back the issues that there are many venues, currently, for measure patient communication and patient satisfaction. At least in our hospital, we measure HCAHPS scores for the inpatient, and we have similar scores for outpatients. But we have to be careful of who we establish measurement of quality because there is plenty of evidence that shows that there is really no correlation between good quality and those HCAHPS scores.

And you mentioned that there are a lot of patients that want to bring back their stories. Actually from a sample of patients that get these surveys, only 10 to 15 percent sent back these reports. And there is evidence also out there that shows that a fair amount of those reports are more negative than positive. So patients have more tendency to report when they have a bad outcome than a good outcome. And I think that it would be unfair to necessarily make assumptions that someone is providing poor

quality just based on those scores.

So at the same time there is a very relevant important tool, I think that we have to be savvy enough and cautious to necessarily establish that that has to be an absolute measurement of quality.

CO-CHAIR DANFORTH: David.

MEMBER HASKELL: Could I address that? I mean I think HCAHPS, and I am not as familiar with the other CAHPS scores but I think HCAHPS is really specific enough that it might useful in this context so that you could associate, for example, did the nurse come when you rang the call bell? Was your environment sanitary? You can sort of associate those things with cause and effect. I don't know, that is just my thought that those would be a useful tool in the context of other things.

MEMBER SEIDENWURM: So first off, I want to thank you so much for sharing your story and emphasizing the perspective of the patient experience because sometimes we all get bogged

down with some of our technical concerns.

But with respect to the point that Carlos made, which are all true, I regard those as features of the system rather than bugs because most of the time, I hope, medical care is pretty good. You know it does what it is supposed to do. And what we want are enriched samples of the times that it doesn't work so well because if we start out with a low prior probability, we are going to falsely -- we are going to have false positive assignments of poor care.

So I think that by starting with some of the problems that you discussed, I would argue that instead of being problems they are actually features of the system that we ought to try to exploit and that maybe in our measurement efforts, we can focus on the denominator there as an enriched source.

Now, there will be other false negatives because some of the bad results don't get reported in those ways but at least we might

start with a higher prior probability and, therefore, might wind up with a greater predictive value for identifying correctly the areas where the system didn't do its job.

MEMBER MCDONALD: Thank you, Helen, very much, as usual, for your insights on this.

I wanted to ask you to maybe explore a little bit more about this side about explaining to the patient what the health problem is, so the definition of diagnostic error where it actually is a diagnostic error if the explanation wasn't there.

How do you think about what would be sort of the gold standard of explanation and sort of patient involvement in that piece of the definition, based on all the patients you have talked to.

MEMBER HASKELL: Well, I think they would be a discussion with, rather than an explanation to. So, I don't see that as just semantics. I mean I think it needs to be sort of interactive. Does this fit with your symptoms,

as you understand it, not a fiat from above, but a conversation to make sure you have got it right.

You know one of the things we worry so much about whether the patient understands us, and I think it is really critical to know if we are understanding the patient just to make sure that everybody is singing from the same hymnal, which is a pretty darn hard thing to do.

MEMBER MCDONALD: I would like to say I was thinking, as you were talking to Jen, that there is the part that the patient knows something that may not be picked up and so giving that opportunity more room to flourish. There is also the chance that the patient is missing something about the context that would be important to the patient's next actions. So, an emergency room not getting a particular diagnosis doesn't mean that the diagnosis doesn't exist. It just means that the emergency room was just trying to figure out something that was going to be really urgent like right now and patients may

not recognize that if they continue to have symptoms that it is perfectly fine to continue to seek care.

And so the explanation that says that is part of that interaction with the patient that is important in the diagnostic journey for the patient and for the system to have the patient get the diagnosis that they need to get. So, it is all sort of all sides of that.

So, I like your adaptation of that. It was nice.

MEMBER HASKELL: This is a tangential point but something else that I meant to say and I somehow overlooked it, in the context of my son's case but in the context of many, many other cases, the misdiagnosis of drug side effects and reactions, which is just rampant. And I think that I don't know exactly how it could be measured but something that would really encourage more learning about drug side effects among both providers and patients and providers at all levels. I mean everybody misses it when

there is a drug reaction.

CO-CHAIR DANFORTH: Thank you so much, Helen.

I want to turn it over to Tracy now to talk about something that we started talking about this morning, which is terminology and definitions.

DR. LUSTIG: Thank you. And I am just going to queue this up to get you all talking about it again, although we have talked a lot about this already.

And just to reiterate from this morning, and as a reminder, we are not looking to recreate what was done in the Academy's report. We are not looking to redefine or come up with a new model for the concept of diagnosis. We are really trying to focus here on the measurement part of it.

And just as an example, as you all know, we already have multiple definitions and there is probably even more than these. Next slide.

And so I think what our conversation needs to be, we don't need to necessarily decide this here and now and today, which is we have the terms diagnostic error and diagnostic accuracy. In our last call, there was a sentiment that we didn't like using diagnostic accuracy for the purposes of this project. And one of the questions we raised and want to talk about now, and we already have started to do this is whether we really need to focus more on quality.

And just as an example, and these all came up this morning, we could call it diagnostic quality, the quality of diagnosis, the quality of the diagnostic process. And I know those are all nuances and that is why we wanted to kind of open it up to discussion.

And then one of the other things that did come up this morning as well and we have talked about is whether we need to explicitly call out safety to call it diagnostic quality and safety or quality and safety of diagnosis or do something to reinforce the IOM definition of

quality that already does include safety as part of quality.

MEMBER NEWMAN-TOKER: Could I just ask that either David, or Helen, or you, Tracy, give us a capsule summary of what you perceive to be the major differences between the umbrella of quality and the specific narrowly-defined issue of safety just so that we are all on the same page about kind of the scope and scale of how they differ?

DR. HUNT: I can start in. I am a big fan of a previous IOM report that was authored by Paul Tang. Actually, it was commissioned by AHRQ where he wrote early on that safe care is indistinguishable from quality care. And by saying that, what he said and that whole group said, is that safety is a subset of quality, that all things in safety are related to quality but not all things in quality necessarily have to do with the specific issue of patient harm.

So I am a big believer, and not that that has to be the guiding voice throughout this,

but I am a big believer in that idea that the safe delivery of care is indistinguishable from quality care.

DR. BURSTIN: I completely agree with David, and I think that our perspective has been that safety is very much part of quality but, at times, I think my personal opinion, like in this instance, calling it out separately has an additive effect. That may be something we want to consider.

MEMBER SINGH: And I am just going to quickly add from our perspective in our work, safety is sort of the foundation. You have to get that right.

You know United is terrible as an airline -- terrible. It is like one of the worst airlines but I haven't crashed yet. So that is what is important to me, for instance. So that is sort of the way I would say.

MEMBER SEIDENWURM: Sorry. I'm sorry. I lost my train of thought in your remark. That was such a great remark.

MEMBER SINGH: I didn't mean to be that dramatic but you know.

MEMBER SEIDENWURM: But I guess when I think about the safety of the diagnostic process, I think about the harms of diagnosis. There are harms in achieving diagnostic certainty from false positive diagnoses, incidentalomas, the harms of the diagnostic process itself, a pneumothorax from a lung biopsy, for example.

So I think that when we talk about safety here, we want to talk about making sure that we calibrate our degree of certainty required with the degree of harm. And there are numerous example that I see in my practice on a daily basis. And we can debate about whether dizziness in the ER is one of them, or we can talk about something that is more clearly the case -- for example, incidental adrenal nodules. You know you see we probably caused more cancers following these up than we detect net-net-net, timing differences and blah, blah, blah but we really have to be cognizant of that in the quest

for diagnostic certainty. Are we being safe?

DR. HENRIKSEN: You know it is a matter, I think, of -- you know we are at the National Quality Forum here. It is not called the National Safety and Quality Forum. It is the National Quality Forum, and so it is probably important to recognize the other components of quality besides safety.

And the second IOM report, Crossing the Quality Chasm, identified five other components of quality, in addition to safety. And that was effectiveness, efficiency, timeliness, patient-centeredness, and equitability in terms of care.

At the same time, just as Hardeep mentioned, but you know a focus on safety can serve as a spearhead and create the coattails for issues of quality to be discussed and measured and focused on as well.

So it is a matter of -- you know the Joint Commission's Journal is of quality and safety. Our Center at AHRQ is the Center for

Quality and Improvement and Patient Safety. And so we like patient safety because it is the spearhead that sort of gets the attention above the centerfold probably on our national newspapers and so it can sort of usher in quality issues with it.

And so it has value, and that is why the two are oftentimes joined together in different types of organizations.

CO-CHAIR DANFORTH: Martha and then I see Helen and David.

MEMBER RADFORD: I am also a huge fan of the IOM construct of six domains of quality, of which safety is one. We are asked to focus on national measurement here and I think it would be -- this is my own personal bias -- I think we should focus first on measuring safety. In other words, diagnostic errors. I think that will get the ball rolling in measuring diagnosis in all the other domains as well, in support of your view as well.

So I just think this is where we

should start and acknowledge that it is a start and not an all-encompassing solution to anything here. That is my bias.

CO-CHAIR DANFORTH: Helen.

MEMBER HASKELL: So my concern is with these definitions is blurred boundaries. How unsafe is the delayed diagnosis? I mean particularly in the case of diagnosis. I think it is an issue all the way across healthcare and I am sure it is an argument that you all have had ad nauseam, but I particularly worry about it in the case of diagnosis.

I just think that you can call it whatever you want, I suppose, but you need to focus on the whole package because it is all about safety, and it is all about quality.

CO-CHAIR DANFORTH: David.

MEMBER NEWMAN-TOKER: Yes, I just want to make this point about efficiency and effectiveness. I do not think it is possible to have a robust national discussion about diagnostic error and safety without talking about

the flip side that David keeps bringing up of overtesting and overdiagnosis.

I think the concern when you do that is, of course, that you are just going to spur people on to do the wrong thing, which is over test everybody in every situation.

So I think you have to make sure that kind of both are part of that package because they are two sides to the same coin. We could call solve the problem of missed diagnosis tomorrow by simply ordering every test on every patient in every situation. And it is clear that that is not good diagnosis. So from a measurement standpoint, you would not want that to be your measurement objective was that everybody got every test in every situation.

So, clearly, somewhere there is some concept of good diagnosis, better diagnosis that has to factor in. You know most of us think of, when we think of clinicians as being good diagnosticians, they are people who kind of get to the heart of the matter quickly and

efficiently with a minimum of testing, and effort, and time and they get to the right answer. I mean that efficiency component of actually getting the right answer without doing lots of unnecessary tests and without taking three months to do it is, in fact, core to the notion of good diagnosis, and I don't think you can kind of pry those apart by saying well, we only care about whether we got it wrong or not. We have to actually talk about whether we are doing the right thing.

CO-CHAIR DANFORTH: Hardeep.

MEMBER SINGH: I want to underscore what Mark was saying about focusing on safety because that will lead to conversations about everything else, which is very, very overlapping.

And I will give you a very concrete example. So we started doing work in the VA where we found a lot of delays in cancer diagnoses and test results that were abnormal and were not being followed up. They were all safety issues. So the conversation started with a real

safety problem, where patient harm was occurring.

And then that led to discussions about communication of test results, where we came up with actual measures on how we are communicating test results to patients. And in that, the timeliness came out. So now we say 14 days and seven days. Efficiency came up where we had this same discussion. Making it patient-centered came up.

So I think all the conversations were propelled and stimulated by a concrete foundation because nobody would argue with us anymore that this is a problem that we don't want to measure because we had such a strong foundation for safety at the core of the issue and it was easier to pass all those measures.

CO-CHAIR DANFORTH: I'll just add -- I am actually going to go back to one of David's earlier points. I was a little surprised, David, that you just said what you said because on one of your earlier slides you said you know we really need to focus on B, which were those

preventable diagnostic errors that we know results in harm. And I think when you talk about it in that way to start the conversation with providers, and hospital leaders, and even patients, this was a preventable diagnostic error that resulted in harm, it is something that everyone agrees shouldn't happen. When we start to move into defining efficiency of the diagnostic process, that is harder.

So I just wanted to add that point. I know you have your card back up and I am going to let you respond and I do see David back there. So, I am going to go to you after.

MEMBER NEWMAN-TOKER: So just to clarify what I meant by that. So I think we have to separate our mission and our process in some sense.

So I think the mission here is to develop a measurement framework that gets us to a point where patients don't suffer harm from diagnostic errors. And I think that should remain at the core of the mission. Like as we

start to get around the periphery of every other imaginable detail in the quality spectrum, that we need to have that be our touchstone or our North Star.

That having been said, I think we risk, by being solely safety-focused in a developing a measurement framework, setting ourselves up for a degree of failure. And what I mean by that is let's make it concrete. Let's pick the dizzy stroke thing where I know something about it.

So, what you heard before is what I hear all the time. The first question that David asked today well aren't you going to just over test lots of people and not just waste resources but harm people from incidental findings? You know in the end, you are going to harm more people than you save, so on and so forth. I actually think if we don't at least address that in some way kind of head-on, we risk having people say these people don't understand what diagnosis is in the real world. And I think it

could happen in one of several ways.

I think you could say okay, look, we are going to actually explicitly say that for x class of diagnostic error measures, we have some associated diagnostic quality measures that kind of counterbalance on this issue of overdiagnosis or whatever you want to say. Because you could literally say okay well, look, for the stroke thing, you should never measure the frequency of stroke errors without constantly monitoring your utilization of neuroimaging. You can pair those two things together conceptually and say the quality comes out of getting both of them right, not out of just focusing on the one that patients care about as such.

So I think somewhere in there there is a hybrid. And I don't know if for every safety measure there is a quality measure or whether it is saying this time, for this year, we are going to focus on safety and the next time it is explicit it is going to be about the rest of the quality measurement. Somehow we have to address

it at least.

MEMBER SINGH: Yes, but I think that is what Martha was also trying to say that you start with the safety problem and then you start thinking about some of the other things that are in the same equation.

MEMBER RADFORD: Right. For this particular issue, I think once you have the, in a sense, safety outcome measures, to me it is very easy to design that family of measures that is going to give you the quality. Because for example, just to take the point of overdiagnosis, you could say okay you have diagnostic errors and what does it cost. That is very easy for me. They follow one from another easily. Then there is the process measures that also can be developed, et cetera.

CO-CHAIR DANFORTH: Why don't we go David, and then Mark, and Helen.

DR. HUNT: Yes, I think what we are saying, I don't think anything that has been said would exclude us continuing on the foundation

that was laid with the IOM. I mean those six -- when you read the Quality Chasm report and read the deliberations that they went through to come up with those six, I think you have a real good appreciation for the time and the thought that went into them. And they are not six independent degrees of freedom. I don't see any way that you can be equitable without being patient-centered. You know you can only be as efficient as you are safe.

In today's world, timely treatment is equivalent with effective treatment. So I think that --- looking also, I think it is important that we, as a group, look toward the larger process that we are a part of, and it is easier to have policy discussions down the road when the products from groups like this sit on a foundation that was already laid. It is easier to say this is completely consonant with what we have done before, not that we have to walk lockstep in. I mean if we find obvious reason why there should be a difference or that we are

somewhat skewed, I think we should go ahead. But if it is at all possible to stay on that same foundation, it is a much easier policy discussion moving forward when we think about how we are going to go to that next step and that step after that.

CO-CHAIR GRABER: You know I agree with that. And in terms of next steps, for this to actually work, we are going to need all the stakeholders to buy into this and be interested and maybe excited actually to tackle this process.

So I am a little worried that we will alienate physicians when we start talking about errors, the point that David made. It comes up all the time. And I think where we could go astray, I think we all like these six dimensions of quality but it kind of conveys the impression that we know what these are, that we know what timeliness is, and we actually know what the real diagnosis is and we don't. Even the gold standard is highly fallible and not often even

conclusive.

So I am a big fan of including all the six elements but I think in the next sentence we have to acknowledge what a complicated process diagnosis is and there is uncertainty at every step of the way and that it plays out over time. And it is very hard to judge at one point in time whether you are really on the right track or not. You might get a different answer if you waited a day or a week.

So just pointing out that complexity and the uncertainty of the process I think would help us get some parties onboard to go with a comprehensive definition like this.

MR. LYZENGA: And just to quickly add to that, we can sort of -- we can flesh out some of those sort of nuances as part of our report and what we put out with this. And in addition to doing things like emphasizing maybe that safety is very important and should be focused on first, in addition to all these other things, these are kinds of things that we can sort of

build around our core recommendations and add nuance and context, too, as we work through this project so we are not sort of -- you know we don't have to end the discussion here. We can kind of keep working on these things.

DR. BURSTIN: Pretty much just to build on that conversation, I feel like, at times, these discussions get sort of unnecessarily complex. I think we know what we want to focus and safety will largely be the biggest domain that we are talking about. At least for me, I don't want to lose the threads we heard earlier from both Helen and Jen about patient-centeredness, timeliness, communication.

There may, in fact, be new patient-focused kinds of surveys that may not focus exclusively on the safety of a diagnostic error but may give us the kind of information that will be invaluable to really help transform the health system. So it may be that those could still be part of this framework but I don't want us getting caught up in this.

It is a prioritization issue more than it is an exclusion or inclusion issue, and I think we probably could just proceed with safety as the priority but not lose sense of where there are in fact significant issues in this space around equity, significant issues about patient communication, as we just talked about.

MEMBER IRONS: I wanted to follow-up on what Mark said because I think you make a very important point. And I think that we talk about things -- when we look at large data sets, we are looking at things from 35,000 feet and it is hard not to look at those figures and say, "Oh, my God, what is going on in those emergency rooms across those nine states."

But I would like us not to forget that there is a midpoint between underdiagnosis and overdiagnosis because, clearly, it is easy just to default to the imaging study because if you can identify more patients by doing that, it is timely because if you can identify more patients by doing that, it is timely.

But what is really important is knowing what questions to ask. You know taking a good history, actually touching a patient. My two elderly parents have been in the hospital for the last month and they are still both hospitalized and it is amazing how infrequently someone has laid hands on them. Three emergency room visits, this doctor never touched them. Everybody was looking at machines.

So, knowing what historical questions to ask, knowing how to do a physical exam, actually having the time to think about the patient in front of you and having the system, I am a big advocate of making sure that the systems actually support the physicians. You know doctors shouldn't be on the phone talking to insurance companies about preauthorization for imaging studies. That is where it becomes unequitable.

And so having the systems -- I don't want that to get lost in this conversation because otherwise we are going to be talking

about sort of the larger issues and forget about the important pieces of the physician actually interacting with the patient.

MEMBER NEWMAN-TOKER: So just to be concrete about the question that is on the slide, I personally would vote for calling it measures of diagnostic safety and quality, acknowledging that the group wants to call safety out as something important but, at the same time, we want to acknowledge that it can't be the only thing.

And if that meets with the kind of broader political objectives, if it fits that framework by having them both there, then maybe that is a direction that we could all potentially get behind.

I will strongly advocate against using diagnostic accuracy or error as the main header title of what we do; error because it is poisonous, and accuracy because it so laser-focused on the issues of tests and results that it misses the big point of the whole process and

everything else.

CO-CHAIR DANFORTH: Lavinia.

MEMBER MIDDLETON: I put my card up just as you were saying -- I don't want to waste time and I am in complete agreement with what you are saying. But I don't think that error is such a pejorative term. I think in this sphere where I work, physicians are aware that the system is broken and I think having as much focus on the process and the tools that we have available in order to even make sense of all the tests that have been ordered or actually be able to capture them and to create a meaningful differential diagnosis or treatment is as relevant as making a diagnosis.

So I don't think that pulling out accuracy and error or taking it away from the conversation is that bad, but I would include the focus, as you did in the Venn diagram, and not just the diagnosis label but also the process failure and error as well. And I think that is where you get the buy-in from physicians,

especially it is extremely timely as more EMRs are being rolled out, more groups are being consolidated and more patients are having an aggregate of examinations and evaluations from multiple caregivers.

CO-CHAIR DANFORTH: Helen.

MEMBER HASKELL: I just wanted to put in my two cents. As much as I just used it, I don't like the word error for the opposite reason. I think it sort of narrows the definition so people will say well, I met the standard of care so there is really anything wrong with this. It wasn't an error.

So I think when you talk about safety, quality, broader words, you remove the limitation.

CO-CHAIR DANFORTH: Martha.

MEMBER RADFORD: We are asked here to develop a measurement framework, really, not to develop measures, per se. The framework can also suggest priorities here.

I think that reflecting back on 20

years in the quality performance measurement reporting and improvement field, where do the really good new measures come from? They have come from various ends of things over the years, including national initiatives, say the AHRQ measures, the quality of those you can debate but anyway, they are there, the CMS measures. But really right now I think the really innovative measures are coming from the field. They are coming from the provider community, doctors -- Dr. Singh and David -- sorry, I can't remember your last name -- David and Hardeep have given us a few of those actually really good examples.

So if we are asked to develop a framework, it is really potentially a framework for everybody. And what I am saying is that I think at the national level we need to get these safety measures because then the others can flow, as I explained before, but it also can inspire the other parts of the healthcare system to develop the kinds of measures for the other domains, the efficiency domain, patient-

centeredness, et cetera. There is certainly some work that needs to be done on new ways to solicit patient reaction to the diagnostic process or whatever. There needs to be more work there. What do you do? The whole movement of patient-reported outcomes, which is a whole separate discussion, may not be covered here.

And I just think that this -- I am going to agree with calling it diagnostic quality and safety and just say that, as far as a tactic, from the national level, that we do the most service to the country by defining the kind of big picture safety items earlier on.

CO-CHAIR DANFORTH: Jen.

MEMBER CAMPISANO: I just wanted to agree, one, agree with Helen that I think -- sorry. When we talk about the safety is important and quality is important, to put on my lawyer hat for a minute and not my patient hat, I would just caution against the word error, which I am sure every physician in the room is aware of but I think it is incendiary to some degree and

it invites blame where maybe blame doesn't need to exist.

DR. LUSTIG: Can I jump in for a second? Because I think we are getting really far afield from what we were trying to do here.

We weren't ever going to use the term diagnostic error. The title of this project currently is Improving Diagnostic Accuracy and a lot of people said they didn't like that. And so we were trying to just in general name the project something different. It sounds like we are landing on diagnostic quality and safety, which is fine. We also didn't intend to get here into what we are prioritizing because that is going to come up later in the project. And it does sound like we are already leaning towards safety being a priority.

This was simply meant to be we don't like the title of the current project. And so I don't think we necessarily need to keep going, if people seem to be okay with us calling the project, in general, Improving Diagnostic Quality

and Safety.

CO-CHAIR DANFORTH: So let's let Hardeep and -- I can't see your name tag, I'm sorry -- and Marilyn their comments and those will be the last two comments we can take on this topic. Then I am going to open the line to see if there is public comment.

MEMBER SINGH: So even though we have had lots of good conversations on quality and we believe it is important to think about some of the ways we are going to approach safety and some measurement issues, I do think that the title of the project and the report that comes out ultimately should only have the word diagnostic safety in it.

And there are several reasons, I would say for that. One is another quality report and more quality measures is generally going to be, the physicians say we already have quality measures in healthcare. We have got diabetes and all those other ones, hypertension. Why do we need more?

So the thinking of the front line physicians, who we really want to reach out to because, unless we engage them, nothing is going to change, is not going to be more quality measures of diagnosis.

Second, and we have had this issue with some of the work that we have done with Office of National Coordinator, is we are still having integration problems of sort of health IT related safety with sort of the patient safety because they think our patient -- you know these are just two different things. If we are talking about patient safety, we are only talking about readmissions, and missed diagnosis, and falls, and so on and so forth, and infections, but health IT-related safety issues are totally different. We don't want that; there is a wall.

So let's think who is going to be our audience. Who are going to take these measurement concepts forward? Who are going to be the improvers of what we want to propose?

So the improvers, to me, are people

like Martha's crew, who are going to be looking at sort of risk management, sort of the patient safety officers, managers. And I don't know if some of the routine quality measurement concepts are going to be that appealing to them because that is not what their areas are focused on, the foci of.

CO-CHAIR DANFORTH: Marilyn.

MEMBER HRAVNAK: I guess I just wanted to make the comment that if we are looking at the next step, then, which is once we decide then to come up with measures, I think we get into a little bit of trouble with quality and safety because I think safety is more a dichotomous setup. You know you are either safe or not safe and there is a little bit of gray in there but there is not much of a spectrum as there is between low quality and high quality.

And you could argue on a low quality end that a lot of the unsafe is going to be there, but I'm not certain that the same measures are going to translate across both of those.

CO-CHAIR DANFORTH: So just so I may try and understanding, so what you are suggesting is that we, for the project title we use the broader title of Quality or Quality and Safety.

MEMBER HRAVNAK: I think that the measures, it depends on where we want to start. I think that in terms of the measures, I think that the measures for quality and the measures for safety may not be overlapping. I think safety is going to be the subset and it is going to be much easier to put in a measure for a practice that is safe and unsafe, versus a practice that is of the lower quality or a higher quality, when you get to the measurement tools.

DR. LUSTIG: I actually think this might belabor it more, but maybe this is something people can look at and give us some feedback on. So based on sort of our assumption of starting with a title of Diagnostic Quality, these were some definitions -- these are different variations on each other, based on the academy's report definition that we could be

using for our project. And so it was based on that, one of those three bullets, and then making it clear that it had the six dimensions of quality implicit in that definition. But again --

CO-CHAIR DANFORTH: Maybe we can leave this up through lunch because I think when Andrew talks about the framework, a lot of this is going to come back up. Is it going to be the framework for diagnostic safety and quality, or the framework for diagnostic quality?

I am always sensitive if folks are waiting on the phone to give public comment because I have been one of those folks before. So I am going to ask the operator to open the line to see if there is anyone in the room or on the phone, or Tom.

OPERATOR: Okay, at this time if you would like to make a public comment, please press *1.

And there are no public comments at this time.

CO-CHAIR DANFORTH: Paul?

MEMBER SEQUIST: Hello?

CO-CHAIR DANFORTH: Oh, hi, Tom.

MEMBER SEQUIST: Hey. So, it sounds like you are going to break for lunch. I don't want to hold everyone up. I would just throw my vote in, I guess. I think it would be sort of labeling it not as error but something in the framework of safety and quality, I would favor more.

CO-CHAIR DANFORTH: That's helpful. Thank you.

Paul.

MR. EPNER: So as a member of the public, I have to reflect on the entire morning, but I will be very brief and have three comments. Hopefully, I remember what I was going to say.

So it is Paul Epner, Society to Improve Diagnosis in Medicine.

In the discussion of process failures and label failures, and we talk about process defects, the cognitive issues sometimes manifest

themselves as a process defect. I didn't think to order a test. You can look to see if the test was ordered and, therefore, it may translate into a process defect. But there can be cognitive biases, et cetera, that lead to inappropriate conclusions for which there may not be anything measurable in the process. And I would hope the committee would consider process clinical reasoning as a part of a process that should not be ignored and think about that in the total context.

The notion of completing -- this is the second comment. The notion of completing the task versus the effective completion of the task is something I hope you will all reflect on. I am thinking that Hardeep was talking about the ability to measure the communication of test results to patients, but giving the patient the physician's report with Hs and Ls and asterisks may not mean anything to the patient. Half the time, it doesn't mean anything to the physician.

So I think we need to always think

about when we look in big databases for evidence that something happened, just because it happened, didn't mean that it happened effectively, and that is the difference.

And then my final comment is on the notion of patient engagement and measuring patient engagement. And the notion of patient preferences I think may have been mentioned earlier, but I didn't hear it strongly. So some patients, for reasons of financial need, for reasons of being afraid of not being able to handle the information, knowing they have a disease that is going to cause or knowing they may have a disease that may cause them to miss work and they have got a family depending on them, so the notion in the process of offering patients or evaluating patient preferences and documenting those patient preferences in some way I think potentially has some value in this whole issue of the effective communication of diagnosis.

Thank you.

CO-CHAIR DANFORTH: We will break for lunch. We are 15 minute behind schedule. So I will look to the NQF staff to decide when we should come back.

DR. BERNOT: We can make that time up between Andrew and my section. So I think if we come back at 1:15 or even 1:00. Yes, 1:00 would be fine for us to come back.

CO-CHAIR DANFORTH: So Tom, we are going to drop off. If you want to drop off, we will email you around 1:00 if you could call back in.

MEMBER SEQUIST: Okay, great.

CO-CHAIR DANFORTH: Thank you.

(Whereupon, the above-entitled matter went off the record at 12:32 p.m. and resumed at 1:07 p.m.)

CO-CHAIR DANFORTH: Okay, everyone, I think we are going to get started. So we are going to start actually by having Andrew review the proposed framework for measuring diagnostic quality --

MR. LYZENGA: And safety.

CO-CHAIR DANFORTH: Don't start, Andrew. And then we are going to have John and Andrew review the preliminary results of the environmental scan that they did. And so let's start with the framework.

And would you like committee members to hold questions to the end or as they come up?

MR. LYZENGA: No, you can ask me questions, if you like or hold them to the end. Either way. I think we had actually allocated a good chunk of time at the end of the day for discussion of both the framework and the environmental scan and the measures we found through that.

So if you want, you can hold your questions until that time or just interrupt me if you would like to. I am perfectly happy with that.

So as Christy mentioned, one of our major deliverables for this project is to come up with a framework for measuring diagnostic

quality/diagnostic safety, whatever we end up wanting to call this.

And just to sort of reiterate what the purpose of this framework is, there is a number of purposes -- and this isn't actually a comprehensive list, either -- but among the major purposes for this framework are to provide an organizational scheme for us to identify and categorize diagnosis-related measures, to facilitate the systematic identification of measure gaps related to diagnosis and diagnostic quality, accuracy, safety, and then to facilitate a systematic approach to prioritizing those measures or measurement areas and gaps that we identify, and then finally, to serve as sort of a conceptual tool or guidance for the development of diagnosis-related measures in the future.

So essentially this is intended to serve as a tool for us to do our work moving forward and for others in the field to -- we will use it to sort of frame our recommendations in terms of what we think is important to measure,

how it should be measured, and what the sort of key areas of measurement are with respect to diagnosis and diagnostic quality and safety.

What we are proposing here is very similar to what we saw with the National Academy's framework and to Dr. Singh's framework. What I am proposing here is, in some ways, is sort of a deconstructed version of the Academy's framework pulled out into -- or embedded within, rather, Donabedian's structure-process-outcome model. And I actually got that idea from Hardeep's Safer Dx framework.

The elements of the National Academy's framework and Dr. Singh's framework seem to fall actually fairly nicely into these categories. So it seemed like it made sense, given that this is the traditional way of looking at measures and organizing measures.

Structure measures address aspects or attributes of the work system, the sociotechnical system -- to use Hardeep's terminology -- in which diagnosis occurs. Process measures

addressed those actions or processes or interventions or things that commissions or others are doing to support accurate and timely diagnosis. And then outcomes, obviously, are what comes out of that process. And I will talk a little bit more about how we have broken that category down.

So again, the structure domain. Pretty much, if you look at the diagram here, the Academy's model addresses those -- that circle there of the work system elements. You can see there are the diagnostic team members, tasks, technologies and tools, organization, physical environment, and external environment.

If you go to the next slide, those are the same elements rephrased slightly. In the first, these are the subdomains of the structure domain and we would hope that any structural measures that -- concepts that we come up with or measures that we identify, would be able to sort of be categorized into one of these dimensions or domains.

And then the process domain encompasses -- again within that circle of the Academy's model -- from where the patient engages with this health system. So including that part of patient engagement, that will include things like access to care and other aspects of that, that cycle of the diagnostic process, and then communication of the diagnosis. All of those elements are part of the process domain.

We have sort of tweaked the names of these a little bit or added on to them with some feedback from Dr. Graber as we were developing this. The first is still patient engagement.

The three elements that relate to that sort of cycle of the diagnostic process we are calling -- were called information gathering, information interpretation, and information integration in the Academy's report. We are sort of renaming them a little bit or we are adding a little more detail to sort of reflect the concept of diagnosis being an iterative and -- cycle, and something that is ongoing.

But you need to sort of categorize a given measure somewhere, so we have the first being information gathering, that being those steps that you take to conduct a diagnostic evaluation, the next step being information interpretation, where you take that information and generate an initial hypothesis, and then information integration, being that hypothesis confirmation and revision over time. And that I think will -- we intend to incorporate things like communication between providers and other ongoing activities as part of that diagnostic process.

The next element is -- sorry, go back -- is communication of the diagnosis to the patient. Again, we think that is a really important aspect of both the National Academy's framework and their definition of error. We want to incorporate that into our own definition and work.

We also added on another element that is acknowledged in the Academy's framework but

not, I guess, in some versions of the model. We added on quality improvement and learning activities. We wanted to make sure that was incorporated here, any processes that organizations or clinicians are taking to learn from what is happening and improve the quality as its own dimension that could be measured through process measures.

The outcome domain, again, at the very end there, if you want to skip to the next slide.

How we've broken this -- what we are including here in outcomes is both what are typically sort of construed as outcomes with that being patient outcomes, clinical outcomes. But also we are including intermediate outcomes which, in this case, is what we are labeling diagnostic errors, for example, a missed diagnosis or a late diagnosis. Those sorts of diagnostic outcomes, I guess you might call them, we are bucketing them into the -- or calling them, for our purposes, intermediate outcomes. And we can discuss whether that is appropriate.

Another next dimension is those patient -- actual patient and clinical outcomes. This is kind of, I imagine, maybe a tough area to address, but we would like to make some progress toward identifying some clinical outcomes, actual patient outcomes that could be associated with diagnostic performance and the quality of diagnostic care.

Patient experience being the next, obviously, really important aspect of this as well.

And then system outcomes being things like cost, resource use, efficiency, and -- as acknowledged again in the Academy's report -- things like the patient and consumer confidence in the system which is, again, maybe a hard thing to reflect in measurement, but is something that we wanted to account for here.

So that is the basic outlines of that. We will get more into this as we -- we sent a document to you that laid out those domains and also had some measure concepts associated with it

to sort of illustrate how measures might be associated with each of those domains. Those were taken from Hardeep's paper and some concepts that we were suggested by our co-chair, Mark Graber. And as we get into our environmental scan results -- we actually didn't include those concepts in our analysis there but we will -- as we move forward -- incorporate both those concepts and any concepts that we come up with as a group in our exercises in terms of identifying measures and potential measures and where we can identify gaps.

So maybe I will just take questions here. Yes.

CO-CHAIR DANFORTH: Martha.

MEMBER RADFORD: Could you go back to the thing where you had a circle around a bunch of it?

Yes. So I was a little concerned because you are leaving off treatment there. Because response to treatment is a very key diagnostic tool and I just would incorporate that

in there. And perhaps there is monitoring --

MR. LYZENGA: Do you think that should --

MEMBER RADFORD: -- the diagnosis on an ongoing basis, refining it, et cetera. That is an important piece.

MR. LYZENGA: Sure. I think we intended that to be included in the information integration step. If you go to the next slide, it includes, again, sort of hypothesis confirmation and revision, that being sort of that continuous process of learning from your treatment approaches.

But if you and others think that it may need to be called out as a separate domain, we could consider that.

MEMBER RADFORD: I'm not sure it is a separate domain but it's just to mention somewhere that response to treatment is part of the diagnostic process.

MR. LYZENGA: Okay.

CO-CHAIR DANFORTH: Mike.

MEMBER DUNNE: Maybe I had better wait for Act 2. This is just general right now, right? This isn't a plan for what you want to do, what you want to measure, how you are --

MR. LYZENGA: It is supposed to, I guess, serve as, again, sort of a framework for filling in that detail in the future. We would hope that this would sort of identify the major categories in which measurement might occur. And there could be all kinds of measures that would fit into this but we would hope that any measure that comes up could be, in some way, categorized into one of these domains just for the purpose of sort of organizing our thinking and identifying again where there are gaps in measurement or where we want to prioritize some particular area.

MEMBER DUNNE: Because I keep on coming to the conclusion that all of this is going to be very, very different for each clinical entity.

MR. LYZENGA: Oh, absolutely. Absolutely.

MEMBER DUNNE: I mean you are almost going to have to define separate parameters for each -- you know, for sepsis, for stroke, for this, for that.

MR. LYZENGA: Yes, and so there will be -- yes, there may be many measures that fit into this. And that is -- yes, you wouldn't -- I expect, at least, you would have a measure that is information gathering or diagnostic evaluation. It would be something much more specific, specified out, likely for a particular condition or set of conditions that would then reflect the quality of the information gathering process or the diagnostic evaluation, as an example. So these are, yes, just very broad categories that are intended to serve as an organizing framework for much more specific measures that would kind of fit into that.

MEMBER DUNNE: So do you see, in the long-term, that these will be used to develop diagnostic pathways for all of these clinical entities?

MR. LYZENGA: I don't think so. I mean -- what do you mean by pathways?

MEMBER DUNNE: Well, let's take sepsis, for example.

MR. LYZENGA: So like the process measure around sepsis that you have to sort of take certain steps and things?

MEMBER DUNNE: Yes, what is the differential diagnosis? Once you learn the diagnosis, what is the appropriate treatment? You know, what type of testing is involved? How is a septic patient engaged?

MR. LYZENGA: Yes, so I imagine -- I mean, you could, potentially -- I don't think this is what we intended it necessarily to be but we -- that is an interesting idea that you could sort of apply this entire framework to, say, a clinical area or condition like sepsis and, across each of these domains, come up with structure elements that reflect the quality of diagnosis for sepsis or the sort of conditions for correct and appropriate diagnosis of sepsis

and then process steps in each of these domains.

MEMBER DUNNE: Right, because you can measure that.

MR. LYZENGA: Right. Well, we hope so, and that's what we --

CO-CHAIR DANFORTH: David, David, Hardeep, and then Helen.

MEMBER NEWMAN-TOKER: So, just to follow up on Mike Dunne's point before I get to my other question. I think this is an important issue, is that a lot of the measurement in this space is likely going to be problem-specific because we know that the problem with diagnosis is highly problem-specific. And there are some generalizable elements of it, but we have to be mindful of that as we construct the framework.

And in particular, I think we have to be mindful of the fact that when we are talking about measuring outcomes or diagnosis label failures and correct diagnoses, you are focused on diseases. But when you actually start talking about correct process and the diagnostic process

being high quality or low quality, or associated with error or not, you actually have to focus on clinical-presenting problems or symptoms.

So it is not what did you do in sepsis patients. It is what did you do in people in whom sepsis might be suspected who had a fever or low blood pressure or this or that or the next thing, not -- because you don't know the diagnosis up front. So you have to look at diagnostic process performance in the context of a high-risk situation or clinical-presenting complaint or problem, not the reverse. It is only when you are talking about the outcomes that you can sort of take that disease frame.

Can we go back to the outcomes one? My question here, when you started talking about -- sorry, go up one more, the subdomains. So you said that diagnostic measures sort of mostly fall into the intermediate outcomes. I would like to just maybe hear a little bit more discussion about that and get some reflections of folks -- including Helen and our patient representatives -

- to have a sense for where that shakes out.

So, like you can think of an intermediate outcome measure that says like was this an assay for whether you had this gene defect in a colon cancer right or wrong. Was it a false positive/false negative? I think everybody would kind of agree that that might be thought of as kind of an intermediate outcome. But is it an intermediate outcome when we get to the point where you are given a diagnosis and then it affects your treatment? Maybe you want to reflect on your experience there?

MEMBER CAMPISANO: Sure. Where do you want me to start?

MEMBER NEWMAN-TOKER: Well so, to you, was that an intermediate outcome or was that a patient clinical outcome? In other words, when does a diagnostic failure of some kind fall out from being an intermediate to being a patient outcome? Is it when the patient is harmed? Does that harm include the psychological harm of being given the wrong diagnosis? Where is that

transition point from intermediate to patient-centered meaningful outcome happen?

MEMBER CAMPISANO: I can sort of go over the discussion that we were having in the break a little bit to review my background.

In August of 2011, I was diagnosed with what we thought was metastatic breast cancer. And I had a five-month-old son at the time. Discounting the fact that my OB/GYN had thought it was mastitis and dismissed me several times, by the time I was diagnosed and then treated for four and a half years for metastatic cancer, I thought my oncologist knew what he was doing.

And I still think that he is an excellent doctor, but earlier this year it came out that I probably had stage II breast cancer and, at the same time, sarcoidosis, which can mimic cancer on scans, apparently. And that came out because I had a lung biopsy this spring, finally. And so I don't know where I would consider that error to have occurred or, you know

-- and obviously my outcome is better than I thought it was going to be. So I don't know where I would fall into the framework.

MEMBER NEWMAN-TOKER: So but Jen, I guess what I was trying to ask was -- obviously, it is a horrible story. At what point did it cease -- so, an intermediate outcome is sort of some kind of accounting tabulation or along the way that we didn't get it right but it may or may not have had any downstream impact on a patient outcome.

And I guess what I am trying to figure out is how soon in the process of that problem did you experience the real-world harm of your situation?

Was it at the point where you started getting side effects from chemo for a disease that you didn't have, or was it the moment they dropped the bombshell of the C diagnosis and said that you were going to die? Or when was the -- when did we cross over into meaningful harm that we want to be preventing?

MEMBER CAMPISANO: I would say that the majority of the harm that I experienced was probably psychological. And that came within the first week, when they went from saying yes, it is breast cancer to we did a PET scan we see spots in your spleen and your lungs and your chest wall and your abdomen, and you are probably going to die in the next couple of years.

MEMBER NEWMAN-TOKER: So I think the issue there that this brings up is whether getting that diagnostic label, in and of itself, may be a patient or clinical outcome, never mind what happens to their medical condition. And I think we have to be careful about how we define what is an intermediate outcome versus a real outcome for patients.

MR. LYZENGA: And I should say that we -- I think Mark brought this up in one of our recent conversations, that we should be considering those psychological effects and I think we had intended those to be in the patient and clinical outcome area, not in intermediate

outcomes. That would -- I think the intent was for that to apply to those, I guess, label failures and that any effect of that label failure on the patient -- whether that is mental or psychological, physical -- would be considered a patient outcome or clinical outcome. And I don't know if that makes sense to everybody else. This is exactly what we want to talk about.

MEMBER HASKELL: Would a better distinction be near miss?

MR. LYZENGA: Yes, you could consider those near misses, the intermediate outcomes where there was a label failure.

I mean, well, it would -- even if it did result in harm, I think then you would want to account for it I guess twice, in some sense. You had the label failure -- the diagnostic error, if you want to call it that, or the missed diagnosis -- and then the patient harm that occurred if it was not a near miss.

MEMBER MCDONALD: Actually I think what you are saying is you could have a near miss

-- whether we count that as an intermediate outcome or not could be debated -- but that there still could be a patient outcome from a near miss.

MR. LYZENGA: Right.

MEMBER MCDONALD: I mean somebody doesn't actually have to -- if they knew that there was a near miss.

MR. LYZENGA: Right. If that caused some -- yeah.

MEMBER MCDONALD: That could have an effect, yes.

CO-CHAIR DANFORTH: David.

MEMBER SEIDENWURM: Two points. One is that I would like to add a little bit to what Martha said. And I think not just treatment as a diagnostic tool but also just observation, just doing nothing and waiting as a diagnostic tool. And I think it would be very important to include that explicitly in our thinking.

The other thing is --

CO-CHAIR GRABER: Can you say that

again, David, a little louder?

MEMBER SEIDENWURM: Sure. I'm sorry.

So what I was wanting to say was that I wanted to add a little bit to what Martha said about counting treatment explicitly in our evaluative process. And sometimes doing nothing is the best diagnostic test and just watching because many processes are self-limited, many symptoms are self-limited or many disorders are self-resolving. So they don't all need to be investigated.

And then the other point I wanted to ask about is how do we conceptualize the idea of a negative diagnostic test when, in fact, the patient doesn't have anything but we have given them the idea or they have, themselves, gotten the idea that they might. And so we see this all the time in headache, back pain -- perfect, another one. I mean there are a hundred great examples.

And so I feel sometimes like we are the arsonists who set the fire and then get the

medal from the mayor because we heroically put it out.

And so I don't know exactly how to think of this in our taxonomy here.

MR. LYZENGA: We could consider it a -- well, intermediate -- I don't know, an intermediate outcome, in some sense, a system outcome maybe.

MEMBER NEWMAN-TOKER: Well I think we're straying into the territory of overdiagnosis and overtreatment and so -- with overdiagnosis being defined as sort of a condition that it actually is but you don't -- the patient shouldn't care about because it is not really going to make a difference.

The issue here -- in terms of the harm to the patient -- you know, the same way we just said, if an overdiagnosis -- like you say that somebody has thyroid cancer and it was something that was immaterial -- you may have gotten the correct diagnosis label in some sense but you have made a diagnostic process failure in another

sense because you found something -- if you pursued it in an overly aggressive way -- that you weren't supposed to find.

I think it is when it becomes psychologically harmful or harmful to the patient in terms of overtreatment that it then sort of falls off into this clinical outcome bin. So I do think that there is a way to kind of make that transition from diagnostic error or overdiagnosis or whatever to harm from that.

CO-CHAIR DANFORTH: Hardeep.

MEMBER SINGH: So I want to pick up from what Mike said and make another related point.

So you asked for disease-specific measures. Is this framework going to inform every -- you know, measures for every single disease? I don't know if that necessarily needs to be the case right now. Maybe another 50 years or 100 maybe that would be more relevant because I think the problem is there is too many things going wrong at the same time.

So in one of our studies, in 190 diagnostic errors in the primary care setting, there were 67 different conditions involved. None of them was more than six percent of the entire subset and then some were just like two, one and two -- you know, like right at the end. So, there are so many things to fix.

So, you could take a disease-specific approach if there is a really high-risk condition like sepsis and use this type of a conceptual framework to make sure you are covering ground and then maybe focus on that as an outcome measure.

But as far as the process measures and the system structure measures that we are sort of thinking about, I think they will be fairly generic, that you don't need a disease-specific approach for every single structural, every single process measure. I think disease-specific approach, for instance, in one of our papers that you took was for cancer. So for colorectal cancer, high-risk condition, have a measure as an

outcome.

So does that sort of help you sort of --

MEMBER DUNNE: Sure but when I think about things like timeliness, you take sepsis and then you take chronic lymphocytic leukemia and the timeliness of diagnoses is a world apart or even --

MEMBER SINGH: Oh, yes, absolutely.

MEMBER DUNNE: -- tuberculosis.

MEMBER SINGH: Yes, well, lung cancer and colorectal.

MEMBER DUNNE: It has got to be within the framework of when therapy is going to provide a benefit and when you have lost that.

MEMBER SINGH: Very true. And that is why I think outcome measure. So an outcome measure for like maybe 60 or 90 days for colorectal cancer diagnosis but shorter time for lung cancer diagnosis would be more relevant. So that was sort of my reflection.

The other thing I was just going to

mention quickly is there is a very important measure that is on the previous -- the one about the quality improvement and learning activities that we shouldn't sort of lose sight of. I'm not sure which slide it was on -- but that is really important because I am going to ask this question to you, Jen, is what did your system -- your health system and your providers -- learn from your experience and what did they change?

MEMBER CAMPISANO: I hope that they have changed how they approach new patients who present with seemingly metastases that they investigate further. You know, I think it is not 100 percent accurate, but there is a blood test for sarcoidosis, for example. You don't have to necessarily biopsy every lesion.

But I am not entirely sure. That is a good question. I haven't asked.

MEMBER SINGH: And I would say most missed diagnoses we learn nothing and we move on, in most cases like this we move on.

So I think really that quality

improvement in learning activities needs to be a very strong component of trying to sort of think about measurement in that area to make sure that physicians, nurses, the healthcare system, are making progress to what I think is actionable measurement.

MR. LYZENGA: And just --

DR. BURSTIN: Just one quick thought on the process outcome question.

MR. LYZENGA: Yes, go ahead.

DR. BURSTIN: It is a very interesting question, Hardeep. Just going along with the comment about sepsis, there are, in fact, three-hour bundles for sepsis so you don't miss anything, that you don't, in fact, miss the early identification and treatment.

So I'm not sure there are always going to be outcomes but I understand where you are going. But I don't think we want to be so exclusive to say you would then not have, for example, appropriate bundles around diagnosis to make sure you don't miss anything for some of

those three very high-risk areas you mentioned earlier.

MEMBER DUNNE: But getting back to what David said -- and most of the time you don't know what you have to start with, but there is going to come a time where you are going to want to work backward, when you have the final diagnosis and you want to retrospectively evaluate whether it was done appropriately from the start. And in that case, now you have got disease-specific guidelines.

MR. LYZENGA: And in fact in our environmental scan -- you will see this in a moment -- is mostly that type of measure to the extent that we found things like intermediate outcomes. And most of the process ones as well are focused around specific conditions or diseases.

One of the measures we found was persistent indicators of dementia without a diagnosis that -- over a period of time in a nursing home a patient has been exhibiting

behaviors and characteristics that would suggest -- and even undergone assessments that should have prompted a diagnosis of dementia, but that diagnosis was never made is the sort of measures that we are finding and that we, I think, would expect.

I think you are unlikely to get -- and something I kind of wanted to say about the sort of the definition of diagnostic error, too, in some ways is -- you know, I think maybe we can make some recommendations around what we think the most important way to think about diagnostic error is. But in terms of specific measures, you are unlikely to have like a measure of diagnostic error. It is more likely to be some particular condition or set of circumstances that reflects a diagnostic error in one of those ways that is very sort of detailed and specified.

So in some ways, the sort of question of how you define diagnostic error is kind of moot for those considerations, although I do think as part of our recommendations broadly, we

probably want to have some thoughts around how we think diagnostic errors should be conceptualized. Even though once you get to the specific measure level, it kind of becomes irrelevant in some ways.

MEMBER NEWMAN-TOKER: Can I just follow up on that? I know Helen has been waiting, but just on that point, that issue of how disease-specific we need to get, I do think that there is a middle ground there. Which -- for instance, the example you gave about carrying an undiagnosed condition for x number of visits, encounters or whatever, is kind of a generalizable idea. The same way that what I showed you about the stroke stuff is a generalizable idea. You can do that symptom-disease para-framework with chest pain and heart attack. And you can do it with dyspnea and pulmonary embolus, and fever, and sepsis, so on and so forth.

So when we think about developing a measurement framework -- rather than individual

measures -- maybe what we want to do is get granular enough that maybe it doesn't apply to everything, but that it applies to sort of a class of problems, where people can start to plug in -- fill in the individual things. Maybe for dementia it is a slightly different set of criteria than it is for missed cancer or something else. But like the undiagnosed condition for three or more times is, in and of itself, maybe kind of the framework level that we want to get to.

CO-CHAIR DANFORTH: Helen.

MEMBER HASKELL: No, I just wanted to sort of put my stake in the ground again about making this patient-centered. So instead of explaining to the patient, you have discussions with the patient. And the primacy really of the patient and experience outcomes, you seemed a little dubious about that. I think that is really central.

And then the quality improvement learning activities, again, patient involvement

really changes those and makes them much better and they also are areas that are of critical importance to patients to know that there has been improvement.

CO-CHAIR DANFORTH: Prashant.

MEMBER MAHAJAN: So I have to confess I am getting more confused. And you know you can all just put it down to my ER background but the way I look at it is we are coming up with a conceptual framework that provides the best way to explain the patient's condition. So that is the definition that we are going by. It is to explain the health problem in a timely and accurate manner and communicate that to the patient.

So I am with Hardeep, in that sense, that if we go down the path of having individual measures from the provider perspective, I think it just adds to a layer of confusion.

For instance, in the sepsis issue, there a bunch of measures already out there. So is accuracy and diagnosis of sepsis and added

measure in how we are going to respond to that and how people are going to take this on. It has become a little bit difficult to operationalize.

So the way I was thinking about it is we could consider measures what can be done in the three domains of structure, process, and outcome that will enhance the ability to provide that accurate explanation. So, rather than going towards individual conditions -- I am just throwing this out -- if it could just involve x number of providers trained for x number of years in certain conditions, something like that, versus timely access to patients to clinics. In that sense, it sort of takes away -- because whatever that will allow the diagnosis to be made in a timely manner, rather than going down individual path.

Because in sepsis, my only argument toward sepsis is there are a bunch of bundles out there but it is a very heterogeneous condition and most of the bundles have not been shown to be effective because it does not necessarily treat

the original condition.

So what now has happened is certain states have now institutionalized those bundles but yet there is no clinical acceptance and you have this dichotomy of some measure of quality but not being accepted.

I was just throwing that out. Maybe it is just my confusion, but I am --

MR. LYZENGA: Yes, and I think the -- I wasn't saying necessarily that we ought to be focusing on specific conditions. That was what we found in the environmental scan so far. I think we should certainly try to push towards sort of broader measures if we can make those sorts of recommendations of things that capture more cross-cutting issues around diagnosis and diagnostic quality.

I think in particular you mentioned a couple of ideas and those seemed like structural measures to me, and my inclination would be to say most of the structural measures are going to be much more cross-cutting and about sort of

creating the conditions in which diagnosis can be made successfully and that is very unlikely to be conditioned or disease-specific, at least in most cases.

And you know in terms of the processes, those may get more specific or not. These are exactly the types of things I think we want to discuss over the course of this project. And right now we are kind of creating this broad framework for conceptualizing diagnosis and what elements we might be able to measure. But as we move on, I think we ought to get more granular as we can and make recommendations around what we think measurement should look like. And I think this is exactly the sorts of conversations we want to be having moving forward.

CO-CHAIR DANFORTH: David.

DR. HUNT: Just to -- this is probably a reiteration of how this product will be consumed as we identify and create the cubbies, if you will, that the individual measures will fit in.

From a policy framework, organizations, groups at HHS may then look and say: well, you know we have a whole bunch of measures in this domain, but it seems as though we need to have someone develop measures in say patient engagement or hypothesis confirmation because it seems as though we are missing a lot when people do studies of root cause analysis. And that will help us understand what areas we may want to improve our measure portfolio, if you would.

In which areas do we have enough measures? And if we have a somewhat balanced set of measures in a particular domain, that might lend you to say well this might work. These group of six might work well for a composite measure.

So this is, in large part, filling out the little cubbyhole places where then we can work on areas of improvement.

CO-CHAIR DANFORTH: I am going to do two more comments from the Davids and then I am

actually going to turn it over to John to review the environmental scan because I think that will actually help some of us better wrap our head around the concept of the framework when you see some of the example of measures currently exist and where they would fit in the framework. You will see that it actually ranges through a bunch of settings in clinical areas and conditions.

So, the last two, then John. So, David Newman-Toker and then --

MEMBER NEWMAN-TOKER: Just quickly, in terms of Prashant's comment. I think it brings up a really important issue, which is that we should be careful not to stray into the notion that says we know which process defects that are out there are kind of causally linked to the specific diagnostic errors, and we should focus our attention on all of these specific process defects that are clearly responsible and, if we fix them, they would reduce the outcome of interest.

I think it is tempting to do that but

I actually think we have no idea pretty much for anything because, to my knowledge, there has not been a single intervention that has shown that you could actually decrease patient harms -- which is sort of ultimately where we are getting to -- from an intervention focused on some specific aspect of the diagnostic process to get to the point of decreased patient harm.

And Hardeep can correct me if I am wrong, but I am not aware of any such studies. And so even if there are one or two out there, they are not across any meaningful set of domains and we know that the process defects that are responsible are going to differ across problems.

So I think we need to measure process and we know that that is the place that we can change in order to influence outcomes, but we should be a little agnostic about which process problems are the causal ones. We should be very careful about drawing too many inferences about which ones matter the most.

MEMBER SEIDENWURM: Thank you. And

this is not to be considered the last word in any way. I just wanted to make sure that we considered the problems caused by the EHR technology itself in the diagnostic process and how some of the human factors, perhaps -- or lack of human factors and their design -- contribute because I think that a big problem that we see is that there is a lot of information that is hiding in plain sight and there is a lot of debris in the medical record that obscures the ore that you are trying to mine.

And the famous example of the Ebola casein Dallas was quite likely exacerbated or even caused by inadequate either design or implementation of the EHR.

MEMBER SINGH: I actually wrote a whole series on this case. So it is actually not just was the EHR. The physician had clear access to the nurse's note and could actually read -- if he had gone on a different screen -- that the patient had travel history to West Africa. So, I just wanted to --

MEMBER SEIDENWURM: This is the last word. I rest my case.

CO-CHAIR DANFORTH: No but I mean just to go back to David's point, there are I think some sort of discrete measures that aren't perfect or ideal measures that do sort of get at what you said. I mean I think CMS tried to implement or maybe even for a limited time implemented a DVT measure that looked patients that were at risk for DVTs didn't received the prophylaxis and then got a DVT.

So I think within the diagnostic process, where the risk is identified and the treatment clear and then not given, and then the bad outcome happens, there are some small discrete examples, potentially, to learn from.

MEMBER NEWMAN-TOKER: So an intervention was introduced that increased that quality measure and then it decreased the harms from DVT like PE in patients?

CO-CHAIR DANFORTH: Yes.

MEMBER NEWMAN-TOKER: Can you send me

that?

CO-CHAIR DANFORTH: I will. And again, so we looked at this measure carefully at Leapfrog because it was a great example of you have this defined population that is at risk and you have a defined care protocol for prophylaxis, and then you just didn't give it them and they got the DVT.

You know what I mean? It sort of --

MEMBER NEWMAN-TOKER: But that is a treatment thing. I mean is that a diagnostic thing?

CO-CHAIR DANFORTH: But they were identified. So they were diagnosed as at-risk for a DVT, and then they didn't get the correct treatment for that risk that they were diagnosed with, and then they got the DVT.

MEMBER NEWMAN-TOKER: Well that is actually its own little interesting can of worms that we should discuss. I mean is that what we mean by diagnosis? I mean --

MEMBER MCDONALD: Diagnosing the risk

and following up with a treatment for that risk, as a pair, right? Because that is what you are saying.

CO-CHAIR DANFORTH: Right.

MEMBER MCDONALD: And it is the pair. The pair didn't happen.

CO-CHAIR DANFORTH: Right and it is a paired measure. I mean that was what in the measure.

MR. LYZENGA: More, I think you are asking if just the assessment of risk, are we considering that diagnosis.

MEMBER NEWMAN-TOKER: So there is a whole domain of a prognosis and risk stratification. There is like an entire scientific field around this sort of idea. And this is a murky area because it is somewhere in-between screening and symptomatic diagnosis.

And I think we can all agree that when a patient comes in complaining of a headache or back pain or whatever and we don't get the diagnosis right, that that is diagnosis, but it

is murkier and murkier the more you get into the issue of asymptomatic screening and asymptomatic risk assessment and so on and so forth. That is potentially scope creep that could be huge for the framework, and we should at least know where we stand on that.

CO-CHAIR DANFORTH: So I guess the reason I brought it up is because of the conversations we have throughout the day about sepsis. I mean some of that is early identification of the sepsis patient, not responding appropriately when the patient presents with those symptoms of sepsis.

MEMBER NEWMAN-TOKER: Well it is slightly different. So there you are talking about recognition of sepsis signs. That is really symptomatic diagnosis.

What you were saying is this is a patient who has no symptoms of DVT but has the following high-risk setting or behavior. They are in the setting of the hospital. They are post-operative from an orthopedic or

neurosurgical procedure. They are such and such, and such. And if that is part of diagnosis, then I am not saying that it shouldn't be, but I am saying that if it is, this is an even bigger and more thorny complicated problem than it already was.

DR. HUNT: I think it is as big and thorny as you might suggest because in the case of the DVT, if I do a low pelvic surgery, I am automatically putting that patient at risk for DVT. I know that. So part and parcel of doing that procedure is that I should take the steps to confirm or at least make a diagnosis or exclude it because I know that I put that patient at risk.

MEMBER NEWMAN-TOKER: But it is precisely the fact that you know, by having done the lower pelvic surgery, that to my mind makes that a treatment error issue, at treatment safety issue, rather than a diagnostic one.

The main thing that makes diagnostic errors different from treatment, so those that

actually consider them different because are some people who think it is all just sort of one big thing, but to the extent that people compartmentalize between the two, the diagnostic piece involves that uncertainty in not knowing, as opposed to --

DR. HUNT: Right. Not all my patients get a DVT.

MEMBER NEWMAN-TOKER: No, no, but that the procedures that you put in place after a procedure that are based on whatever risk factors, they are based on age, or they are based on the type of surgery, or they are based on the immobility, or this or that, those are all entirely predictable events that are -- there is just a probabilistic association with whether or not your patient will get a DVT. But your path, as a clinician, is actually quite clear.

The diagnostic trouble is that your path is not clear. And I do think that this is a murky and important issue for us to settle before we do this for a year and then realize that we

weren't clear on whether risk stratification was part or not part of the mission.

CO-CHAIR DANFORTH: So let's do the environmental scan and see if other examples of measures that address this question come up.

DR. BERNOT: Okay well really I don't want to cut off -- the discussion is really good.

I do want to start. I think maybe it would be helpful to take just a half a step back and get to what David said because David and I are right on the same page except cubbyhole is way better than what I came up with.

And when we are looking at this from the NQF and how do we take this huge topic and really try to get it down into something at the end of a couple of meetings is valuable. There was a few things we were thinking of. One of them is just we are going to have all of this data or these processes and these frameworks presented so we have some sort of data dump. And I think we tried to do that this morning. And then come up with what these cubbyholes might

look like. And that was what Andrew was going over. So those may be cubbyholes.

And I think one of the things that is important is do those make sense. Are you able to put your measures into those cubbyholes or not, regardless of whatever measure it is you come up with? So that is one thing.

The second thing is: what do measures or concepts look like in those cubbyholes? And what I am going to do here is go over and say: when we look at what is out there, where do they fit in the cubbyholes, to get to what David's point was? We might be overloaded. If we are overloaded in places, I am sure we could all guess where we are overloaded.

And then the last thing, whether it is tomorrow or throughout the course of this process is starting to get some priorities. So now we have the cubbyholes. We know what could fit in the cubbyholes and then we can prioritize them and hopefully, that is a framework or foundation for future measure development. And so it is not

to say all this discussion is not great because this discussion is fantastic. We are already getting to the priority point, I think, to make sure we are focusing on the patient and make sure we are having a method that connects diseases.

So that is what I just wanted to take that half step back so that when I am going through the environmental scan I can show you how some of the measures that are out there already hang on this and then we can say is this a sufficient framework for measure development.

So I hope that makes sense.

DR. BURSTIN: Just one thing to add. Just keep in mind also a really important piece of this that David mentioned. There is a lot of Davids here.

MEMBER NEWMAN-TOKER: This is like the David committee here.

DR. BURSTIN: I am just going to call him Hunt. David Hunt said earlier was we also the reason to have a framework like this is to be able to say this is a new area of measurement.

We don't have a lot of measures. We are going to try to fit some into some of the cubbies as appropriate. But really importantly, where are the empty cubbies? Where are the areas you guys are prioritizing as being some of the most important areas, like let's say, for example, the discussions with patients that we have just talked a lot about so far today. That is likely going to be a pretty empty cubby and maybe one that then rises to the top, in terms of saying if you are going to develop measures in this space, make sure you focus on those.

DR. BERNOT: Empty cubbyholes.

All right. So what I will do is just go over the environmental scan. There is two parts we did on this. The first one was a literature search and we did ask research questions to try to look at what all literature is out there. The second part is looking at the actual measures that existed that we thought. And I am sure we can go around and say you missed a measure or if you have this. I don't know if

you categorized this exactly but I do think, though, the order of magnitude will be the same and we certainly want to get it accurate so I am not trying to say we want to be imprecise but I don't think it is going to change the order of magnitude.

So when we look at the overview, what were we really trying to do? Again, these still use the diagnostic accuracy wording, because that is what we had in the original proposal but say diagnostic quality and safety, just for the sake of this presentation. But what metrics exist? So what measures are out there? What are the approaches? What has been written in this space? What already exists? So that is really what we tried to do here. Can you go to the next slide?

So the literature review I think is an important end component to this. And it was very important to us, at least internally, to be reviewing this to make sure we were identifying the areas. Not surprisingly, some of the most important literature is from people in this room

and we have heard about this literature already this morning. So I think it is an important part of the end product but it is not something just in terms of keeping the scope reasonable for this two-day meeting, we are not going to focus as much on the literature, although I think experience with literature should influence how we go forward. And we really tried to do that.

But just to let you know, these are the sources we looked at. We tried to be comprehensive and we went through the PubMed, the Grey Literature, measure inventories, again. Can you go to the next slide?

So I think this is probably, since we are really trying to get to a measure development framework, where we wanted to spend our time, at least, and so we went through all of these explicitly and looked at some key words. There is other places we looked at measures or found or they were referred to us but these are the ones that we looked at very explicitly and very systematically, I should say. Can you go to the

next slide?

And these are the key words. We talked about this at the web call, too. So I don't want to belabor it. Some great feedback came back from committee members after the web call and said hey, these ones in blue, can you add some other terms and go back and take a look and see what that does to the inventory that you can find. So, we did that. We added all those key words.

And this is a living list. As we come up with more things, we are happy to continue to refine the overview.

A couple caveats on this. Just as we looked at things, and when I get the sheet in front of you and we go over what we hung in these cubbyholes, there was a couple assumptions we made that I do think should be brought up for discussion after we are done with this. One of them is we did not include asymptomatic screening. We had to make a cut at some point. Is that a diagnosis of colon cancer for a person

who just came in for their -- they are 50 years old. So, I am just putting that out there that we did not include these. These can be up for discussion.

A couple of other caveats of things we did not include were strict overuse or appropriate use measures, where it is simply looking at a utilization, not necessarily the end diagnostic tool. So, we made those decisions, and they can be reversed.

What we did include were comorbid conditions. So, you have a diabetic patient and did you diagnose their peripheral neuropathy or did you screen for their peripheral? We did include that. We didn't believe that was as much of an asymptomatic screening as a disease progression. And we are trying to get to that best diagnosis of the disease. So, we did include those.

And then any staging of a disease. So different staging of cancers, especially that might have had a different treatment, we tried to

include those measures. Go to the next slide.

So with that said, there is not a lot of measures out here anyhow. And what you see here is the summary data for the form that you have, the colored sheet you have in front of you.

So just looking, that is actually -- I'm sorry there is a typo on there. It is structure 1, process 45. We will fix that. Sorry about that. And outcome -- yes, very impressive. We are forward thinking.

So process 45 and outcome is 8 and that is of the 54 measures that we found and felt that they fit into the key word search that we did.

Any questions about that part, so far, like just how we got to this part?

CO-CHAIR GRABER: Would this have included Joint Commission measures?

DR. BERNOT: Yes, any measure that we could, yes, that we could come across.

And again, I am certain that there are still other measures out there that we did not

include. And there is a couple other little tiny things we didn't include.

For example, one place might have a COPD measure that looks at spirometry diagnosis over 40. One of them has it over 18. We included that only once in there because I think from a conceptual point of view it is not two different measures. I know it is from the actual measurement science but the concept was somebody who was looking at spirometry and COPD for the point of this.

MEMBER MAHAJAN: So can I ask this question? Were these measures that you found largely disease-specific or rather condition-specific, or were they more cross-cutting?

DR. BERNOT: They were largely disease-specific. And actually I am going to hand it off to -- well, after I finish these, I am going to hand it off to Andrew, who actually put together just some qualitative analysis on the measures, these 54 measures -- again, just as a way of looking at the measures and getting a

feel for what is out there right now.

And we could cut it a million different ways and probably still be correct, but we are just trying to come up with something that makes sense.

Go ahead and go to the next slide. So now I am going to relate the measures back to what Andrew went over. So, if you thought about he had the three -- he had structure, process, and outcomes. And then with the structure, he listed some subdomains, which are the ones across the top of the first page and what you see out there. Structure is a little bit strange, since there was a total of one measure that we came across and you can see which of the buckets that it fell into. So we are short on structure in general and there is only one of the categories. So, that is not necessarily even needing a graph.

But go to the next slide. I think it will make a little more sense when we get into the process measures.

So no surprise. Where is our patient

engagement? We are already seeing may be a hole in what is out there. Patient engagement has one but when we get into the information stuff, the stuff that we, as at least presently as clinicians, know what to do. We know how to take stuff. And did we run it? Did we not run it? Did a process occur or not occur? That is where you are starting to see the measures.

Now this does not equal 54 because we believe that some things can fit into multiple categories, just to let you know that.

And again, so we thought for structure there were certain aspects, certain cubbyholes that might make sense that are different when it comes to process. Again, all these tried to come from Hardeep's framework, at the time IOM framework. So that is where we came up with these.

But all of this is up for debate. You may say there is a seventh category or a fifth and there should only be five. That is the discussion I think we want to make sure we have

the right buckets to put these measures into.

Go ahead to the next slide. And for the outcomes you can see, again, our label of intermediate outcome, you make a good point. Maybe we don't need an intermediate outcome. Maybe that is a patient outcome and we say we want to get this down to three buckets and that better visualizes the data as to where we actually have measures. So that is our hope in this.

And I want to stop because I know I went over a lot of stuff, and I want to make sure that I was clear because I think a lot of the future discussions are going to drive off of what we are hoping to get. So I will stop here and see if there are any questions.

CO-CHAIR DANFORTH: Hardeep, do you have a question?

MEMBER SINGH: I don't know. Maybe I will wait for the last one.

CO-CHAIR DANFORTH: Okay. David.

MEMBER NEWMAN-TOKER: Can you go back

one slide for the process measures list? So could you tell us a little bit more? You talked about there being overlap between categories. Did you find a lot of overlap between those three ones that are talking about information gathering, interpretation, and integration?

DR. BERNOT: What we found was, again, based on our use of the words, we found that the most was over the information gathering and the information interpretation that a lot of times they were one in the same. This was run and it was determined to be blank.

MEMBER NEWMAN-TOKER: Right. So I mean I guess I think one of the questions here -- you know this process of developing the framework is all about deciding what level of granularity you want and how much lumping and splitting to do but I do think that it is not surprising that there was a lot of overlap in those spaces because they are kind of inextricably linked. So one might consider condensing and then consolidating them into clinical reasoning skills

or something else that can kind of weave all that together. I don't know exactly what, but rather than splitting things that are almost impossible to differentiate from one another.

MEMBER MCDONALD: I might argue the other direction though because there is like a lot of work that has been done sort of in the laboratory context that is really about the information gathering, getting that as good as it can get. Of course there is interpretation within the lab but that is still constrained to that piece. So it might be helpful to keep things in their separate buckets if there are actions that would be taken in separate buckets.

MEMBER NEWMAN-TOKER: Well, I don't feel strongly about it.

DR. BERNOT: And if you look, just as you glance through, hopefully this is pretty easy to read. But if you look at maybe pages 2, 3, 4, and 5, you will be able to see how some of the measures hit one, some hit both of those categories. There are a number that do fall in

independently but I thought that was the most overlap. But these are good points. We want to make the most -- the simplest amount of buckets or cubbyholes that we can put these measures in and actually have some value as to say this one is empty and this one is full and know where our priorities should be.

MEMBER NEWMAN-TOKER: And maybe to Kathy's point what we need to do is actually not just ask whether these things are theoretically different but whether they can be measured separately.

So in the sense of if information gathering can be measured by looking at charted documentation as a discrete entity and saying if it wasn't written down, it wasn't gathered, then maybe that is its own thing that is discrete and separate from interpretation.

CO-CHAIR DANFORTH: I think, too, there are some examples of those measures that have the check boxes in both. So we might even after want to look at some of the examples where

the NQF team has said we think it falls into both buckets and see if that is -- you know based on these descriptions, is that what we would think?

I'm going to go to Mark and then David.

CO-CHAIR GRABER: We had this discussion internally, whether these are the right buckets. An alternative way to do it would be to say that there are things that happen early in the diagnostic process versus things that play out over time. Would that be a more helpful way to classify?

MEMBER NEWMAN-TOKER: Maybe or things that happen at the bedside versus things that happen in the interactive process with the rest of the healthcare team. I don't know.

MEMBER SEIDENWURM: Just to understand a little bit more about how the categories were separated, would something like interoperability be considered information gathering? In other words, you could look in somebody else's -- when I say somebody else's, the same patient's but

some other hospital's or whatever information system, or would that be in the information integration, or would that be part of communication? How would we think about that? Because I think that we do need to get -- that would be back a notch under structure.

And there is a couple of those out there and I am not sure they were captured in the document.

CO-CHAIR DANFORTH: Hardeep and then Martha.

MEMBER SINGH: So I am going to just quickly reflect on the categorization and I am going to revisit some of the earlier discussion probably at the same time.

So I would think this is where those five dimensions that I walked people through this morning are really useful because even though they have overlap, they are sort of patient-centric. So you have got the patient engagement.

But there is doctor-patient interaction. So, this is where all the cognitive

stuff happens, the information gathering, the refinement, and you order some tests, you talk to the patient.

Then you get the test done, which is the lab or the radiologist. So if David, the neuroradiologist reads the MRI wrong, that is that dimension, which is the information interpretation.

The third one is follow-up of test results. So, he reads the MRI and I miss it. So I never followed up or I didn't communicate to the patient the third dimension.

The fourth is you can add subspecialty referrals and all that.

So I think that categorization avoids a little bit of this artificial distinction between some of the other integration and hypothesis generalizations because you just have one doctor-patient interaction as one dimension and then you put everything related to that within that dimension. So in this diagram, it is the brown one for instance. So that would be the

only way you could refine, if you wanted to, the categorization.

The second comment was sort of more revisiting the earlier discussions on are we trying to improve the 10,000 diagnoses or diseases, the diagnosis of those 10,000 diseases that WHO has or are we actually trying to focus on some narrow high-risk areas that we want to fix? And again, the point being patients care about delayed and wrong treatment, delayed and wrong tests.

So with that in mind, I was thinking a lot about the DVT example. And then when you said that 50-year-old with the CRC, with the colorectal cancer, I had actually written down earlier, just to given an example. So if your 50 year old man presents to the primary care doc and does not get offered a colonoscopy and a year later they have colorectal cancer diagnosed, what are we going to call that and why? Anybody have any strong preferences here?

How many would call it a diagnostic

error? A 50-year-old came -- okay, how many of you would call it not a diagnostic error?

Okay, how many of you want to know whether the physician asked whether there was a family history or not? What if I tell you that that patient had a brother who was diagnosed at the age of 35? They needed a screening colonoscopy but did they need a diagnostic colonoscopy or not?

So if I missed a screening opportunity because they had no family history of colorectal cancer, that is a screening opportunity. But if they had family history of colorectal cancer, that was a diagnostic opportunity. So I am kind of reflecting on David's point earlier that I think we are going to need to sort of think through some of these intricacies.

And then to point out there is lots of screening type measures here. So, are we done? I mean this is 56. Why do we need more?

Why do we need more? Why can't we just stop there and pick the four or five or

seven that we --

MR. LYZENGA: We could but they may not be the right --

MEMBER SINGH: Exactly! That is the point I am trying to make. They are not the right measures because they do not talk about the things that we all are interested in, which is missed and delayed diagnoses, which is, again, the framing point that I was sort of trying to make earlier.

We are going to have to come up with some priorities in what are the types of measures we are going to focus on.

So was there a missed opportunity to diagnose a DVT in your patient, for instance? That is the question. And if there was an asymptomatic patient, there was no opportunity to diagnose that patient with a DVT at that particular point of time. I would say you could call it a screening failure, or a surveillance failure, or a preventive failure but I'm not sure we can plug everything into a diagnostic category

and we are going to have to be really careful; otherwise, everything is diagnostic, which is the reason why we have not made any progress in this for the last several decades.

CO-CHAIR DANFORTH: Martha and then Kerm.

MEMBER RADFORD: Just a quick comment about the one structure measure. It probably is the same as David's but I just want to say it my way, I guess.

So it is helpful to sort of bucket the structure measures in the way that you have done but I also think it is helpful to somehow reference where in the diagnostic process this -- if this is an actor.

And again, on this one, the one that you have here, it is information gathering and then communication is where that acts in the diagnostic process. So I just would ask that if we find anymore structure measures, we at least reference that option.

And then I would also suggest that

maybe suggesting new structure measurement areas or measurement concepts might be a real service as well.

MR. LYZENGA: Sorry just to clarify, did you say that the structure measure we have here could also be -- it would be --

MEMBER RADFORD: Right, you have binned it in technologies and tools, which is absolutely correct. But it also could be binned in where in the diagnostic process. They are not mutually exclusive, that's all.

CO-CHAIR DANFORTH: Kerm.

DR. HENRIKSEN: Yes, one thing that the discussion has sort of made me think of is something that we really haven't talked about and that is measure usability. And these three areas of information gathering, interpretation, and integration. If the researcher and if the thought leaders that are on top of this subject matter can't really easily separate this cognitive process in these three stages that are very intricately linked and interactive, if you

can't come up with a useable measure that doesn't take two days to figure out and we actually agree on this, then it is a measure usability issue.

And so one thing that the framers of this could possibly do is see what the user and our reliability is in being able to categorize things in these areas. And if there is a lot of confusion and core interrelated reliability, then you need to collapse the category and create a higher order category, or hybrid, or a combination, or a composite.

MR. LYZENGA: You might have just created some homework for the committee. If you guys are open to that, we could try something like that and see how others assign these measures into the various buckets and see what kind of integrated --

MEMBER NEWMAN-TOKER: I would vote that when you do that exercise you try to also use Hardeep's four categories and see which one ends up being more accurately, reliably classified.

MEMBER IRONS: So I would just like to add my voice to some prior calls for consideration of a category for what to do with patients who are at risk for conditions that then require screening afterwards. You know if next-generation sequencing becomes more and more pervasive, as it is not targeted, we are going to have children that are currently being screened for a targeted reason and they are going to be identified with BRCA1 mutations, other adult onset conditions, and how that is followed. And if it is not, what we call that missed diagnosis, 10, 20, 30 years later is going to create a whole other category of concerns.

I mean it is probably going to be the biggest tsunami that is coming our way, in terms of presymptomatic screening.

MEMBER SINGH: Hang on. So you mean to say after they have had the next-generation sequencing?

MEMBER IRONS: Right.

MEMBER SINGH: Okay.

MEMBER IRONS: Yes, so you know for example, a child may have intellectual disability and anomalies and if it not targeted to the symptoms, there is actually a panel of adult-onset diseases that they are also screened for and not only them, but their families.

CO-CHAIR DANFORTH: Kathy and then Mike Dunne.

MEMBER MCDONALD: That makes me think it something with the comment. So I was just going to extend what we were hearing about the structure and the tie-in to the process.

I know we don't have a lot of structure measures but they said that this will be a whole. Some of the slides before and just the discussion makes me think that we need to be clear about these structures tying to kind of what the vulnerability is diagnostically, so it is not, obviously, just any structure. And I know you guys are thinking about that. It just wasn't in the text. You know they get a process tied to the diagnostic problem or diagnostic

quality structure tied to that. And I think with structure, particularly, it is going to be important to be a little more explicit about what that tie is, whether it then gets binned into some process piece or not, just tie it in.

The reason I was thinking it related to your comment, you know you can imagine that there would be a structure that would be some sort of registry that would allow some sort of population monitoring of a group of patients who are now known to be a higher risk and that that would be a system approach, not just the one clinician at a time approach to that potential need to sort of engage more proactively in using the diagnostic information that is available. So that could be a structure as well as then there could be gaps in terms of using that information at the clinical level.

MEMBER IRONS: But then just to follow-up on that, it will also require us to shift our definition of whose patient is really the doctor's patient. Because if the sequencing

is done on the child and the child has a BRCA1 mutation, then it either came from the mom or the dad. So is the child -- whose responsibility is it to follow the parent, when the child is somebody's patient? So it is a whole other can of worms.

MEMBER RADFORD: Just to focus also on your very good comment about the structure measures, very few structure measures have a structure outcome link and that is what I like to see as well on structure measure, even this one.

CO-CHAIR DANFORTH: Mike.

MEMBER DUNNE: I was just going to mention a possible additional source of information. I was just fooling around here and went to the LTRC, which is the full free text online legal review journal search system, and I typed in diagnostic errors and got some really interesting results back, one of which had to do with patient interaction.

So it might be another source to look for additional materials for your environmental

screen. That's it. Thanks.

DR. BERNOT: All right. Do you want me to turn it over to Andrew?

Just to wrap this up, and just to let you know, this is the -- the things we are going over is pretty much the last of what I consider the data dump to the group. And really, from here on out, it is trying to take all of these things we talked about and really come up -- do we have the right buckets, the right cubbyholes? And we can actually make those changes or get those suggestions today. And then how do those work with measure concepts?

So Andrew is going to talk a little bit about the measure concepts, as well as just a little bit of qualitative analysis he did, just to give a little more flavor to some of these measures.

MR. LYZENGA: We didn't actually include the concepts in here.

DR. BERNOT: Oh, sorry about that.

MR. LYZENGA: These are still just

environmental scan.

This is just very briefly just another, again, as John said, sort of another way to slice it, just trying to look at themes that emerged from the measures we identified.

And there is a range of different sort of topic areas. A couple of the major ones, as you might expect, we called a type of measure appropriate use of diagnostic criteria or diagnostic tools. And you might expect that to be the largest area of measurement currently because that is sort of the low hanging fruit when you have something very discrete you can measure. There is a test or a process you are supposed to follow for this particular condition. Did you do it? Again, kind of low-hanging fruit. Not necessarily where we want to be in really getting at the quality of diagnosis, at this point.

A number of measures that we have called sort of care coordination, communication between providers and appropriate documentation,

and that sort of thing.

Sorry, David did you have a question?

DR. HUNT: Is the category of availability and/or deployment of diagnostic resources part of appropriate use of diagnostic criteria and tools?

MR. LYZENGA: No, sorry. So there is a zero there.

DR. HUNT: Okay.

MR. LYZENGA: This is a little bit confusing.

DR. HUNT: So those two aren't combined?

MR. LYZENGA: No.

DR. HUNT: Okay.

MR. LYZENGA: And this is a little confusing. These categories were done with the concept, so there is a number of the measure concepts that did fall into these categories but they are a zero for the actual measures that we found through the environmental scan.

Completeness of diagnostic assessment

was another sort of the larger. And basically, trying to determine whether -- and that had a lot to do -- and maybe that is the wrong name for it but whether, again, things were appropriately and fully documented, whether, in some cases, the diagnosis was granular enough to be useful or whether it was recorded in an appropriate way and communicated. Next slide.

I just wanted to go through that very quickly.

And then to what we were talking about before, we tried to break it down a little bit into which sorts of clinical conditions the measures applied to.

And to Prashant's question I think there is not a lot, four I think here, that you could consider cross-cutting, that didn't have a particular clinical condition associated with them. Lots in the area of oncology. Again, as you might expect, it seems to be a prominent area.

There are some of these like dementia

and stroke that could probably be considered all under neurology. That was another sort of large category.

And I should note that many of the ones that we considered, oncology, and a number of the others, as well, were maybe also more -- you might more preferably consider them lab-related measures. There is quite a lot of those of whether you did the appropriate laboratory tests and then reported those laboratory results in the appropriate way. They just happen to be focused on cancer conditions.

So, that is it. I just wanted to give you sort of another slice at the data and way of looking at it.

But again, to your question before, most of them were condition -- were procedure-specific, not a lot of cross-cutting measures, at this point, although the concepts that we received and that we pulled from Hardeep's paper and that Mark suggested, there were a lot more cross-cutting ones that we may want to propose

for future development.

CO-CHAIR DANFORTH: Prashant.

MEMBER MAHAJAN: I just wanted to mention, at this point, that as we think more about this, we could consider that condition-specific, which would be certain high-risk conditions, specialty-specific, which could be by different age range, but I would also want to control by healthcare setting-specific because certain conditions may be diagnosed specifically only in that situation versus not.

CO-CHAIR DANFORTH: David, and then Hardeep, and then back to Mike.

MEMBER NEWMAN-TOKER: So just in terms of the last slide of how they were aggregated with a lot of sort of discipline-specific stuff, I would discourage us from doing that too much because I think it bespeaks of the compartmentalization and siloing of the care and we really are trying to kind of -- that the diagnosis is the outcome of the team process. And I think when you say well, this is your

piece, and this is your piece, and that is your piece, I think it takes in a bad direction overall.

I am really surprised -- I'm not surprised but I am noticing that more than half of your process measures are not what I would call diagnostic. I would call them either screening, risk assessment or staging.

Now, I think, at some point, we have to come to either an agreement or you guys have to just say that those things are in or out but we have to know that when we are arguing about what the framework should look like and how many things fit into a cubby and so on and so forth because I would not have pulled these out and said these are measures of diagnostic performance in the traditional sense.

I realize diagnostic testing is involved but there are a host of these things, including monitoring for treatment complications and other things that involve diagnostic tests but don't fit the sort of what comes to mind when

most people talk about missing or not getting a diagnostic error, or not getting a diagnosis right in a patient.

MR. LYZENGA: Yes and we had this same debate as a team as we were looking through the measures and we made some decisions but we said should we or should we not include these types of measures and wanted to bring it to the committee to let us know whether we ought to include those kinds of measures.

Those were some of the major questions measures where you have a diagnosis, diabetes say, and you are screening for likely comorbidities or common comorbidities. We initially included those, said that those were diagnostic-related. But if the committee thinks that those should not be included, we don't feel strongly about it.

Again, we did not include the asymptomatic screening measures that we found. We figured that was too sort of far afield.

Some of the other ones, again, like

you said, the staging of disease, we did include those but we don't have to.

I can't remember what some of the other sort of categories were but we would welcome any thoughts from this group on anything you see in that inventory, at this point, if you think it is appropriate or inappropriate to include and why or why not.

CO-CHAIR DANFORTH: Hardeep and then David.

MEMBER SINGH: So I was just thinking I think this is coming together very nicely.

So I am just going to sort of reframe some of what you said and then maybe propose some actionable next steps. So, I am just thinking.

So everybody is agreeing that the structure process outcome is, in general, a good way to approach. I think this is good. So that is like the Step 1 of what we are going.

The Step 2 is I think we are coming up with subdomains that we want to try to include within structure process outcomes. So I say

within structure, we have got the IOM, lots of tools, technology, work environment kinds of things. We can add the 8-dimensional Socio-Technical Model just to complete that.

But I think more of this is this is an approach to make sure that we are not missing anything, rather than having every one of these be accounted for for a measure.

So in the same way, in the process one, you have got the IOM processes but we can include say for these five processes, we are getting to the point of reminding ourselves that this is just sort of a thinking, to inform our thinking of measures.

And for outcomes, it is generally patient, providers, and systems; patient, providers, and care teams, and then systems. So that would be sort of the organizing framework.

And then is the important one. I think that is sort of Step 3, which is sort of what we are struggling with. And I just quickly wrote down, especially after what Prashant said,

you know I think we are thinking about this is where we are going to need to have discussions. What are the high-risk conditions we should put forward for proposed areas of measurement?

So they could be high-risk because they are certain conditions or diseases, so infections, cancers, and cardiovascular conditions came up, or they could be high-risk populations. And the high-risk populations could be children or it could be other vulnerable populations that we have talked about, other sort of patient disparities.

There could be high-risk settings. So we know emergency rooms are high risk. We know primary care is high risk. And so we come up with a list of the third layer, if you will.

And then I am wondering we had a prioritization exercise at the committee, HIT Safety Committee, and we recently had one at the Diagnostic Error meeting in Los Angeles, where the 20 measures, 19 or 20 measures or measure concepts that we had in the Journal of Patient

Safety paper that came up this morning, we just asked people, I mean asked basically rate them A through F and the ones we discussed would be where people mostly ranked Fs or As just to get the other's perspective. There was some diversion.

So like half the room said A and the other half said F. And then we would say okay, let's understand why. And I think that prioritization exercise really helped. I haven't looked at the data but I have the data. Of the 19 I think the committee or the group, rather, agreed to about six or eight of them being really superstar measures that they thought should go forward for further development. And some they just ruled it out out of the list of 19.

So I think that is sort of the way I am thinking that we are bringing together some of the concepts.

And some of the screening stuff we will just weed out because people won't agree.

MR. LYZENGA: Right. And sort of a

sneak preview, tomorrow we are intending to try to get you guys to brainstorm sort of just any concepts that you can come up with for measures of diagnostic quality or whatever you want to call it, just to come up with as many things as we can. We will then sort of map those against our current framework as sort of a test case.

Again, I think it would be useful to see if when we come up with a large number of concepts or measures, are these fitting in well to the framework. Do we need other categories? Is this not adequate or appropriate?

But then as we get -- we will have those concepts and we will, moving forward, probably in the next meeting or after that, do a prioritization exercise and say which of these concepts or measures do we think are really important. Which are less important? And try to sort of things out that way, very similarly to what we did in HIT Safety.

CO-CHAIR DANFORTH: David, Mark, and then Martha.

DR. HUNT: I don't want to get stuck on this. And I definitely don't want my voice to be the loudest but I just want to point out that the discussion of risk in screening, they are really the same thing. Screening is what we have identified as high-risk areas.

And to break the tie, and I don't know which way we should fall, but to break the tie, would it be at least reasonable to consider, from the patient perspective, what would a patient say? A 58 year old man who has never had a colonoscopy, should my doctor have done some screening to diagnose that colorectal cancer? So from the patient perspective, would they consider that a misdiagnosis and could that be a reasonable way to sort of break the tie? Just offering up there.

MEMBER NEWMAN-TOKER: We should hear from our patients but I think -- I should think so. I mean I think, again, patients care about getting harmed. They don't really care about what we call things or which buckets they are in,

I think is my general sense.

And I think, for instance, if we ask Jen whether staging was part of the diagnostic process, I think we would get a very clear answer, that the diagnostic error she suffered was part of the staging process.

CO-CHAIR GRABER: Yes, we know one of the biggest buckets for diagnostic error is delayed diagnosis of cancer and the biggest bucket within that is failure to screen or failure to follow-up on screening. So I would feel very bad if we left screening off the table.

I would like to get some sense of the group. Is it okay to leave screening inside our bucket? We are going to include screening.

MEMBER SINGH: We have screening measures on cancer, don't we?

CO-CHAIR GRABER: Yes. There seemed to be some sense that we wanted to exclude screening. I didn't quite follow that argument.

Is everybody comfortable with including screening within our domain?

MEMBER SINGH: I mean in VA our rates of colorectal screening are so sky high, we are not investing as much. It is like plus 90.

So I don't know.

MR. LYZENGA: Again, we made the distinction between asymptomatic screening and screening for, again, likely or common comorbidities. And we kept the latter but did not keep the former. And I don't know if others have thoughts on that.

Should we be keeping asymptomatic screening?

MEMBER NEWMAN-TOKER: I actually think asymptomatic screening is more important to keep than what you have called symptomatic which, in my mind, isn't exactly the right way to refer to it. It is standard disease-specific asymptomatic screening.

You are saying if someone knows that they have diabetes, there are some known complications that they should be monitored for, which is slightly different than sort of general

population screening of people of a certain demographic group. So it is symptomatic in the sense that they know that they have diabetes. It is asymptomatic in the sense that they may not know that they have numbness of their feet. They may not know the actual symptoms of the complications.

I am fine keeping screening in. I do think there is an even bigger rung of clinical decision-making around risk stratification that we got into earlier, which was the run around the DVT, which that is neither are you screening an asymptomatic population with a diagnostic test to look for a specific disease, nor are you investigating a symptomatic complaint. You are actually just applying treatment based on a risk stratification.

Now you could argue that if you don't do that, if you don't ask them the questions about the risk factors, that that somehow is a diagnostic screening process that you should go through to get the point of treatment but I do

worry that if you actually draw the confines around what that includes, I think you will find that almost every clinical reasoning decision will fall into that bucket in any treatment scenario. Like you are in the middle of a code and you don't accurately risk stratify which pathway you are in in the code. Is that a diagnostic error? In some sense it is. It is a clinical reasoning error, whatever. But I don't know whether we want to go there.

MEMBER SINGH: I mean here is an NQF measure. Colorectal cancer screening, percentage of patients 50 to 75 who are appropriately screened for colorectal cancer. So why are we going to pursue something that is going to reinvent the wheel?

MR. LYZENGA: Well we are not proposing new ones, just whether that should be included in our inventory and then included in our prioritization exercise. Or should we put those measures off to the side and say they are not diagnosis related, we are not even talking

about them?

DR. LUSTIG: I think what we were thinking, and again, this is just our initial decision, was there were so many screening measures that we were thinking about conceptually as, to use the words of the Academy Report, a patient comes to the office with a health problem of some kind and what you do from that point forward. And so we were trying to eliminate all the general screening recommendations, not just you have a 58-year-old male in your office and you should be recommending screening to get to a diagnosis. But this is what we wanted to present to you and you could say no, we do want to include all this.

But we were starting from the point that someone has come to you with a problem and what you do with that. I don't know if that helps clarify.

MEMBER MCDONALD: I think in the text somewhere, we would have to check, but I think the text actually suggests that coming with the

problem can include coming with the need for sort of asymptomatic screening because you are a certain age. So I don't think it was off the table from the conceptual perceptive at that point.

DR. LUSTIG: I thought heard the term health problem.

MEMBER MCDONALD: I know but the health problem in the text I think is described as potentially including this.

MEMBER NEWMAN-TOKER: Well I do think there is an intermediate possible solution, which is to include it but explicitly put it off to the side and say look, screening is part of the big diagnostic process but there is a whole set of U.S. Preventative Services Task Force recommendations and so on and so forth and we were addressing a different piece of the measurement, probably where there aren't robust measures and so on and so forth.

CO-CHAIR DANFORTH: Okay, we are going to go in this order, on the left side of the

room, my left, Martha, Helen, Prashant, we are going to come back to Mira, and then Kathy, based on how I saw cards go up.

MEMBER RADFORD: I am going to agree with the last comment and just say we were asked to develop a measurement framework. And I think we can certainly call out screening as part of that and acknowledge that there is a fair number of screening measures already. And if other ones need to be developed, it is pretty clear how to do it because it has been done so many times.

I really think that the biggest contribution we can make is in the empty cubbies and to point out where we need new measure development work, where we need research into the structure and process outcome links that support measures, and where you think -- where we think, really, that the measurement, that the new measurement low-hanging fruit might be.

CO-CHAIR DANFORTH: Helen.

MEMBER HASKELL: Well, I agree with that. I don't support the inclusion of

screening. I think we can say there are screening measures -- the U.S. Preventative Task Force Services. My other concern about screening is overdiagnosis. So, I agree, really, with everything that Martha and David have said.

I am not at all uncomfortable with this small measure set and identifying gaps and ambitions.

CO-CHAIR DANFORTH: Okay, some cards went down. He is thinking about putting it back up. He is not.

So, Mira and then Lavinia.

MEMBER MIDDLETON: So I am going to suggest that one way of bringing this altogether may be to focus on risk. And patients are either at risk for a condition because they come with symptoms, they have physical findings or symptoms and they are ill, or they are at risk because they have a family history of a condition that gives them increased risk, or they are at risk because they have a condition diagnosed in childhood that has late onset problems, or

because they have some screening test that is either a genetic screening test. And maybe that is one way of prioritizing it, keeping the screening in there.

Because I think as we go on, more and more patients are going to be at-risk for symptoms that lead to misdiagnoses based on conditions that are diagnosed early in childhood, rather than those presenting with symptoms. We may be able to presymptomatically diagnose that.

So focusing on risk and why you are at risk may be one way of bringing it all together.

CO-CHAIR DANFORTH: Lavinia.

MEMBER MIDDLETON: So I think that many of the process risks related to pathology and diagnosis have been well-established, to the point where if you look at the delta of the hospitals and physicians who report out these measures that is very small. So like last year, greater than 99.6 percent of all reporting hospitals reported that their Barrett's esophagus reports had a statement on dysplasia.

I think the real opportunity is with the communication of the diagnosis. So for each one of these that you have for the measure, I would focus more on communicating, making sure that there is processes in place and documenting how this information is communicated to the patient.

MR. LYZENGA: We will keep that in mind for tomorrow when we are coming up with concepts.

CO-CHAIR DANFORTH: I was so preoccupied with left side I didn't see on the right side who put their name card up first. So I am going to trust that one of you will let me know.

MEMBER HRAVNAK: I just wanted to speak in favor of including risk assessment and screening. And I think my reason for that is that I think we need to think sometimes about what it is that we are diagnosing. Are we differentiating between a primary diagnosis of a disease process versus catching a diagnosis of a

complication? And again, I just worry that if we leave out the risk assessment and screening that we are going to lose that kind of -- do we wait for somebody to develop symptoms before we start jumping on the bandwagon or do we try to prevent those?

There are many screening tools for that long list of 15 complications leading to failure to rescue, you know infection, DVT, GI bleeds, skin ulcers. So I just feel like we are really losing that end of the diagnostic spectrum if we don't include that.

CO-CHAIR DANFORTH: David.

MEMBER SEIDENWURM: So I agree that we should include screening with sort of a sidebar that there is enough done on it. The one area in screening, and I think we should maybe suggest this as a population health type of metric, is in the area of overdiagnosis which I believe is principally noted in large groups of people. And I think that we would be remiss, perhaps, if we didn't include some kind of community or some

unit overdiagnosis metrics with respect to perhaps thyroid, breast, and prostate cancer, which I think are the most suspicious for that in our society right now.

So maybe we could propose that as one gap but, otherwise, I think the focus should not be on screening.

CO-CHAIR DANFORTH: Helen.

DR. BURSTIN: Yes, just this has been a great discussion. It raises a lot of good questions for us.

I think one of the other ways, besides cubbies, I tend to think of frameworks as trees. And in some ways, it may be very illustrative to show the volume of measures that fit only into this one branch, only this one cubby to the exclusion of everything else. So by actually showing how many are really just about diagnostic testing, et cetera, with no then connection to communication, with no connection to how it is used, may actually be very illustrative to make the case of why the other branches shouldn't be

there.

And so actually loading this one up may not be a problem but it might be really important for us -- I forgot who said it. I think it was actually you, Missy, who made the point of who often in fact some of these measures cross some of those domains and how often they are just lone wolves out there and did you just collect that data, information gathering phase, where we have got plenty. But to Hardeep's earlier point, I don't think this is what many of us think of when we think of looking at the safety and quality of diagnoses.

MR. LYZENGA: Just one more sort of nuance that I just remembered. We also did not include measures that were focused on overuse of diagnostic tools like MRIs for back pain. We, again, excluded those and I just wanted to throw that out there and see if others wanted to include them on that sort of overdiagnosis theme.

MEMBER NEWMAN-TOKER: I think what you want to do is not include all of them. I think

you want to marry those to underuse in those domains. I think you want to make an explicit pairing for people so that they understand that they have to monitor both sides of this argument, both false positive, false negatives kind of side of things.

CO-CHAIR DANFORTH: Okay, actually --

MEMBER MAHAJAN: Can I? Sorry. Can I just --

CO-CHAIR DANFORTH: Yes, one second. David Grenache has been so patient, actually. So, we are going go David on this side of the room, then Kerm, and then Prashant.

MEMBER GRENACHE: Right. Thanks. I don't to belabor the risk versus diagnosis issue. I agree with David Seidenwurm.

I'm not a physician. I am a Ph.D. clinical chemist. I do a lot of screening tests in my laboratory. I am not convinced that assessing someone for risk is the same thing as a diagnosis. You can be at risk for a disease but never develop the disease. That's obvious.

And there is a slippery slope here because, as you screen people, you are going to end up with lots of false positives, this whole premise of screening, and you are going to end up with overdiagnosis and potential harms. And there are very well-known harms from screening.

CO-CHAIR DANFORTH: I almost wish we could develop, I think it is Hardeep's term, a calibration measure, a structural measure focused on calibration. What are you doing at a facility to balance overutilization with underutilization related to diagnoses? I mean I think everything I am hearing sort of gets that issue that Hardeep brought up very early this morning around calibration.

Herm.

DR. HENRIKSEN: Just going back to Helen's comment, whatever we do I think screening and staging and those types of processes have to be acknowledged and the rationale.

The screening issue is unresolved but it, indeed, has to be acknowledged. And whether

it is in a sidebar or an appendix, or some other way of treating it, it has to be certainly acknowledged because it would be deficient not to acknowledge it.

And so the way you treat it in any final report or any final classification system, as long as you justify it and explain it clearly, will be serving a valuable purpose I think.

CO-CHAIR DANFORTH: Prashant, and then Helen, and then Hardeep.

MEMBER MAHAJAN: Yes, I just wanted to make a nuance to what David -- DNT said, is if we are going down the path of overdiagnosis and overtesting, the real-world situation is patients who are seen in the emergency departments and some other places frequently get overtested because of the environment that they practice in. So we have to nuance that factor. If we are going to do certain disease-specific high risk, high reward type of a condition, then that approaches better but if you are just going to go into this wholly of overdiagnosis and

overtesting, we have to be a little careful.

CO-CHAIR DANFORTH: Helen.

MEMBER HASKELL: So my concern with this, aside from the overdiagnosis, is with all these little measures on these very specific topics that it really detracts from the overarching themes we are trying to emphasize.

I am just concerned that it is confusing. Am I saying something that has already been said?

MR. LYZENGA: No, I was just saying we will try to fix that tomorrow.

CO-CHAIR DANFORTH: Hardeep.

MEMBER SINGH: So I was just going to add, sort of building on what Helen just said, I think it is probably okay to sort of just say yes, we have thought about screening but one thing that sets us apart, we should all agree that we are focusing on a specific or defining the problem that we are trying to solve.

And I thought, putting in the context of risk, I think what we are trying to do is

looking for high risk situations for either missed, delayed, or wrong diagnosis. So whether you look at it from any perspective, if our rationale for doing some kind of prioritization exercise tomorrow or whenever is around is this situation or measurement concept really relevant for conversations around how you reduce the high risk situations for missed and delayed and wrong diagnosis. I mean if you got a shared understanding, that is useful.

I am not sure why we should spend too much time on screening when we can't even do basic history and basic physical exam anymore. We just had a study that is going to come out hopefully in a few months that about half of patients who have spinal epidural abscess were seen by multiple physicians with red flag symptoms, including fever, neck pain, and I could go on and on, neurological symptoms. I mean no matter which conditions we look at, we are finding we can't even fix the very obvious stuff and we can't even recognize red flags anymore.

So I am not sure how much emphasis on screening will help.

CO-CHAIR DANFORTH: I guess I think maybe I would like verification on one point. So I understand the concept and the uses of the framework. I am wondering if this is the right time to ask the question that based on the comments that people have had and the discussions we have had about additional work we might do to the framework is are we saying that potentially if we develop a framework like this for improving diagnostic safety and quality that the framework would be appropriate for diagnosing diseases -- so they come to the emergency room with dizziness and it is a stroke but would it be appropriate for diagnosing for complications? So back to Marilyn's point.

And if we are not saying that, then can we leave the framework open for both kinds of measures or are we saying that framework is appropriate for one and not the other? I think I would like clarification on that.

MEMBER RADFORD: I would say that, again, this is a document that we are producing and it is a lot of recommendations. And part of the recommendations can be basically the NQF full employment act, where we have subsequent groups that are going to look at aspects of this that we are not focusing on because we are prioritizing what we are going to focus on, which will happen, I guess, over time.

So I don't have a problem with the document setting a framework for not just this group, if you will.

MR. LYZENGA: Yes and I think that is the intent is both to serve as sort of an, again, organizing tool for us but then as a framework for others moving forward to identify what the gaps in priorities for measurement are? I mean as a tool in both of those senses.

I mean I suppose we should have some specificity about we are including and not including. But in terms of the specific measures I get a sense that these screening measures are

going to be a fairly low priority anyway. And to Martha's earlier point, I think the real important work that we are going to be doing is seeing those gaps and trying to come up with ideas to fill them. But I don't know.

Any more thoughts to help us clarify what we should or should not include would be welcome. But again, I think the more important part will be coming up with new ideas.

CO-CHAIR DANFORTH: Okay, Nicholas.

MEMBER KUZMA: I was just going to say that for the complications of the diagnosis, in some ways I think that is a communication error. Because you have made the diagnosis. You also need to be communicating what are the downstream effects of that diagnosis. So, if you are looking where in the framework to fit that, I think that is kind of where I would think about putting that.

CO-CHAIR DANFORTH: Kathy.

MEMBER MCDONALD: I think some of these discussions are showing us that there are

some places where it gets a little squishier. You know and if we are thinking sort of patient-centeredness, this idea of what the consequences of a miss is, whether it is from a screening perspective or a risk assessment perspective, or a straight out more clear area of diagnosis. So this idea of kind of the write-up and the framework, I think we are still going to be challenged. But even writing up or framework-wise, as long as there is the desire to have cubbies and split, it is tougher where there is these tensions and where it is squishier. So, that will be something to have -- you guys will get to figure that out but it exists in this area. It definitely exists in this area.

CO-CHAIR DANFORTH: David Hunt.

DR. HUNT: I just want to repeat what Hardeep said also that I think it is more important if we come up with a system that somehow or another down the road doesn't find a place for something that is regularly understood to be a diagnostic error. I think that is the

better -- that is the worst error we can make. If something fits into multiple places, or we have a little bit more, having a cubby too small, I think is the biggest problem.

And I guess we are honing in on the idea that this won't be among the very long list of perfect frameworks that NQF has developed before.

CO-CHAIR DANFORTH: Carlos.

MEMBER HIGUERA RUEDA: I keep listening around the room and I think we are all saying the same thing all over and over again. And I don't think that we can pretend to resolve all the issues in medicine in this where it is creating -- because at the end of the day the only thing that we do is diagnosis and treatment. Well, you can mention maybe some prevention as well. It should be very important.

But regardless of that, I think that we are losing a little sight. I mean we have very good feedback from patients here. And what I am hearing is that communication is a big issue

and I feel, at least in my clinical practice when I see with my colleagues that that is certainly true. I think that we see a huge area of measurements that come out every day that make the practice of medicine miserable. I mean like a lot of physicians here, you won't disagree that is very cumbersome to practice nowadays because are doing all this -- we have a list of I don't know how many measurements here that we have to fill in the computer and paperwork and so forth and we are losing the sight of practicing medicine. And I really would like to see in this group to be very pragmatic about the measurements that we are going to come out with to -- I like some of these examples here, where a structure and process are really important around the patient and to give their providers some measurements that are really going to give support to enhance that communication. I think that that is really where we are failing because everything else, if we go to the minutia of data, a lot of these things we don't have enough data

to make recommendations. Maybe we can make recommendations of research questions that need to be answered on a lot of these things. That is for sure. But other than that, I really would like to see that those measurements rely more on the communication, the process, and, again, give the providers what they need to practice good medicine.

CO-CHAIR DANFORTH: David.

MEMBER NEWMAN-TOKER: So several people have said things that allude to this issue of how granular our task should be. So when we talk about this issue of lumping and splitting, if we said our measure framework is structure, process, and outcomes, and our measurement concepts within that framework are the things written on this page, you know these subdomains, as you call them, we could probably be done and we could all agree within a few minutes that that is a good conceptual framework.

But if we want to go one rung deeper than that, which I think we do, to say within

each of these subdomains, what are the sorts of measures that could be built or developed. We have to find that kind of middle level between the subdomain you have listed and the concrete measures you have listed that doesn't get us stuck in the weeds around well is this diabetes retinopathy thing an important thing or not an important thing because we are going to be in deep trouble if we are there.

But saying okay, this is a class of measures about asymptomatic screening in patients with known conditions that have disease complications. The measurement concept is about monitoring for disease complications. And there is a different one for measuring for treatment complications, when you give someone a drug or you do this, or treatment complications after surgery, you know the DVT in post-op, or whatever. I think that is the level that we want to hit and I don't think we have got the words on the page yet. Hopefully, we will do that tomorrow. But I think we have to be careful

about not getting all the way to the level of granularity of what these individual measures are and that we dial it back a little bit to get something that is both informative but not unmanageable.

CO-CHAIR DANFORTH: Prashant.

MEMBER SINGH: Would Andrew mind showing the committee just maybe like an example from the HIT Safety Report? Because I think that is what sort of David is also suggesting that giving examples and showing high priority areas and then maybe just a concrete example. We have been calling them measures. We call them measurement concepts. Examples may be worth it. I don't know.

MR. LYZENGA: We'll see if we can pull that up. But David, I think you put it really well. I think that is exactly what we are trying to do. We are trying to get a level below these subdomains and try to flesh out a little bit what are the types of things, sort of measure concepts. Again, conceptually, what do we want

to be measuring within these domains? What is important to be getting diagnosis right and avoiding errors?

To the extent that we do come up with anything that can be a little bit further specified, not to the point of a fully implementable measure but if we come up with here is a potential numerator and denominator for this measure, that would be helpful, I am sure, for the field and measure developers moving forward but I don't think we need to get to that level. But we want to get to the level you were talking about, at least.

MEMBER NEWMAN-TOKER: But I think if you wanted to do that, you might want to do it for the specific domains that are known to be high risk.

Like you might want to say okay, look, for cancer, infection, and vascular events, we want to do two rungs deeper than where we are with the subdomains, rather than just one but we want one for sort of the big picture.

MR. LYZENGA: Or even identifying, again, what those key conditions are and saying these are some conditions or areas where we think it might be fruitful to pursue specific measures within.

CO-CHAIR DANFORTH: While the NQF staff look for the NQF staff look for the HIT example, we are going to take about a 10- or 15-minute break and then come back.

So Tom, if you are still with us on the phone hanging in there, we will be back in 10 or 15 minutes.

MEMBER SEQUIST: Okay, thank you.

(Whereupon, the above-entitled matter went off the record at 3:08 p.m. and resumed at 3:38 p.m.)

MR. LYZENGA: All right, so for the next 20 minutes or so, or however long it takes, we are just going to very quickly run through the concepts that we received from Mark and that we pulled from Hardeep's paper, just to give a sense

of what we are looking for tomorrow, as we are trying to brainstorm concepts. I just thought it might be useful to quickly run over this so you can see what we are looking for and what you might want to be thinking about tonight and then looking toward tomorrow.

So, let's see. Can you blow it up at all? Can you scoot over at all? There, just to those, yes.

So again, just to give you a sense, these are the sorts of concepts that we were hoping to have you come up with tomorrow. These are not fully specified measures, like we have in sort of this other preliminary inventory. They are generally a bit more vague, sort of conceptual. This first one, for example, staff involved in diagnosing patients have appropriate competency to do so.

If you were to create an actual measure out of this, you would have to define a whole lot of terms in what competency means, in this case, who the appropriate staff are to be

measured, who is involved in diagnosing patients, lots of details that you would have to flesh out and define and specify for that. But for our purposes, we just kind of want to get some ideas of what a good thing to measure -- if you were to pursue a measure like that, we would like to see a measure assessing that the staff involved in diagnosing patients are competent to do so.

The same with the next. The provider mix is appropriate for the complexity of the case.

So just a fairly high level, not at the high level of our framework domains like information gathering. That is a little bit too broad but not quite so specific that you can actually go into a hospital or a clinician office and implement the measure but an idea of what a measure might look like, if we were to pursue it.

If you would scroll down to the process one so we can see some examples of those.

I think we were talking about it a little bit on the break and we are not going to

try to have you assign these to our framework subdomains. We are just going to stick with the broader overall topics of structure, process, and outcome. We are going to break you into some smaller groups and have you just come up with as many of these types of concepts as you can. So just think about what is important to get diagnosis right. What are some things that might be measurable in that area? And then to think about -- you know you can use the subdomains that we have as sort of a mental framing. Try to spread your concepts across those different subdomains to the extent that you can so we don't have a lot bunched under tools and technology, although if we do, then that's okay if it is a really important area but sort of use that to guide your thinking a little bit. But you don't have to say this measure is going to be in the information gathering stage. We won't make you do that at this point. But just think of structure, process, and outcome measures that would be important to evaluating the quality and

safety of diagnosis.

And again, I think you have these documents in front of you, if you want to take a little bit of a closer look.

So this is just to give you an idea of what we are looking for tomorrow.

I don't know if we have anything much else to cover right now. We could --

CO-CHAIR GRABER: Andrew, could I ask a question? This is all new to me, writing measures. And it seems to me that certain things can be stated as either a structure or a process measure. And I don't know how you decide which is better.

Like here it says second opinions are available. So as a structure thing, if you could rephrase it and say 50 percent of cancer patients get a second opinion. So, is there a preference for how you would like us to do those?

MR. LYZENGA: No, and in fact you could do both, if you like.

CO-CHAIR GRABER: We could do both.

MR. LYZENGA: Yes, I think we -- and then we can -- the idea right now I think and tomorrow is just to get as many ideas as we can on the table, basically, as many ideas for measuring and measure concepts as you can.

We have some subsequent meetings, where we will go through and we will kind of refine those ideas. If we have things that are kind of competing like that, maybe we can decide we would prefer a structure measure over an outcome measure or, rather, a process measure in this area where we think we can combine these two or we can rephrase these a little bit and then we will prioritize. We will have an exercise where we try to decide which are the most important and which are the least important but we will do that in a subsequent meeting. Right now, again, just trying to get everything we can out on the table so we have some material to work with, essentially.

Do you have a comment, David?

DR. HUNT: Yes, basically to that end,

we don't necessarily have to identify, even if it is structure.

MR. LYZENGA: Fair enough.

DR. HUNT: If you have a great concept you know fits into this domain, that could be work for others later down the road.

MR. LYZENGA: Absolutely.

DR. HUNT: Because that fine splitting of hairs is sometimes very difficult.

MR. LYZENGA: Yes, I think that is a fair point. So we don't even have to think about it at that level. Just anything you can think of that might be worth measuring, some idea of a thing that might be measurable or something that is important in a diagnostic process or as a diagnostic outcome that we should consider.

MR. LYZENGA: Any other comments from our co-chairs or anybody? I think we can probably end a little bit early, then.

Will you pull up the dinner slide again? So just to remind everybody, 6:00 p.m. dinner at P.J. Clarke's. There is the address.

I hope you all can attend.

Otherwise, we will see you tomorrow.

MR. EPNER: Do we have any public comment?

MR. LYZENGA: Oh, Paul, you are right. Public comment, absolutely. You are doing my job for me.

So yes, we will get comments in the room and then maybe we can turn to the phone after that.

MR. EPNER: So I have to say two things, even if the train has left the station.

So on the issue of how broadly to go and whether to include screening, et cetera, I would encourage the committee to go very broad in its description of the framework, even if it says in this body of work we can only focus on complaint-specific.

I think of all screening as, building on Mira's comment, is risk-based, whether the risk is because of age or gender or because I am having surgery tomorrow or because I am having a

certain kind of treatment. It all influences pretest probability or pretreatment probability of success.

So, again, I would just go very broad at the high level so everything has a place, even if you can't do that, focus on it now.

And the only other comment is on the term overdiagnosis. Some people have heard me talk about this before. I think it is a term that gets lots of use and it is an inappropriate term in most cases. In most cases, it tends to be a bad treatment decision. You are acting on something that you didn't need to act on. Or it is a misdiagnosis. You are calling it something when it is really something else.

The case where you are diagnosing accurately and has nothing to do with the treatment as some kind of syndrome x and it is an overmedicalization, from what I have read, that seems to be a pretty small percentage of what is going on when people talk about overdiagnosis.

I think we know words matter. And I

would hope and encourage the committee to be careful in joining the bandwagon of this overdiagnosis to bring focus to an area of overtesting, where people do things with it they shouldn't be doing, which is an overtreatment problem.

So, I would just ask us to think or ask you all to think about that. Thank you.

MR. LYZENGA: Thank you. Yes, great point.

Operator, do we have any comments on the phone?

OPERATOR: Okay at this time, if you would like to make a comment, please press * and then number 1.

And there are no comments at this time.

MR. LYZENGA: Okay, thank you.

Yes, Tom, do you have anything to add?

MEMBER SEQUIST: No, I'm all set. Thanks.

MR. LYZENGA: Okay, great. Thank you.

All right, well we are adjourned then, unless there are any other remarks.

See you at the dinner, if you go, and if not, see you tomorrow.

(Whereupon, the above-entitled matter went off the record at 3:48 p.m.)

C

Q

R

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Improving Diagnostic Quality and Safety In-Person Meeting

Before: NQF

Date: 01-10-17

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter