[Side conversations from 0:00:00 to 0:02:09]

Dan Perry: Good morning, everyone. The doors have been closed and you're locked in for the next few hours, but we promise a very interesting and possibly even exciting and inspiring meeting.

My name is Dan Perry. I'm the president of the not-for-profit Alliance for Aging Research. And that's a quick blurb on the Alliance for Aging Research. We have been at this for approaching 30 years as an advocacy group seeking to increase the national investment in research and development of treatments, therapies, preventions, better health-care modalities for the growing older population.

And that's the last you'll hear today of the Alliance for Aging Research, because I am also here as the chair of the Aging in Motion Coalition. And Aging in Motion or AIM is your host for today's discussion of patient-reported outcome measures in sarcopenia.

The AIM Coalition is a collection of over 30 not-for-profit organizations working to press for more research and innovation to develop treatments and preventions against age-related muscle wasting, sarcopenia, and its associated functional decline.

We all know that to lose one's independence, the inability to climb even a few steps, to move from room to room, to travel, to be an active member and a contributor to one's own family and one's own community is indeed a great diminishment of our quality of life. And given what we know lies ahead in terms of the aging of our population and, indeed, the populations of most of the world, we know that this condition and the functional loss, functional decline associated with aging is going to be a great challenge and threat to our entire economy and quality of American life.

Yet, at present, there is no FDA-approved measures to guide the development of potential drug therapies and other treatments for sarcopenia, and no pathways for approval nor for Medicare reimbursement for what we hope to see as a new generation of interventions, treatments, preventions of this condition.

Now, any coalition, by definition, is a gathering of diverse groups. And the AIM Coalition is indeed made up of a great diversity of organizations. It includes patient organizations. It includes medical experts in physiology and nutrition. It includes caregivers and their representatives, provider organizations, health

professionals, employers and health-care industry, ethnic interest health groups. And AIM also has international ties through its members especially to Europe.

I want to give everyone a very hardy welcome and a thank you for being part of this first ever AIM allies meeting seeking to be the icebreaker to plow the way ahead toward a way to develop treatments, therapies and interventions for this condition.

Of course, none of this could be made possible without the generous support of our sponsors. And I'm delighted to say that each of our sponsoring organizations will be represented at today's meeting and discussions.

And those sponsors are Eli Lilly and Company, Regeneron Pharmaceuticals, Sanofi, Novartis, GlaxoSmithKline, Myos Corporation and Nutricia Advanced Medical Nutrition. And I thank all of you and appreciate your having representatives here today.

It's also the mission of the AIM Coalition to assist governments and public policy to prepare the way for more effective treatment and prevention of sarcopenia across the population. Therefore, we are delighted to have among us representatives, the key people actually from the National Institutes of Health including the National Institute on Aging and the National Cancer Institute and from the Food and Drug Administration and from the Foundation for NIH. So I welcome all of our representatives from government and look forward to your active participation.

This Coalition's strength goes to the science that is brought together by our Science Advisory Board, the people that have come to us from academia and from industry to really keep their eye on what is happening from the research community and help bringing that forward to inform the values and the policy goals of the Coalition. And the goal of that Science Advisory Board, whom I'd like to recognize now as well as the members that are with us, the chair is Dr. Jack Guralnik of the University of Maryland and an alumni of the National Institute on Aging, Jack Guralnik, thank you for being here.

[Applause]

Dan Perry: And the others, who are with us on the Science Advisory Board today are: Dr. Bill Evans, formerly of the University of Arkansas and other academic postings and currently with GlaxoSmithKline;

Dr. Stephanie Studenski, formerly with the divisions of geriatrics at the University of Pittsburgh and now heading a very important aspect of the National Institute on Aging, and the well-known Baltimore Longitudinal Study on Aging; Dr. Ronenn Roubenoff, head of Global Translational Medicine in Musculoskeletal Diseases at Novartis Institute for Biomedical Research; and Dr. Carl Morris, Senior Principal Scientist at Pfizer Incorporated. And I ask you to give a round of applause to the members of our Science Advisory Board.

[Applause]

Dan Perry: So I want to thank everyone for being here, thank the sponsors, thank the science advisors, thank the government representatives.

And, now, to say just a word about the issue of patient-reported outcomes and its potential to serve as a measure and possibly an endpoint or part of endpoints for sarcopenia drug development and therapeutic development, I'd like to ask my colleague, Cynthia Bens – she is the Vice President for Public Policy at the Alliance for Aging Research. And Cynthia has been the spine, the bones, the essence of leadership at the staff level behind the AIM Coalition. Cynthia, do you want to say just a few words about today's meeting?

Cynthia Bens: Great. Thanks, Dan. And I know what you're all thinking, oh, no, it's too short a meeting for unscheduled speakers, so I'm going to try not to take up too much of your time.

Dan, actually, asked me yesterday to help him set the stage for the meeting today, and so it's really my pleasure to do that. And I'm just going to have three quick points to make.

The first point I'd like to make is I think we all need to acknowledge that there's growing consensus about what sarcopenia is, but we need to go into this discussion realizing that there's not really an accepted definition at this point.

And I was at a very important last week that the FDA held. Donald Patrick and Ashley were a part of that. And it really became clear to me that FDA really views any discussion about measures for clinical trials through the lens of the defined diseases they apply to.

So regardless of what positive points may come out of this meeting on how PROs are going to apply to potential clinical trials for

sarcopenia, I think we really do have some work to do in supporting those measures with clearly defined clinical definition. And we're going to hear a bit about progress that we're making toward that that Stephanie and Maria have been leading the charge on.

Second, I think we need to recognize that there is an affected population while we don't know necessarily what the population is at this point. Because of the extent of their muscle loss, they're suffering. They have poor quality of life. They have a harder time recovering from injury and they have a tougher time managing the comorbidities they get as they grow older.

So that population is going to continue to grow for all the reasons Dan mentioned earlier, because of the aging of the baby boom population, the aging of populations globally, so this is something that we need to really attack with urgency. And I think we need to get past of the point of discussion about defining the disease and deciding what endpoints to put in the trials and just make some decisions on the best science we have available to move forward.

And, lastly, I'd just like to say that up until this point, development from industry has really been at the earliest stage, but we're really fortunate that a lot of companies are developing later stage treatments or moving into later stages of the development for sarcopenia. And, particularly, there's a lot of momentum in Europe that we'd like to see momentum building here as well.

And it's not going to be an easy road ahead for these companies. I'll be honest. From what I know from the conferences I've been going to lately, there is still a lot of uncertainty. And so we really applaud you for taking the lead to be doing what you're doing.

But we're all here to provide you with some help to get it right. We're hopeful that we're going to get some help and advice from the FDA today, that we're going to get some help from the academic community, but, mostly, I think that you need to look at patients as a resource and how they can provide you with some help in defining what your outcomes are going to be for treatments.

And so Dan asked me what I would like to see as the goals for the meeting. And so what I'd really like to see is that we all leave here with a clearer sense of how FDA views patient-reported outcomes for sarcopenia as they currently stand and how we can really move to a point to where we're clear on how they can be employed in

clinical trials. And so those are my two major goals for the meeting.

And I encourage you all, if you came here with different goals for the meeting, to articulate that during Q&A. Jack, who was really intent on keeping this discussion focused, he and I will know the appropriate times to break out our hook if you're taking us too far off track. But I encourage you to make this a real dialogue, because ultimately this is really here to serve as a discussion to kick off what we hope is a really successful partnership.

And I want to mention a few housekeeping. I have the unfortunate pleasure of breaking the news to some of you, who had the intention of having Internet access. We do not have Wi-Fi in the room, so, hopefully, we can separate you from your email and your offices for the next four hours and have you really participate in the discussion.

And I'd also like to let you know that the discussion is being recorded and we're going to have a transcript that's going to be posted on the AIM website probably within the next three weeks.

And we also have a science writer here. She's sitting right in front of me. And she's going to be preparing a summary that we'll also post to the AIM website. So if there are any colleagues that weren't able to make it here today and they'd like access to it, we'll be sure that they're able to get a copy of that.

So, with that, I'm also going to join Dan in thanking you for being here. I know that some of you have come from great lengths to be here, particularly Erwin Meijer from Nutricia, who made the trip yesterday from Amsterdam. But it really shows that this is an important discussion and, hopefully, we'll be able to make it worth your while for coming.

And Dr. Maria Vassileva from the FNIH is going to be our first speaker. And so I'd like to introduce Maria up to talk about the work she's been doing to –

Dan Perry: I still have a couple of slides I would like to go through.

Cynthia Bens: You do?

Dan Perry: I do.

Cynthia Bens: Oh, Dan *[laughs]*, sorry about that. All right.

Dan Perry: And then we'll introduce Dr. Vassileva.

Cynthia Bens: Great. Thank you all.

Dan Perry: Thank you. Thank you, Cynthia.

[Applause]

Dan Perry: Thanks very much and I think setting those goals at the outset is very wise.

I just want to take just five minutes to stand back and look at the biggest picture, the widest context for the issues we're going to be talking about today.

Clearly, the unprecedented experience of population aging across the globe is something that no one forecast a hundred years ago when we started developing antibiotics and started cleaning up the air and the water supply and started working indoors, but that's what's brought us here. And our attitudes and expectations about what old age would be have gone through some radical changes in a historically short period of time.

In the middle of the 19th Century, which this cozy photograph, or not photograph, cozy image indicates, old age was a gift from God bestowed upon those who had lived a virtuous and a righteous life. And it was a time of great comfort and it was a very short time and very few people got to experience it.

So, now, let's fast forward in an American context, a hundred years, the United States has emerged from World War II. We are a superpower. We can do anything. And we've got ambitions for a great, long life with great vitality.

And the next image will show the cover of a very popular self-help book that captures all of that, but you may notice that there's a distinct limitation in the view of long life. In this case, we have this hip, svelte-looking couple striding bravely into the future, aware that, of course, at 40 their lives are pretty much washed up. They're no longer so attractive. They don't really have a future in their career. They're kind of sidelined. But this author will tell you the 10 steps of the power of positive thinking. And with all of that, you just might be able to eke out another 10 years or so. So that was the way we looked at this in the 1940s.

Needless to say, we've got a bit broader view today. I will admit this is driven by Madison Avenue and a certain senior citizen organization that kind of promotes this idea that we're all going to be leaping nonagenarian athletes, but this is at least the way we look forward to it.

And it all started with this lady, Kathleen Casey-Kirschling, born a few minutes after midnight, December 31, 1945, and into January 1, the very first baby boomer. That person turned 65 a few seconds after midnight December 31, 2010.

And beginning in 2011, the United States population made a significant shift. We moved from what had been the pattern for decades of producing about 6,000 new 65 year olds every day in the United States. We went from 6,000 a day to 10,000 a day with the leading edge of the baby boom generation. That was in 2011, 10,000 a day. In 2012, 10,000 a day – I could go through this, but it goes to 2029, we will be producing about 10,000 65 year olds every day. That's the wave that we've all been hearing about.

Now, 65 is not exactly a geriatric case. But with that doubling of the 65 and older population, we are also going to experience a near quadrupling of the 85 and older population. And with 85 and older folks, we really do start to get into multiple chronic diseases, geriatric syndrome, significant burden of health care, significant costs of health care. And right where it turns red from blue is about where we are now and that's in millions of 85 plus. And you can see where we're headed by the middle of this century.

And, as I said, with that is – tracks with a significant and really unprecedented increase in chronic, disabling, life-limiting, life-shortening diseases associated with the underlying biology of aging.

And to get this right, which is what we're all trying to do today in terms of sarcopenia and functional decline, its physiologic and its nutritional aspects is the goal of wise people everywhere, who see that as Robert Hormats, who is our Under Secretary of State for Economic Development, has said, "The countries that get this right will be the countries that succeed, lead and dominate in the 21st Century." It is no small thing that we are doing today.

Now, the Alliance for Aging Research waded into this area years ago. And in 2006, we put together and today still lead a coalition of multiple caregiver organizations, patient groups, scientific organizations to meet with the FDA on a regular basis with its

neurology branch to talk about what are the barriers in drug development for Alzheimer's Disease. I won't say any more about Alzheimer's Disease. I don't think you need to hear that. But this was the model for what has now been carried over beginning in 2011 into the area of sarcopenia, muscle wasting, functional decline. Some would say frailty. And so that was the essence of the creation of the coalition that now have brought us together today.

And here is our board of scientific luminaries that guide our every step. And we've recognized those that are with us today. And here are the organizations that comprise the Aging in Motion Coalition. And you should know them, because we are your host today. And we look forward to the outcomes of your discussion.

And, with that, I thank you again. And we'd like to invite Maria Vassileva. And I'm sorry, Maria, for stalling this for a moment. She's with the Foundation for the NIH and has been working on an initiative to get us an accepted definition of sarcopenia for some time. And we look forward to getting an update on that, Maria.

[Applause]

Maria Vassileva: Thank you, Dan, for the kind introduction. Thanks to the organizers from the Alliance for Aging Research and the AIM Coalition for giving me this opportunity to present on the work of the Sarcopenia Project Team managed by the Foundation for the NIH Biomarkers Consortium.

I had the daunting task of not only following the person of the Alliance for Aging Research in being the first speaker today, but also presenting in front of the project teams here, Dr. Stephanie Studenski, and one of the principal investigators on the project. So, hopefully, I get this right. And if I don't, they'll be correcting me.

And so we really appreciate this opportunity, because it's with great pleasure that we can announce today that we are finally ready to publish the results of the last three years of work of this project that I will be telling you about today.

But before I begin talking about the project, I just wanted to say just a couple of words about the Foundation for the NIH so you can understand why this public-private partnership was housed

within this institution and why this was a good neutral convener to enable this kind of an initiative.

So the Foundation for the NIH is an independent, non-profit organization. And it was authorized by an Act of Congress about 17 years ago now, and since then it has been able to raise $560 million and has been able to support over 400 projects. And so we're very proud that 94 cents of every $1.00 that we raise goes toward program development, which makes us a very lean organization, but also very efficient.

And the way we perform our function is by being a neutral convener between government and the private sector involving a lot of disease-focused foundations as well as the prominent key opinion leaders from academia in formulating and executing public-private partnerships.

This is just a snapshot of our website. And since we have such a diverse audience here today, I thought I'd just give you an idea of some of the partnerships over the last ten years that the Foundation for the NIH has managed. And the project I'm going to talk about today, of course, falls, as I said, within the Biomarkers Consortium, which is just one of the many partnerships of the Foundation for the NIH.

And the most recent one, which you may have seen in the news several times last month, is the Accelerating Medicines Partnership. It's a new initiative led by the Office of the Director at NIH in close collaboration with ten pharmaceutical companies and it focuses on target validation. And a lot of the companies represented in the room today are involved in this initiative, which is why I wanted to list it on here.

But, as I said, the Sarcopenia Project falls within the Biomarkers Consortium. And the Biomarkers Consortium is a public-private partnership of the Foundation for the NIH that was launched in 2007 with the idea that we need a lot of exchange of knowledge and expertise between industry, primarily the pharmaceutical industry, academia and government leaders in the area of biomarker development and qualification.

And so, once again, the Foundation for the NIH was able to facilitate cross-sector partnerships in a pre-competitive space, which means not one partner benefits from the interactions. All scientific results are made publicly available and we address unmet medical needs that cannot be addressed anywhere else other than in

a pre-competitive space and via partnership. And, hopefully, the idea at that point was that the results we produced through our collaborative projects would really be able to affect the lives of patients and affect regulatory decision making in clinical practice.

So the reason I put this slide out there is just to show how the Sarcopenia Project was a perfect fit toward this model, and that's why it was housed within the Biomarkers Consortium. And it was actually managed by the Metabolic Disorders Steering Committee of the Consortium, which, in addition to the Sarcopenia Project, manages a lot of other projects.

And, as I said, Dr. Stephanie Studenski, back then at the University of Pittsburgh, came to the Foundation for the NIH with an idea. And that idea was further developed by a big team.

And you will see the team listed in the acknowledgements. And, as I said, Dr. Peggy Cawthon is in the audience today as well as Jack Guralnik and Judy Hannah, other project team members.

So the problem statement, as we already heard some of this from Dan, is that loss of muscle mass is common in aging and wasting conditions and is frequently associated with weakness, poor function and lower survival. Despite the fact that this is such an important clinical condition, it is currently unrecognized.

And, as you heard already from the first discussions, there are multiple potential interventions being developed by drug companies that aim to treat or prevent muscle loss, but there are not necessarily any regulatory guidelines about which measures can be looked at when you evaluate the effectiveness of those interventions.

So the field is really in need of a clinical definition in order to proceed with a specific regulatory formulation. And, of course, as we heard from Cynthia, one of the reasons for today's meeting is to address some of those issues in the area of patient-reported outcome development.

It is a huge unmet medical need. Currently, we're unable to identify patients that require treatment. And we really need clear and valid diagnostic criteria and outcome measures in order to fulfill the regulatory demand and in order to support the investments that the drug companies are making in testing the interventions.

As you heard, there is a growing number of older adults throughout the world. In the U.S., it's predicted that by the year 2050 we're going to have 86.7 million adults over the age of 65 years. And, as Dan already commented, that fact is expected to result in a huge economic burden, because the growing population of older adults is going to be experiencing a lot of comorbidities and there's going to be an increased need for institutionalization, so we really need to address this issue.

And so the Sarcopenia Project was launched in 2010 with the goal to create an evidence-based definition and, as I said, to address the need for a clear and valid diagnostic criteria and outcome measures that could be acceptable to clinicians, regulators and health insurance companies.

That development of an evidence-based definition would provide opportunities to develop and test the potential interventions that are currently in the pipeline addressing low muscle mass and strength and, ultimately, improve the health of older adults.

So I just wanted to mention that, as you know, when we started out the work in 2010 and shortly thereafter, there were several other definitions of sarcopenia, and we have tried to summarize them here in chronological order. There's no other meaning behind the listing. So those four different definitions that came out in 2010 and 2011. And I just wanted to point out that the major difference between those and our definition are that those were based on the opinions of experts while our definition is based on data analysis. And you will see how they compare with each other.

But when we started out our work, the one question that remained is when is low lean mass empirically grounded to its relationship of strength and function? And that's the question we set out to address.

And this Dr. Studenski's slide that represents the current clinical paradigm that when a patient presents with poor physical function, we're not sure if it is caused by weakness and then we have to look at whether the weakness is caused by low muscle mass or by other causes of poor performance. And then we have two options. One is that the low muscle mass is the possible cause of the weakness. And the other one is that we really could have other non-mass causes of this poor function.

And that's why in the Sarcopenia Project we decided to use cross-sectional analyses and pull prospective data from several

aging studies in order to evaluate criteria for sarcopenia diagnosis by using shared operational definitions of performance, strength and body composition.

Then, also, as a part of the Sarcopenia Project, we wanted to present some of the results of the preliminary analyses to a broad professional audience in order to receive feedback and recommendations. And we organized a big consensus meeting in May of 2012, so that's almost 2 years ago, in Baltimore, Maryland, where we invited not only all the stakeholders in the consortium, but also all the health organizations and any other academics from all over the world that are in this space in order for them to see some of our preliminary results and weigh in on some of the findings and tell us what they think some of the next steps are.

And so, in the papers that will be coming out in the next two weeks with the results of our work, we also have a summary of this consensus meeting and the recommendations that came out of it.

Here's a list of all the studies that participated in our project. We had a request for proposals in which we outlined what kind of measurements we wanted certain studies to have. And then a lot of investigators applied suggesting to analyze some of their data. And so the MrOs study was represented by – Peggy is the principal investigator. The Rancho Bernardo study was represented by Dr. Tien Dam from Columbia University. And we also had six clinical trials from the University of Connecticut. Most of the other studies are observational as you very well know. Dr. Alley from University of Maryland was the principal investigator bringing to us the In Chianti study. Dr. Bob McLean from Harvard brought to us both the Framingham original and offspring cohorts. And then from NIA, Intramural Program, we had Dr. Tammy Harris that brought to us the Health ABC data and the AGES data.

And as you'll see from this table, and I've tried to summarize a little bit on the next slide, is the strength of pulling over 26,000 patients' worth of data is that it was a very large, diverse and well characterized set of population that we used for our analyses.

57 percent were female; 43 percent were male. And the median age was 78.6 years for the female population and 75.2 years for the male population. So we had both genders. We had a diverse range of ethnicity. Not as large as one might want. There was definitely African-American population in the Health ABC study. And then, of course, the Boston Puerto Rican study was primarily Puerto Rican. Also, we had multiple geographic regions including some

European populations and a range of health and functional states. However, you will see in some of my conclusions that there were certain limitations of the dataset which we are aware of.

We also, at the outset, had an explicit conceptual framework and did really extensive sensitivity and cross-validation analysis. And the reason we could do this range of sensitivity and supplementary analysis is because we could look at alternate measures, the physical function, strength and body composition, and evaluate whether the findings would differ substantially if we had chosen different cut points and measures and that really helped us confirm and strengthen our conclusions.

And, finally, based on extensive work done by Dr. Studenski and others, we could set our primary indicator gait speed less than 0.8 meters per second. And we chose that because it is associated with reduced survival and increased disability, which has been widely accepted.

And so, as I said, this work resulted, after the pulling of all the data and the analyses and setting some preliminary cut points, and the publication of five different manuscripts that are coming out in April in the *Journal of Gerontology and Medical Sciences*. As a part of the papers, we also included the recommendations from the consensus conference that I mentioned took place two years ago.

And just to give you kind of a very short summary for the sake of time of what we were able to do is we were able establish criteria for two important components of the condition. And here are our candidate criteria. For men, grip strength of less than 26 kilograms and an index of less than 0.789 for appendicular lean mass adjusted for body mass. And for women, grip strength of less than 16 kilograms and an index of less than 0.512 for appendicular lean mass adjusted for body mass.

Based on this consensus, we believe that now further biomarker analysis will become possible and they should be enabling for future clinical trials.

In addition to this presentation and the publication of the papers, there's going to be a series of public announcements and then a webinar organized by NIA. FNIH is planning a press release to coincide with the publication of the manuscripts online in April. And we've already presented those findings at the ASBMR annual meeting and as innovator presentation at the FasterCure meetings in New York.

So we're really trying to fulfill the mission and really disseminate those findings. So we appreciate the opportunity to present at this meeting as well.

We do understand that a lot more work remains to be done. And the main part that we see as the next steps is the confirmation of those established criteria in populations with more severe mobility limitation. That was something that came out as a recommendation of the consensus meeting that we held in Baltimore in 2012.

But after that, additional things have come out as possible next steps. As was mentioned, in an aging population, there are a number of comorbidities that develop simultaneously. And it would be great to be able to start to address how we can take into consideration those comorbidities as we develop our measures for sarcopenia.

And, also, there are two types of DXA machines that are typically used, Hologic and Lunar, and we really have to account for some of the possible differences or to say this in reverse is to prove the reproducibility of measurements done by the different machines. And that's something that still remains to be done as a project.

So here are some ideas that our group has been discussing for possible next steps for possible partnerships and using the momentum of this completed work and taking it to the next level.

We could look to validate and confirm the predicted validity of our candidate criteria and, as I said, through analysis in populations with higher levels of mobility limitation and also in people with muscle wasting disorders.

And, also, a typical project that we frequently do in the Biomarkers Consortium is look at the reproducibility of a given measure. So we could certainly look to establish the reproducibility of repeated measures of body composition.

And something that's of interest to everyone who develops interventions in this field is, of course, demonstrating how changes in muscle mass really affect function, because that's the key question that still has to be addressed today.

As I said, none of this work would have been possible without the very dedicated team and this is the core team here listed led by Dr. Studenski, who is sitting in the audience today.

It also wouldn't have been possible by very generous support provided by one nutrition company, four pharmaceutical companies. And, very importantly, the National Institute of Aging, they provided us with a U13 grant that enabled the support of the consensus meeting two years ago. And, also, that same meeting was supported partially by funds by the Food and Drug Administration.

I'd like to say that this is one of the teams that is always given as an example of how people from different sectors should communicate and work toward a common goal, people who are very dedicated to achieving the ultimate goal. And, therefore, although they might speak different languages and have different interim goals, they really put their expertise and efforts together and made this work possible.

And since many of you are out of town and you may have come expecting to see cherry blossoms today, I'm sorry, they're not out yet, but they will be soon. So, at least, there was no snow today. That's why I thought I'd finish with this picture and let's hope for better weather this weekend.

So thank you very much and now I'll take any questions. And, as I said, Peggy and Stephanie will also – and Jack and Judy might be able to.

[Applause]

Bill Evans: That was great. Bill Evans at GSK. And it's a fantastic effort. I think it'll make a tremendous contribution.

When we first started thinking about sarcopenia, I think that we had the view that the amount of muscle would ultimately predict some distal outcome, and I think that that's the thinking still. And in the period of time we know that muscle now has a lot of other things. It contributes metabolically to a number of different conditions. And we now know that the quality of muscle changes with age. So even for the same amount of muscle, you may get radically different amounts of force production.

And so to one extent, it complicates things tremendously, but on the other hand it allows us to explore lots of other different possibilities for affecting muscle.

And I'm wondering how the committee really thought about the contribution that muscle has not just to the obvious, which is how strong you are and function, but how it contributes to diabetes and how it contributes to obesity and how the quality of muscle ultimately changes and contributes to poor function?

Maria Vassileva: Yeah. I think Stephanie is going to address that question.

Stephanie Studenski: So, thank you, Bill, those are really, really important issues. I think –

A couple of pieces of this. One, as you know, the wonderful thing about science is it moves forward and grants get written at a point in time and then you're trying to work with changing thinking. So this proposal was written in 2008-2009. It took a year to get through the FNIH. And so to some extent, we are hoping to be part of building toward information and evidence for further thinking.

A second point would be that it became very clear to us, and Peggy was a key part of this, is that we need to be careful about terminology and that there's a lot of dynamics in the room as well as around the world about should we keep using this word sarcopenia and what it is? And we talk about that some in the papers.

And so we chose to say all we're talking about here is a definition of what we think is clinically significant weakness and clinically significant low lean mass. And you guys can decide what that is.

I think there may be decisions to use terminology about poor muscle quality and poor muscle mass in a variety of different ways. And I think it's important for this group to participate in that.

A third thing is we couldn't agree with you more that the impact of low muscle mass in its roles other than function is very important. It's one of the reasons the consensus conference, that I think you were at, said we need to look at how these kinds of criteria play out in different populations and other – one might be a diabetic population. One might think about cancer populations. And so the interface between low mass, low strength and important conditions

is an unexplored space that was a priority for an important new area to study.

I think as a clinician one of the things I'm comfortable with, because, you know, I felt like, well, maybe we just wasted our time doing this whole damn thing, because everybody changed their minds about what is sarcopenia. As a clinician, what I can tell you is there clearly is a population of older people who are shrunken and weak. And I think you can be weak for other reasons besides being shrunken. I think sarcopenic obesity and poor muscle quality are important issues.

But I thought from a point of view of identifying a target population, I think there is a face validity to the little, scrawny, 85-year-old lady with pencil legs who can hardly move. And there may be other very important weak, abnormal muscle group populations that should be targeted for interventions. But I think it is a real population.

And I think what we're saying is it may not be as big as everyone hoped. When you first did those epidemiologic studies and said, you know, 30 percent of old people are going to have this. Just think of all the profits for drug development. It's not that big. The population who are shrunken and weak is real, but I think possibly even a minority of people who are weak. I'm not sure yet.

Ronenn Roubenoff: So from the physiological sort of philosophical to very technical questions, Ronenn Roubenoff, Novartis. This ASM sub BMI thing, so this is now appendicular skeletal muscle adjusted for BMI, not for height squared. Is that correct?

Stephanie Studenski: That is correct. And that came out of the consensus conference. Again, I think you were there, but there was a great deal of concern that we went into it. And Peggy is here and Peggy can tell you about all the ways we tortured her, because she was in charge of the lean mass analyses, and they were by far the hardest to perform.

But the consensus conference was seriously concerned about the effect of body weight and fat on how we define these things. And to some extent, just the lean mass divided by height never accounted for fat or weight. And so what we were asked to do was to use our data to evaluate different approaches, developing cut points.

And I don't know if you have other slides with you, but what was striking was there's a big gender difference. And so, in men, and there's just this wonderful slide where you can see these huge differences in how these various adjustments play out for men and women. The big picture being, as I understand it, you end up with about the same level of explanatory power in men anyway you do it, but you end up with really different populations in women.

And my sense is and I've said this to a couple of people – I don't know how they feel about it – is to some extent men seem to have the capacity to maintain or increase muscle when they get fatter and heavier whereas women don't. So, women just build more fat when they gain weight and don't carry it with any more muscle. And it's not that men get really strong, but they're able to compensate for their increased weight by getting bigger thighs or whatever.

And so the adjusted and unadjusted lean mass estimate explanatory power is very different in men and women. And we have in the consensus paper – I think we're going to have to be very thoughtful about gender and whether we should be using the same or different criteria in men or women, because I think muscle biology is very affected by men hormones and stuff.

Ronenn Roubenoff: This is just a simple division? This is an actual adjustment? This is not a simple division by lean mass?

Peggy Cawthon: It is a simple division.

Stephanie Studenski: And, again, this came out of the CARTs. Peggy, do you want to speak any more to some of the analytic issues that you faced, because again it's not like we just threw in lean mass. And there were about 30 candidate measures that were evaluated in the CART and many ways of trying to look at this fat mass was in there, all kinds of stuff. So it wasn't arbitrary.

Peggy Cawthon: The way CART works is that it's a somewhat nastic approach where we just considered a number of different ways of classifying lean mass including the traditional way that had been initially suggested by Baumgartner. We put all of these different variables in the CART model. We included BMI and weight just alone, because we were concerned that, perhaps, the cut points for lean mass would differ by BMI, so there'd be an interaction with BMI. And, routinely, BMI was not as important as lean mass for distinguishing people who are weak.

And the variable that came up, aside from just lean mass alone unadjusted for anything, was this appendicular lean mass divided by BMI measure, which we had included initially sort of as a measure being complete. We had lean mass divided by height squared. We had lean mass divided by weight. We had other measures of BMI. We said, well, we should just divide it by BMI and see what happens and that's how it ended up arising in the models.

Stephanie Studenski: And am I right that it was about the same in men whether you did it or not, but it was different in women?

Peggy Cawthon: There was still some differential classification based on appendicular lean mass or appendicular lean mass divided by BMI in men, but it was much less pronounced than in woman. And it was about the same prevalence of having quote low lean mass using the two different cut points in men, but it was quite different in women.

Ronenn Roubenoff: So the second question, your working group, it's a flow chart of gait speed and then grip strength and then DXA. Are you proposing the same kind of thing or is this a –

Stephanie Studenski: Yeah. I'd like to be clear that we were very much indebted to the thinking of the European Working Group and the other groups that went before us. And we think this foundation of saying it's not just everybody with low mass, the way it started, that you have to identify a population of functional problems.

And then our thinking was to do the two steps that – the way clinicians think was my argument – a person presents with mobility complaints or physical performance problems and as a differential diagnosis as a doctor I say what are the possible causes of mobility problems? Well, one is that they're weak. So I'm going to evaluate if they're weak. And if they're weak, what are the possible causes of weakness? That's how we think as a doctor. One of them is low muscle mass.

So that's where the sequential thinking and I think there are people who have different senses of the relationship between weakness, low muscle mass and performance.

I'm concerned that weakness and performance are not necessarily equivalent, so I'm uncomfortable having only two of the three and the weight. I think other people have felt maybe some of these –

you know, you could have just two of the three, and I'm a little less comfortable about that.

So it was based on developing a clinical diagnostic conceptual framework which is what we started with is why we have the three.

The only other thing I'd like to say, I don't think it's come up yet, is when the cut points were developed for strength, there were two levels of cut points. The first was 20 in women kilograms, 30 in men, which is just like the European. But there was a second level cut point that was quite discriminatory, which is where the 16 and the 26 come from.

We said because we think that there is some fear out there on the regulatory world, policy world that all the old people are going to get diagnosed, that we're going to say we're going to be really conservative here so that what you have is a low false positive rate. And so we couldn't agree more that we may be missing people who are important to treat. But we're trying to say if you find people that are this weak and this low mass, you've got a really high likelihood that you've got weak, low-mass people and there aren't false positive. And we thought the FDA might like that.

But if I had it to do over – we ran out of time and money – I would have looked at cut points for the 20 and 30 and seen if they were closer to the European group, because I think that may be what you'll see in the last paper.

There's a fair amount of discrepancy between anybody's definition and anybody else's. So the agreement rates are two percent between different definitions except I think there's one that's a little better than that, but they're terrible. They're terrible.

And I think we could have been a little more consonant with the European if we'd done the lean masses at the less conservative strength cut points.

Dan Perry: Other questions for either Dr. Studenski or Maria?

Well, thank you very much, Maria, for that presentation. And Dr. Studenski and all who joined in, thank you all very much.

Before I introduce our next speaker, I want to point out another very important person in the room, and that's Sue Peschin, who is the chief executive officer of the Alliance for Aging Research, sitting over next to the wall. If any of you want to know more

about the Alliance or about the AIM Coalition or be more involved, in addition to Cynthia and myself, please introduce yourself to Sue Peschin, a very important person at the Alliance. Thank you.

Our next speaker previously recognized the chair of our Science Advisory Board. I've known Dr. Guralnik for probably close to 30 years going back. In fact, Jack, I found a photo of you at a press conference in the U.S. Capitol that I was going to bring. I'm sorry. I forgot it. There's this very handsome guy telling members of congress why the aging of America is going to pose significant health financing challenges in years to come. And guess what? He turned out to be right, so congratulations, Jack.

But Jack was with the National Institute on Aging- really the person to talk to about exercise physiology, muscle function for a long part of his career, now with the University of Maryland School of Medical Sciences.

So, Jack, and thank you for your leadership.

Jack Guralnik: It's my pleasure. It was kind of a no-brainer even 30 years ago to see what was coming when you looked at the demographics, but – But we still need to keep telling congress about that.

It's a pleasure to be here. I want to thank Dan and Cynthia Bens for putting together the AIM Coalition. I think it's a really important project. And I think it's going to contribute a lot in the aging world.

We're here today to talk about patient-reported outcomes. So why am I standing up here talking about performance measures? Well, the organizers of the meeting thought it would be of interest to look at how performance and self-report relate to each other. And, also, they asked me to talk a little bit about the FDA qualification process, which the AIM Project is sponsoring our work to try to get a couple of performance measures actually qualified for use in trials related to sarcopenia.

So this issue of performance versus self-report, we've been talking about it for quite a while. We've been at it for 20 years or so since we started doing performance measures in epidemiologic studies, but the use of performance goes back even further than that.

And there's an interesting historical perspective on this. There's actually a performance test that was done in one location on

literally millions of people. And I won't make this a quiz, but I'll show you what that test was.

This is the Great Hall at Ellis Island. And this hall was on the second floor there. And as people entered on the ground floor, they had to come up this set of steps. And the public health service doctors stood at the top of the set of steps and watched people do this performance test of climbing these stairs and screened people on their performance and decided who needed further follow up in terms of their health status.

And this is clearly an example where self-report is not going to work. You ask these people, how's your health? They all say good, because they wanted to get into America. So the performance was quite valuable there.

Now, to think about the relationship of self-report and performance, I think it's very helpful to use a framework related to the path from disease to disability. The WHO had a framework. They started early on. Another one was proposed by the researcher Nagi, picked up by the Institute of Medicine. As many of you know, WHO has come up with another pathway that those of us doing kind of empirical research have not found that useful. So I'm going to focus on the Nagi Pathway, because I think it does help in this discussion.

So these are the definitions. Verbrugge, Jette made some modifications to this and came up with these definitions starting with pathology going to impairments, which are dysfunction, and structural abnormalities and specific body systems then to functional limitations, restrictions in basic physical and mental actions. I think of functional limitations kind of as the building blocks of functioning. And then, finally, on to disability, which is difficulty in activities of daily life, not just activities of daily living, but many different activities including household management, jobs, hobbies, that sort of thing.

This is an example of how we might proceed through this pathway related to the kinds of issues we're talking about today. Starting with a chronic disease or an injury leading to atrophy of muscle, which we can pick up as low grip strength, perhaps, and that's an impairment. Moving on to a functional limitation, which can be slow gait speed, as we've just heard about, or self-report of function, such as are you able to walk a hundred yards or a block or some of these questions we ask. And then, finally, on to disability, which is usual household activities and social activities.

So, in the IOM report back in '91, they asked Nagi, who had originally developed this pathway, to elaborate a little bit on the difference between these steps in the pathway. And I think this is useful.

He gave these definitions, which are similar to the Verbrugge, Nagi definitions, and made the point that both impairment and functional limitation do involve – functional impairment refers to the level of tissues, organs and systems, kind of physiologic function; whereas, functional limitation is in reference to the level of the person as a whole, not one organ system.

And then looking at the differences between functional limitation and disability, functional limitation is a limitation in performance at the level of the whole organism; whereas, disability is limitation in performing socially-defined roles, tasks within a social, cultural and physical environment. So functional limitation refers to an organismic – I got be careful with that word – an organismic performance. And disability refers to social performance; whereas, disability is a relational concept. Whereas, the other three stages really are concepts of attributes.

We use functional limitation as an outcome in many of our studies to study the effects of diseases and impairment on functional outcome. It's a more proximal measure than disability. We use it to evaluate functional consequences of a risk factor or an intervention. We can observe the effects free of environmental influences. It's got excellent psychometric properties. But there's still an advantage to stopping at that low, that step of functional limitation on the pathway. And the changes in functional limitations can be difficult to interpret in relation to the problems in daily life. And, ultimately, if you want to know if someone can go shopping and prepare meals, you don't do a performance test. You ask them or observe them doing those kinds of things that represent true disability.

Now, this is a key point here. And it's when we first got into using performance measures, we felt like they were really going to be measures of functional limitation primarily and that disability was going to be represented by self-report, and that really is not the case.

Functional limitations and physical disability refer to different behaviors not to different ways of measuring the same behavior. Thus, you can measure functional limitations and disabilities using

either subjective or objective measures. So it's not objective equals function and subjective equals disability. They can go both ways.

And, in fact, we're going to hear a lot about a proposed patient-reported outcome battery for sarcopenia. And when you look at the items, many of those items are not disability. They're really closer to functional limitations when you talk about specific walking questions. So that battery itself probably bridges both functional limitations and disability.

So I will be in talking about performance measures describing some results with the short physical performance battery. Many of you know this battery was developed at NIA in the late '80s. It uses three components, hierarchical balance tests in the three positions of the feet that you see here, a timed four-meter walk and timed chair stands done five times. I'll also talk about the four-meter walk alone.

And, again, focusing on relationship between self-report and disability. This is some early work we did where we stratified the population according to ADLs and mobility. And the mobility question we asked was can you walk – in this case, it was asked as a half mile. We generally use quarter mile in questions now. And can you climb a flight of stairs? And you see the population is divided into those who were disabled in both, those who were disabled only in mobility and those who were not disabled in either of these.

And we're looking at the SPPB score on the bottom. This score ranges from 0 to 12 with 12 being the highest functioning. And you can see to start out that the group that is disabled in both ADLs and mobility is very much shifted to the lower SPPB scores. Those who have neither disability have the higher scores. And those with mobility only, and this is a hierarchical scale, so the middle group that has some mobility problems but no ADL problems is really spread pretty much across the whole spectrum. So, within these categories of self-report, you see a fairly wide distribution of objective performance measures. So they're related, but not terribly highly correlated here.

We see this again here in work that Luigi Ferrucci and I did looking at the distribution of the SPPB score and then stratifying it in the dark area by those who have mobility disability and in the light area those who do not. And you can see that among those who do not, there's a pretty wide distribution of SPPB scores.

Now, within this large part of the population, we're not discriminating at least with the self-report measures that we were using.

This doesn't mean that we and you folks can't develop patient-reported outcomes that do discriminate in this whole range the same way that the performance measures do, but that's really I think where the challenge in the self-reported measures are is to start tapping into the higher end of the functional spectrum that the performance measures do a good job tapping into.

These were some interesting analyses that looked at the mean SPPB score in specific subgroups according to self-reported functioning and looked at it by age. So the figure on the left are those who reported no help in ADLs, climbing stairs or walking a half mile, and so they should look functionally the same. But when you look at their performance by age, it goes steadily down. And this means there's something else going on with their functioning that's being picked up with the objective performance measure. It's not being picked up, at least, by these questions. And, again, if you're going to see no age effect with the self-report measures, you're going to have to get much more refined measures that explain some of the changes with age that are being picked up with the performance testing.

And then this is another subgroup, those who need no help in ADLs or climbing stairs, but are unable to walk a half mile. And, again, you see that kind of age gradient in the SPPB even in this very specific group by self-report.

So in order for performance measures or any measures, the PRO measures we're going to talk about today, to be useful, they have to be responsive to clinical events and important clinical changes and they also have to predict subsequently adverse outcomes. So I could spend the whole time talking about data for those kinds of relationships, but I'll just show a couple of examples.

These data come from the Women's Health and Aging Study, a study that we at NIA did in collaboration with Linda Fried and her colleagues at Johns Hopkins. We looked at in this study women with mild to moderate, moderate to severe disability, who were living in the community, to try to understand the natural history of their changes in disability.

Because these women had a lot going on, we actually assessed them every six months. And we looked at the women who had had

a major event within a six-month interval and then looked at their change in SPPB scores from the beginning of that interval to the end. Now, they may have had their event right at the beginning of the interval or very close to the end of the interval and we couldn't really control for that.

But we do see very substantial declines in the SPPB score according to what kind of event they had. So those who had had a hip fracture at any time in that 6-month period had over a 3 point decline on this 12-point scale. You see lesser declines for stroke, MI and CHF. And those who had no events had only a very modest decline over the 6-month period.

So there's a lot of other data to support the responsiveness of SPPB and gait speed alone.

In terms of subsequent outcomes related to performance, we also have lots of data. This is for the SPPB. And in this case, we started with people who were non-disabled at baseline at the time their SPPB was measured and looked at how it predicted disability outcomes in the future. This shows the percentage with ADL and mobility disability. And then these parts of the bar show those with mobility disability alone and no ADL. And then, finally, those who remained non-disabled.

So strong prediction and these kinds of results demonstrate that even in those who are telling us that they don't have any disability that with performance measures you can pick up differences in health status that are predictive of future disability.

Gait speed has shown similar results and important relationships between gait speed and mortality. Stephanie and many colleagues had a nice paper in *JAMA* in 2011 that showed how gait speed predicts mortality in a very consistent way across many different studies.

Also, predicting outcomes here that are important work done again with data from the established populations for epidemiologic studies of the elderly. Looking at the SPPB score and health-care utilization, we have this graded relationship for percent hospitalized, number of hospitalizations and number of hospital days that again shows how this performance measure is predicting outcomes.

Now, many people have proposed that the performance measures and self-reporter, patient-reported outcomes are really

complementary, that they show slightly different things. And when we look at the theoretical model, we kind of get that idea, but there's been some research that really supports this idea.

And these are analyses I did with David Reuben at UCLA again using data from this epi study. And we looked at the joint effect of baseline self-report and baseline SPPB score. Again, the same kind of hierarchical classification for the self-report where (a) is independent in mobility and ADLs, (b) is dependent in mobility only and independent in ADLs, and (c) is dependent in mobility and one or more ADLs.

And you can see that within those groupings, as you go across the spectrum of SPPB, as the SPPB score goes down, mortality goes up as you would expect. So there's kind of more fine granularity that's being picked up by the SPPB score that's not being reflected by these baseline self-reports.

Now, granted, this is a crude hierarchy of self-reported disability. And a more refined scale would probably do more, but it does seem like the two of them are working together to predict the outcome.

This is work done by Dr. Studenski and her colleagues, Subashan Perera at Pittsburgh. And, again, it shows the joint effects of both gait speed and SPPB with self-reporting. In this case, it shows the results for the baseline global health status. This is the question from the SF-36 scale. It just says how would you rate your health? Excellent, very good, good, fair and poor, a five-point scale.

And in a single model here, both the baseline gait speed and the change in gait speed of 0.1 meters per second or more showed significant associations with mortality. So that for each 0.1 meter per second increase in gait speed, the hazard ratio was 0.89, so there was an 11 percent decrease in mortality for each 0.1 meters per second of gait speed. And then a change in gait speed as a dichotomous variable showed that if gait speed declined 0.1 meters per second or more, there was a twofold increase in mortality.

But at the same time in the same model, you see a change in global health status, which is a significant predictor of mortality. Baseline global health status in this model was not. We see basically the same results for SPPB where both the baseline status and the change are significantly predicting mortality and the global health change is also a significant predictor of mortality. So

they're picking up somewhat different things and they're both predicting mortality. That was for global health.

They also did this for ADLs. And, again, the pattern was remarkably similar where you see that both baseline and change in gait speed or SPPB are predictive of mortality at the same time that change in self-reported ADL was also predictive of mortality within the same model.

So I won't talk a whole lot more about the performance test. We could spend a long time talking about them, but that's not what this meeting is about.

I do want to tell you about a new approach that I think of as a little bit of a hybrid. And it technically is a patient-reported outcome, a self-report.

But a group of investigators at Wake Forest, Jack Rejeski, Tony Marsh and Eddie Ip came up with an idea to try to standardize self-report a little bit using video clips. And they experimented with this in different ways and decided that it would be best to use kind of a stick figure. It wasn't a man or a woman or a person of a specific race or body composition and to show videos and to ask people specifically about how their function related.

They wrote a series of papers on this. This battery is called the MAT-sf, the Mobility Assessment Test Short Form and it's already been applied. It was being used in the current life study, so we're going to have a lot of data from a large clinical trial of exercise to see how this particular measure works.

So let's see if the videos work here. Yeah. They basically – people would be shown not all three of these, but in this slide all three are shown together. And they're asked, say for this top one –

Oh, no, I've lost it. There. There we go. Cynthia's about to come to my aid– so you'll see for the top one, they're shown to walk. And so I was hesitant to even try to put the videos in. Something always happens to videos when you go to give the talk.

Jack Guralnik: Yeah. Okay. So they're asked to look at a specific – see I waited too long then. You got to click quickly here. Okay. There we go.

So you see. And for this particular task for the walking, they're asked how long could you walk at this speed, so they're shown that

specific video and asked, and it ranges from 1 minute to 60 minutes.

For the stair-climb test, how many times could you walk up steps at this speed? And you'll see there are very different speeds there.

What I thought was great is that in addition to these very sterile looking ones, they also had – and I don't have this on video – but a picture of an outdoor scene with a rocky terrain. And this was a yes-no question. Could you walk up this hill in an uneven terrain? And it's hard to see here on this, but this stick figure is carrying two grocery bags. Could you climb this flight of stairs with these bags? So these are very much mobility and lower extremity with just a little bit of upper extremity.

And they ended up with 79 items, looking at the same tasks at different speeds. And then they did item response theory work and brought it down to a 10-item battery called the Mobility Assessment, well, Short Form. And this can be administered in less than 5 minutes. They have a version – it could be downloaded from the App Store, so you can put it on an iPad or a laptop computer. It's got excellent relationship of the 10 items to the 79 items and then very good 2-week reliability in a small study.

And so this in some ways is a way of while using a patient-reported outcome also somewhat standardizes the test. Now, if you haven't tried walking up a rocky hill, you're still kind of speculating. And that's one of the limitations in self-report also, always that if you haven't actually done a task, it's hard to say if you can do it or how difficult it is. But, at least, it does standardize the test.

And strong relationship with objectively measured performance. This shows the score on the test for each level of SPPB, a clear and graded relationship. The correlation between the MAT-sf and the SPPB was 0.59.

And then I mentioned that we're using this is in the life study, but I also put this group together with a group that I've worked with from the University of Montreal that's doing a study called, "The International Mobility in Aging Study" (IMAS). And this study is being done in French-speaking Canada and Ontario, in Brazil, Columbia and Albania, of all places. And the translations were quite easy, because there's not really that much text in this test. So now we have these tests done in these sites.

And this is preliminary work done in two sites, one in Brazil and one in Columbia. And the Brazil site was quite a poor area in a city named Natal, where self-reported health was quite poor, but it shows that for each level both functional limitation as measured by self-reporting, the SPPB scores, that in each city that you see, that within a certain category of SPPB score, you get very similar scores on the MAT-sf.

So there are a number of advantages of performance. And so while I still have the podium and being able to speak about performance, I'll talk about some of the advantages. And then we're going to spend the rest of the time here talking about how self-report may be valuable in addition to this.

There's clear face validity. We've seen very good reproducibility, sensitivity to change. They can be used in persons with poor cognitive functioning. When we do these tests, we give the instructions, but we also demonstrate it. So we've done studies where people didn't have severe cognitive impairment, but had moderate cognitive impairment and were able to do the test pretty well. It reduces the impact of culture, language and education.

But there are some disadvantages to the performance test. They do take time to perform. We need to train our examiners. They should be done in a very standardized way. You do need adequate space, although this has not really been a problem with tests like the gait speed and SPPB where we've done them in small apartments in inner-cities. Potential injuries, again, is something to think about, although the SPPB has been done all over the world probably 25,000, 50,000 times and I haven't heard of any serious adverse events from doing it.

And then, again, at a more basic level, these simple tests may not reflect performance on complex tasks where adaptation to the environment in daily life and that can be important as people perceive how an intervention may be affecting their quality of life.

So I was asked to talk about the FDA qualification process.

Now, as I said, the AIM Coalition is supporting an effort to get the SPPB and gait speed alone approved through the qualification process. The reason to talk about it here, I think, is that as people are proposing to use PRO measures, it's worth thinking about whether you should try to go through the qualification process for these measures. And as I'll show, the bar is set pretty high for

qualification of any outcome measures, but it's something to consider.

So the FDA Clinical Trial Outcomes Assessments are used to provide substantiation for treatment benefits claims. And there are two processes for FDA submission and review related to outcomes.

One is the way everyone has been doing it for years, which is part of a drug application review. Often, you do the trial without quite knowing what's going to happen in terms of how the FDA will react to the outcome you've decided on. Sometimes, people do have consultation before they start the trial.

But the new approach that's being offered is under the Drug Development Tool, kind of an unfortunate acronym, DDT, qualification process. And this is a process that could save lots of time and money both for the FDA, who wouldn't have to deal with every drug trial individually, and, of course, also, for the people developing the drugs.

So this is an FDA slide from one of their presentations. They describe this as a new regulatory process to provide publicly available drug development tools independent of the application process. It's got two stages of consultation and advice and qualification review.

So there is a fairly long process for the qualification. The document that they just released in January has a two-page flow chart showing how you get through this process. It starts with a letter of intent proposing what you would like to do.

And we've gone through this with the performance measures. The first time, there were comments on it and it was not accepted. The second time, they did accept it. So we've moved ahead somewhat.

And then you come up with the DDT briefing package and then start working with something called a Qualification Review Team. So the nice thing about this is that this is a collaborative process between the FDA and people who are proposing these outcome measures.

And follow that submission, there's the investigation of this outcome and development of it.

And then, finally, review for the qualification decision.

Dissemination is a very important part of this process. And, again, this is good for the field. And one reason why the AIM Coalition wanted to move ahead with this is this is not a proprietary outcome. When you decide to go into the qualification process, you agree to making the tool known and available for use by all drug developers maximizing the value for public health. There'll be a notice in the Federal Register when outcomes are approved. It directs the public how to access the drug development tool at the location where it's maintained, so you'll have a website which will explain the outcome and how to do the testing if testing is required for that outcome. And then have appendices and supporting documents which are posted on the webpage related to that drug development tool.

So the next step, and the reason I'm showing this is that, again, for this meeting where people are thinking about patient-reported outcomes, if you want to consider going through this qualification process, you have to think about all the things that are required by the FDA to get through it. And, as I say, the bar is set pretty high. You've got to go over a number of different hurdles in terms of having the data.

And we've been working on these performance measures for a good 20 years and there's lots of published reports. Developing a new PRO would mean that you would have to go ahead and do cross-sectional and then longitudinal studies to really document some of the psychometric properties of that.

So you need to describe: the meaningful aspects of the patient experience; labeling and/or promotional claims; the context of use has to be clear; the definition of the disease; the targeted study design; critical details of the measure to the degree known; overview of the current; clinical outcomes assessment development status; description of involvement of external expertise. This is just kind of the warm-up for it.

And then the real methodologic areas: evidence of content validity; cross-sectional evaluation; both score reliability and construct validity. And then longitudinal evaluation of the measures looking at: longitudinal constructability; validity; and then very important to the FDA, ability to detect change; and probably the most important is area 2.4, the longitudinal evaluation to provide guidelines for interpretation of trial results, so evaluation of individual patient change. And then, finally, the ability to translate into other languages and adapt and have content validity similar for versions in different languages.

So quite a challenge, but the benefits at the end, if you do get an outcome qualified, is that it will really help propel forward clinical trials knowing what outcomes the FDA has already accepted and kind of preapproved.

So there's a wide range of spectrum. We think in terms of functioning, of mobility as being very important for sarcopenia. Here is a couple with clearly very high mobility. Along that spectrum, we see a couple who seems to be moving somewhat more slowly, but still getting along. And then, finally, at the other end, this was a sign I found in London. It's a horrible stereotype of what elderly people are, but it is the other end of the spectrum where –

Jack Guralnik: Clearly, a very wide spectrum. And so we need, whether it be performance measures or patient-reported outcomes, to be able to cover this whole spectrum of functioning so we can observe people moving in both directions across that spectrum.

So, with that slide, I will stop *[laughter]*.

[Applause]

Jack Guralnik: Question or comments?

Stephanie Studenski: As we move forward with this, there are a lot of instruments out there that have been used to measure health and function. A lot of them are pretty heavily influenced by mobility like even the SF-36, the physical function, eight of ten of the items, or I can walk a mile, I can walk a few blocks, I can go up three flights of stairs, I can go up one flight of stairs.

I just wonder as we go forward, if we're considering trying to check some of the more widely accepted measures as well as starting over? Because if they're not all that different in their capacity to meet these things, it might be attractive to use something where there's already humongous amounts of data like the SF-36. I think there are some other items and some other standard scales that are awfully heavily mobility framed.

Jack Guralnik: Yeah. This is a terrific idea. And I'm hoping that the panel later will spend a lot of time on that, because we have one item that's been circulated that is a proposed PRO measure for sarcopenia but we do have a lot especially if we're focusing on mobility, which I think is an important domain of functioning to focus on.

Stephanie Studenski: Because I just want to point out, we've tried to do some of this work. And the other key issue is what's the gold standard? Are we going to go back to the gait speed changes, the gold standard for what's a meaningful change in self-report?

Jack Guralnik: Yeah.

Stephanie Studenski: So I think it's important to have a sort of overall measure of whether a person could perceive a change in their overall mobility or some of these sorts of things, because otherwise it's hard to anchor meaningful change in a score somewhere.

Jack Guralnik: Yeah. Good point. And I think that especially if people decide to go ahead and try to get a qualification from the FDA, you're going to have a whole lot easier time if you do something like the SF-36, where, as you say, there's tremendous amount of data. And you really would have a shot at doing it with that, whereas –

Stephanie Studenski: Not that I think it'll necessarily be better. I'm sorry.

Jack Guralnik: It may not be better, but we can do research to see how it relates to other proposed measures.

Bill Evans: Yeah. I think Stephanie is absolutely correct, but I think the guidance from FDA, and I would really welcome FDA comments about this, is that it really needs to be very specific to the indication. So while the SF – you know, however many SFs we have, we really do have a domain that's strongly anchored in physical function just as the quality of life questionnaire, the St. Vincent Questionnaire for COPD. Almost all the questions are related to physical function. And, in fact, some recent studies have shown that that is related to SPPB. However, my guess is that we could never use it in anything other than COPD.

So I think that it would be great to have some guidance, you know? When we talk about the development of the PRO for sarcopenia, at the time, we were developing a PRO for cancer, cachexia and a separate one for prostate cancer. And what was interesting is that while some of the comments and questions were really similar, some were different. And I think that each disease or indication has some specifics about it that may be a little bit different in the way the patient perceives that difference.

So I think that it would be great to use a well-developed, validated PRO that has been used in a number of different indications. My guess is that the regulators have a different view of that.

Jack Guralnik: Yeah. And here we have a regulator coming to tell us the answer.

Stephanie Studenski: I just want to rename the instrument. How about the physical function for sarcopenia?

Ashley Slagle: I think you hit the nail on the head that we really do look at instruments in terms of the condition that they're being used to evaluate outcomes in. I'm not going to speak about whether the SF-36 Physical Function Scale is appropriate or not, because we actually haven't defined the population that we're talking about. So, of course, we're open to lots of different instruments. It just depends on what we're trying to measure and in whom and where in the endpoint model that instrument might be used. It may be that it's relevant secondary – supporting a secondary endpoint, but not a primary. There are a whole lot of things is what we call the context of use that we have to think about before we can determine whether a particular assessment is appropriate or not.

Jack Guralnik: Yeah. And we also have to –

Dan Perry: That was Dr. Ashley Slagle, the voice of the FDA, for all of you who are not familiar with Ashley.

Cynthia Bens: These are just Ashley's opinions. They're not the express view of the FDA *[laughter]*.

Jack Guralnik: There's a little asterisk there. Yeah. I mean one of the reasons that measures of functioning work so well is that function is the final common pathway of pretty much everything that goes wrong with us. And the six-minute walk is responsive to many, many conditions. It's not absolutely specific to the one condition, but it's been used as an outcome for a number of indications for different drugs.

So there is going to be some more general relationship between some of these functional outcomes and multiple conditions and indications. So that's where the challenge lies I think, yeah. Cynthia?

Cynthia Bens: Hi. And this is not actually a question. It's more of a comment, because Jack went through – I think he's been going through the process of putting together a briefing package for the qualification

process and it is a really involved process. But one of the things that I think we can all learn from the roadmap that's really been laid out for what you need to think about moving into the qualification process is actually something that's really helpful for everyone to think about whether or not you're pursuing qualification or not, which is really understanding the disease definition and in defining your context of use.

And one of the things that I thought was really interesting that you brought up this morning about the FNIH work is the differences between men and women and even how there are stratifications within women that may present differently.

And one of the things that I really learned last week was how that relates to labeling and whether or not there is truly the ability to define that subpopulation based on the disease definition of population a company is looking to study.

So these are all really interesting things that I think everybody should keep in mind. So I don't really have a comment necessarily whether or not these should go through qualification, because I know that there really isn't enough data, I think, at this point to be able to put them through, but some of the concepts will be really helpful moving forward.

Dan Perry: Any more questions? Yes. Please identify yourself.

Qian-Li Xue: Qian-Li Xue, John Hopkins. Nice talk. When you talked about the results for self-report for this performance measures, so I have this picture in my mind that's just a simple two-by-two table. So you have the performance on one side, then you have a self-report disability on the top.

So there are two cells that I think would be quite interesting. And the one is do you have a high performance based on a performance measure, then you have a lot of self-reported disability. So there I think is like underachievers. Then the opposite of that cell would be the overachievers. I mean if you measure the physical function by the performance measures is very low, then somehow they did not report any difficulty. I think that maybe you can learn a lot from the overachievers.

And then in terms of targeting whether the priority should be placed on the underachievers rather than just rely on a self-report alone or a performance alone.

So I just wanted to hear what you think about –

Jack Guralnik: Yeah. That's a really good point. And I think that speaks to the possibility of combining self-report and performance to pick up those people on the off diagonal. And one of the reasons we got into performance in the first place is that we knew there were people like that who were misreporting.

And we see evidence of that just when you ask the simple, self-reported health status question, excellent, good, fair and poor. People have their own way of reporting it in relation to how old they are, so you don't see that big a change in self-reported health status as people age, because they're relating it to, well, what should I be at my age, whereas, you see a very substantial decline in objective performance measures as people age. So just in that you see discordance between the two. And I think it can be very enlightening to study that discordance. That's a good place –

Qian-Li Xue: Yeah. And then, also, and in terms of the underachievers, I think that then you can look into the reasons for that. It could be extrinsic, environmental barriers and to the level of self-reported disability, then you can also look at the intrinsic potential explanations and why they're underperforming.

Jack Guralnik: Yeah. Yeah. And if you're dealing with disability, as I pointed out, the environment is absolutely essential in considering disability.

And I tell a story about doing some work with an Italian population, a small town near Rome. I had a post-doc come for a short time with data from this population. And the Italian population looked better in many different things and in their mobility, in their self-reported disease status, but they looked worse in ADLs. And we couldn't figure out why. And then we looked at the individual ADLs and it was the bathing ADL. And then we looked at the environment. And this old hill town in Italy only had these giant old bathtubs that any of us would have a hard time climbing into. And the U.S. population we were comparing to, a lot of people had walk-in showers that they were reporting on.

So there the environment is so different that you kind of need the performance measures to do a standardized assessment or something like this MAT study that I showed you where everybody is looking at the same task and not considering the environment that they happen to be living in.

Qian-Li Xue: Yeah. Thank you.

Jack Guralnik: I think that's a reason for discordance. Yeah.

Rezaul Khandker: I have a quick question along the same line. You showed a lot of correlations between measures. And it's like a technical question actually. For example, SPPB and ADL, one is a continuous measure, so variation is possible – much more variation is possible than some of the other ones, which have discrete, a few categories. So when you look at the correlations, these measurement properties probably come into play. So I was wondering if you'd comment on that.

Jack Guralnik: Yeah.

Rezaul Khandker: I'm sorry, Rezaul Khandker from GlaxoSmithKline.

Jack Guralnik: Yeah. Yeah. I mean I was cheating in a way the way I showed it. We have some self-report on a more expanded scale where you would probably see a better correlation, but as Qian-Li mentioned in the last comment, there are people who really are somewhat discordant and that you can look at people who don't report any disability who then have different performance measures and do have different outcomes in terms of mortality and other kinds of outcomes. So I think that even within a certain self-reported category, when you narrow it down, you do see a distribution of performance.

And you could probably go the other way. If you took a category of performance and you had a really refined self-report, you would probably see a gradient of outcomes in that. So I don't have the data to do that, because we didn't have the right questions.

Dan Perry: Time for one more question.

Brock Beamer: Brock Beamer with the University of Maryland and the Baltimore GRECC. Every time I start to get optimistic, I hear something like – Stephanie said two percent correlation between the different measures.

Stephanie Studenski: That level of agreement. It is not the correlation.

Brock Beamer: Level of agreement, okay.

Stephanie Studenski: There's high levels of agreement on who doesn't have it.

Brock Beamer: But not who does.

Stephanie Studenski: There are low levels agreement on who does have it.

Jack Guralnik: Well, are we talking about sarcopenia now?

Brock Beamer: This is the sarcopenia, right. And I guess –

Jack Guralnik: We're not talking about performance. Yeah.

Brock Beamer: No, no, no. I understand.

Jack Guralnik: Okay.

Brock Beamer: No, my question is posed to for the patient self-report. For all the problems with it and on many things, it does less well than SPPB. If that's the bar that we're aiming at, how close are – I mean do you have a sense of how well there is agreement between patient self-reported compared to physical performance? Let me rephrase that.

Jack Guralnik: Self-report of function?

Brock Beamer: Self-report of function. My personal belief has always been that, oh, one person is depressed, another person has a big bathtub, *et cetera*, that it's going to be terrible correlation, but two percent – I'm sorry these correlation, but two percent agreement between different measures of sarcopenia leaves a big area for agreement between self-report to be a comparable measure.

Jack Guralnik: Yeah. Yeah. No, it's much better than that. If you look at correlations, again, self-report scales that have a little bit of detail in them and a range of scores, you'll get correlations of 0.3, 0.4, as high as 0.6. So they're reasonable. We wouldn't want it to be a perfect correlation, then there would be totally redundant measures. I think from the modest correlation, we also assume that they can complement each other. But it's a reasonably good correlation. I mean people –

Stephanie Studenski: Can I just clarify that in many of the populations we were looking at the prevalence of disability and low muscle mass is pretty low. So the overall agreement rates – if you have two positives, false positive, the whole thing, two negatives was probably 90 percent. It's just that the top, left corner, yes, yes, was where there was a lot of discrepancies.

So I think it's important that you don't get too confused about agreement levels, because it will depend on the populations and what you mean by agreement. I mean, right? So it's not just overall agreement. It's where is your concern?

Brock Beamer: But if I'm targeting a drug and it makes such a big difference which scale I use, which person I get, it –

Stephanie Studenski: Well, I think we're using these as outcomes not as eligibility. In some senses, I would think you want them to be complementary, not the same. I mean I'm not sure. But one is supposed to help inform the other, the self-report and the performance.

Jack Guralnik: Yeah. Well, but we are using gait speed as the entry criteria for sarcopenia diagnosis, so it's part of the definition. So people will be starting with a lower gait speed, but that shouldn't affect using gait speed as an outcome. In fact, it may help, because if you're starting low there's no ceiling effect. You can see improvement very easily if there is improvement.

Brock Beamer: Thanks.

Dan Perry: Great. And thanks for that clarification, Stephanie. Jack, you're through.

[Applause]

Dan Perry: Our next speaker is Dr. Donald Patrick. Dr. Patrick was the first director of Social and Behavioral Sciences at the University of Washington in Seattle, so he comes to us all the way from the other Washington. And we're delighted to have him really delve into the issue of patient-reported outcomes for sarcopenia. He is widely published and highly regarded as a real expert in chronic disease and health disparities, outcomes regarding vulnerable populations and end-of-life care. So he's the man to hear from on this subject and we're delighted that you made the trip out here, Dr. Donald Patrick.

Donald Patrick: Thank you very much, Dan. I appreciate that. And appreciate the invitation from Cynthia and from Aging in Motion.

I had such fun when somebody said – I've been in Washington every week, this Washington every week and the other Washington every week this month. And they said, well, what's making you decide to go this last trip of March? And I said Aging in Motion.

Oh, well, we're all aging in motion they would say, because this is a really important endeavor and I'm really delighted to be here.

Secondly, I was a special government employee for five years. I was very involved in writing the PRO guidance with the FDA, but I'm speaking on my personal opinions today. And I'm really fond of the term, what does the FDA think? And I can tell you after five years, there is no such thing as the FDA. And so we should disabuse ourselves of that notion from the very beginning.

So I'm a PRO person. And I'm dropped into this area. It's not the first time I've been dropped into sarcopenia, but it is the first time that I'm trying to broadly think about PRO measurement development and how do we do a way forward here?

So I'd like to identify the main challenges in using reports from patients about their experience with sarcopenia and evaluate the possible approaches to incorporating patient-reported outcomes and the endpoint strategy for evaluating a treatment benefit.

So I would like to remind everybody that patient-reported outcomes are not the only patient-centered outcomes nor the ones that are only important to patients. Patient-centered outcomes are those outcomes that are important to patients and there's nothing more important to patients than survival. It trumps absolutely everything else then followed by functions or feelings as affirmed by the patients themselves or judged to be in the patient's best interest when the patient cannot report for themselves. So we're talking about patient-centered outcomes and I can guarantee you that the gait speed is of real importance to the patient themselves.

So defining sarcopenia, we've spent a lot of time on this. I loved reading the little piece of brado that comes before that it's a loss of muscle quality during aging characterized by decline in muscle strength, that if untreated can lead to weakness, disability and increased risk of falls and loss of independence. I love this, because it has all the concepts that might really frighten somebody of the term sarcopenia.

So I've had four knee operations in the last two years for severed tendons in my knees. And I was at my physical therapist's office, which I go to twice a week. And I said I'm going to a meeting on sarcopenia. And she says what is that? And I said, well, think. And she says, oh, muscle weakness.

So the term has meaning I think and it can be figured out with anybody with a little bit of health training. It's a little bit weird to the public, because I haven't met anybody in my family that knows what sarcopenia is.

And I was really interested in this relation of comorbidities, because I've gone through these knee operations, because my bone surgeon, my knee surgeon said you better do it now while you're young, I'm 70, because later on when you get muscle weakness, you better have better knees to cope with that. So the comorbidity issue is a really big issue.

So it's clearly a patient important issue, but, as we have been talking about already this morning, there are issues of relating muscle mass to whatever function is considered to be.

So I start out by thinking of clinical trial endpoints and the examples of clinical trial endpoints with survival being the foundation across all of them for many, many conditions, biomarkers, followed by performance measures, of which we've been talking about this morning of the SPPB, followed by some clinician reports including those radiographic readings that include human interpretation or anything with human interpretation versus anything an observer can do – and we haven't talked about that and we might want to pay a little bit of attention to that later – and the patient-reported outcomes.

So I'm interested that we've gone further on the biomarkers, because I went to look to see whether there was a consensus on biomarkers in the published literature and I didn't see it. But I loved all of the possibilities here of the biomarkers, because that's important to consider in any kind of endpoint strategy.

Then we're looking at our clinical outcome assessment and that's what this entire body of these measures are labeled under. And these have human assessment involved or human judgment involved. And it's not just the patient-reported outcome, but what a clinician does or even some performance measures, because of the instructions to perform a defined task.

So whenever my PT says put your back against the wall and squat, I say, oh, I don't feel like it. It's Friday at 4:00. When she says that on Monday at 10:00 AM, I squat right away. So I'm really interested in when the test is taken. Any time I'm given a performance test, my favorite is the Humphrey. I have to go 8:00

in the morning, because I hate the damn thing so much I want to be at my best if you want to excel.

And somebody talked about achievers, under and overachievers, interesting term in relation to performance measures.

So the roadmap is absolutely critical to our entire discussion today, so I want to introduce that a little bit. And this was published this year on the FDA website. And we are going to be thinking clearly, hopefully, about the natural history of the disease and its diagnosis, its range of manifestations as well as the patient subpopulations. And we've already identified some of those this morning, I think most notably gender, but I imagine that there are some genetic issues that are lurking in the background here that probably need to be put into the pie in terms of the phenotype that we are talking about. The health-care environment and the treatment environment as well as the patient and caregiver perspectives.

So what we are looking for is evidence of treatment benefit with any endpoint strategy in a regulatory process or in a clinical trial of any kind of an exercise intervention or broadly any kind of intervention. And our direct evidence is coming from a COA, that endpoints that measure survival or feeling and function and notice this critical thing in daily life. Daily life is a lot different and we need to explore that issue.

Versus indirect evidence of treatment benefit, the surrogate endpoints that measure other things that are related to how patients survive, feel or function, in particular the biomarkers.

So when is a clinical outcome assessment adequate for use and adequate in well-controlled studies? I always love thinking about my time at the FDA, because I went in as having worked for almost 35, 40 years in PRO measurement. And I'm not so sure I learned a lot about PRO measurement, but I sure did learn a lot about adequate and well-controlled. And I learned a lot about labeling, and a lot about what's behind measurement. And I can tell you that it's sort of sobering that we spend all this time focusing on the measure rather than focusing on what it is that we are measuring. And I think that's really important in the field of sarcopenia.

So the concept of interest, and that's going to be something I'll keep on coming back to, has to be measured both validly and reliably in a well-controlled study context of use. And it can be relied upon to measure the concept of the interest and have specific

interpretation. And so the interpretation of whatever we're doing is important in regulation and particularly important for labeling.

For clinical outcomes assessments that do not provide evidence of how patients feel or function, the concept assessed needs to be an adequate replacement for how patients feel or function in daily life. So it's the relationship between those measures and what is going on in the outside world that is important.

And that brings us to the whole proximal-distal thing. I have raised a child with spina bifida, and so I've been involved in the disability community. And I remember the first meetings I went to in which they said I'm not disabled, society is disabled. And, you know, in many ways, that's quite true.

As I was coming in today, this morning, I met the gentleman and I saw the gentleman in the wheelchair pushing his chair down the hall. And I know from experience with my son pushing on a carpet, now these carpets are not too bad, but you try pushing on the carpet at Denver International. That's my favorite one. When I'm with people in a chair in Denver International, getting from terminal to terminal trying to go over a high-plush carpet with a wheelchair, who has the disability?

The distal environment impacts constantly what we are doing in terms of our actual performance of what we are doing. And it is there in every bit of the way. It's also there when you give the performance test. How does it look like in the clinic? How is it set up?

Now, we take for granted that a Snellen is a Snellen and an eye chart, but I'm often fond of going into the ophthalmologist's office and seeing the different sections in which the Snellen is done. It all requires standardization of high – and it's a good thing that we do pay a lot of attention for that.

So the disease defining concepts that we're thinking about, the performance battery and the core signs or symptoms, as we move in that progression – and Jack did a nice job with the Nagi model showing you that that progression brings the social and the environment into each of the definitions to a greater and greater extent. So, to what extent when we get somebody's evaluation of excellent, very good, good, fair or poor from the NCHS, not from the SF-36, came out of the National Health Interview Survey years ago, how does that relate to what's gone on earlier? And what does that predictive have to do?

And I was really happy to see that you used the self-rating of health, because it's one of the most robust general health status measures in predicting functional decline and mortality and we've got years of experience with that. If only we had that much experience with other functional status measures that complicate that issue a little bit more by trying to give specificity to our self-rating.

So understanding the disease or condition is the first criteria. And I found this FDA slide about a good definition is important. And I think we're probably all very well aware of the bottom line that not everyone who has a relatively low muscle mass has the clinical problem.

So I think that what's gone on with the Foundation – and I really was delighted to see the use of CART and all of those indicators to come up with what had a better definition of specificity. And I think a lot of progress has really been made in identifying the population. But the good definition remains important keeping in mind that not everybody that has those indicators is going to have a problem.

So it's a work in process to find an acceptable indirect indicator of muscle mass and muscle strength to define the condition.

So how consensus is it? And consensus by whom? And I would want to point out that consensus by the professional community isn't consensus by the division that evaluates your drug application.

For reasons that are bound by regulatory process, when we're freed and unfettered by that, a little bit more open to being more progressive. When we go to the agency, they have to look at everything they have ever approved in this particular class or this particular division before and what was used. And sometimes it takes an act of God to bomb them off of whatever definition they think of as of the disease when we have moved so fast ahead.

So the next thing is getting the consensus around the entry criteria. And I'm delighted you worked in Europe. We cannot forget the aging in India and China and Japan. And Japan is a huge, huge issue of aging, because I was struck by the characterization of shrunken and weak, as being sort of shrunken and weak. And, boy, whenever I'm in Japan do I see a lot of old ladies that look shrunken and weak walking around that society.

So it's obviously a condition of high importance to the people who have it. Of course it is, because it is defining what people want to do. And this is a need to evaluate treatment benefit and I would say an absolute essential to evaluate treatment benefit using patient-reported outcomes.

So, the next question I would think then is what are the desirable endpoints and can we achieve them?

So that's conceptualizing the treatment benefit and beginning of either use of an existing measure, of documenting the content validity of that measure and going through all the things that Jack has gone through in preparing a dossier for qualifications.

I don't think is any more onerous than trying to qualifying a new lab test, frankly. This is just the social science and the application of science to what humans report for themselves. And it can be approached with the same rigor that we do most of the rest of our development.

The real problem probably lies in the interpretation and that's the biggest issue. When we are in the chemistry lab and we take this Solution A and we add it to this Solution B and we end up with Solution C, we know what's gone in from A, B and C and we can interpret how C came out. In a PRO, that's much more difficult and that's also the need for documentation. Just like lab science, we have to document and use our concepts like a solution so carefully.

And I'm going to make some comments already in the meeting today about that language and those concepts that probably need to be clarified in this particular field.

So, in our context of use, we're going to develop an endpoint model that displays the role and the hierarchy of relevant concepts in clinical trials. And you'll have a primary, a secondary within the hierarchy. And outside of the hierarchy, exploratory outcomes that are extremely important in drug development, medical product development, because often we find something here that slips up here right away after it has been looked at a number of times and eventually makes its way into a labelable concept in the primary or secondary endpoints.

So what is our endpoint model in sarcopenia? Well, I just threw one out here and you probably can tear it apart. But I thought of

change in selected biomarkers as being here. And notice I didn't put primary, secondary or exploratory on this, because I would want to leave that up for grabs based on the data and based on what you could put forward. Change in usual gait speed. Jack and I have been on a committee with NIA and I'm a big fan of gait speed and of the performance battery as being extremely important.

And how I measure my gait speed is the amount of time it takes me from the C concourse at Dulles to the cab. And I can tell you that that's really lengthened with my problems. I used to be the first off the plane, and, my God, I was on that people mover and chomping at the bit when the door opened and I'm in that Washington Flyer. Last night, when I came in at 9:10, I thought, oh, Lord, I'm so slow. My goal is to be at the hotel by 10:00 and I made it at 9:50. So I still felt pretty good that I can push it even walking with a cane. But gait speed is important as measured by who gets to the taxi first.

Change in short physical performance battery. Okay? I think that's all right.

And now I want to talk about the PROs, the change in sarcopenia-related signs and symptoms and in impacts. And I haven't labeled those as PROs, but they are.

So we want reports directly from patients without interpretation or response by anyone else, so it's coming directly from the patient. And it includes signs and symptoms, function and so-called quality of life.

Please note that these two terms, health-related quality of life and quality of life, are not equal to symptoms or function. There's a big difference between how you function and what you term your quality of life. And people with very poor function report a very high quality of life.

All you have to do is read Stephen Hawking's introduction to *A Brief History of Time* when he says, "I don't have a bad quality of life." He uses the term. And that was a big teaching to me that it's a term that's separate from how one functions. It has an awful lot to do with a personal nurse and all the equipment that Stephen Hawking moves around in the world from place to place.

It may also include satisfaction and care and adherence, but this is generally not in the regulatory context, but important for medical product.

So when we're talking about perceived quality of life or symptoms and function, what are we talking about? We're talking about ratings in general about how good, well, of a sensation. There are pins and needles or numbness. And I really don't know enough about muscle weakness to know how much signs and symptoms are a part of it, but I'm interested when the work that we had done in cancer-related cachexia and sarcopenia that signs and symptoms are definitely part of the impacts that are mentioned by patients. And we probably want to keep those very separate from their reports of function.

So within that context, and Bill's already talked about it, we attempted in a study to develop a measure of self-report of reduced muscle strength in sarcopenia for prostate cancer and other cancer-related issues. It was a qualitative research study with adults that were 55 years and older with sarcopenia, who attended open-ended concept elicitation interviews to characterize the functional effects of reduced muscle strength. That's what the qualitative interview guide said. We want to know how reduced muscle strength impacted your life.

The qualitative data were analyzed with a usual qualitative program. And we drafted a PRO measure. And then conducted cognitive interviews with additional sarcopenia subjects to refine the measure.

So, remember, this was only one condition in which it was looked at. And funding was provided by Amgen and is not in the public sector as far as I know. It remains proprietary measure and thus therein is a rub.

I would say the bottom line is that the article provides assurance that PRO development is possible. But publicly-funded development and qualification is possible following a path of performance measures already submitted to the FDA.

I think you have just as high hurdle with performance measures as you will ever have with PROs. Don't put the PROs as saying, oh, my God, you'll never get this down there, because they're going to be looked at under the same criteria.

So we did in this project look at what was a good item. And we were trying to get unidimensional concepts that were relevant to people determined by how often people mentioned the – were easily understood, measured a concept that was likely to change

with successful treatment – because that's one of the big issues is finding something that changes with successful treatment – was not vulnerable to a floor or ceiling effect from baseline to follow up – in other words, can change with not everybody not being able to improve, because they're at the top or not being able to deteriorate, because they're at the bottom – and likely to have semantic equivalents with other languages in a rough attempt to do translatability prior to the development of the instrument.

Again, I want to say that we are operating way distally here in what our questions were in asking them. It was basically impact questions and it bordered over a little bit to general questions.

So what were the signs and symptoms elicited? I went through the concepts that we elicited with a new lens coming to this meeting and they were strength, energy, balance, endurance, coordination and the last one is really sort of a conundrum – all I can say is strength, energy, balance, endurance and coordination is what I do in PT every week. I'm actually tested and I'm given exercise to improve every single one of those aspects. So that's part of a muscle weakness, I would think too, therapy and how well does the treatment improve all of those as done by different endpoints and is done by patient-reported outcome?

The impacts were what you might expect, limitations and activities of daily living, social limitations and emotional limitations. But there were definitely a lot of emotional symptoms that were reported by this population, and I don't know quite what to do about that. Is that simply – is it really from muscle weakness? Is it from the population? Is it the affective issues of aging or where does it come from? Is it really specific to this condition? I would be somewhat doubtful.

So herein is we start talking about the specifics. So these were some of the items that were created. How much difficulty did you have walking a distance, for example, a hundred yards or the length of a football field, which believe it or not the SF-36 people have found an international equivalent to that in every society since they don't have football fields. Difficulty walking in a straight line, walking without stumbling, going up or down stairs. Now, note, how much difficulty? And already this morning in practically every speaker we have not used that terminology. We've asked what can the person do? And this is extremely important for several reasons.

If you ask me can you walk up stairs? I'll say yes. If you ask me did you walk up stairs, yes, no? How much difficulty did you have? It's a different answer.

And this is known as early as 1951 when one of my mentors, the late Professor Margot Jefferys in London, traveled around asking people if they could walk up stairs. She was doing these interviews in their home. And they said yes, I can. And she said, hold on a minute. She went out to her little Morris Minor and took out the stairs and brought them in and said would you mind doing this for me here?

So the self-report is much better asking what people actually do not what they think they can do. And this is all the context with which this is done.

I spent about six years in London. And we had a big meeting at the Department of Health in London to talk about this in terms of national surveys, were they going to use capacity language, can you do this versus performance language? And this was to estimate the number of people with a disability in the United Kingdom. And the meeting concluded that they were going to use capacity language, because the government wasn't interested in what they actually did do only in what they thought they could do.

So for policy purposes, you might actually use capacity. But for trying to correlate with a performance measure asking for exact reports even in the video – I don't know, Jack, whether they're asked whether they can go up these stairs like this or whether they do do that and exactly what the video shows, because that's a different piece.

And I loved the one with the groceries, because that's the hardest thing for me to do. And I do carry them. I carry them for about four steps, put them down, pick them back up, go down another four steps. Of course, it's so slow, but in the end I get there. And I think that's probably what we want is getting there.

So asking how much difficulty is what I would recommend.

So the implications of this study is that it's possible to conduct qualitative research sufficient to provide evidence of content validity of what people say. But I think the real question is what is the concept of interest? What are we actually trying to get at? The daily experience of muscle weakness, the symptoms of muscle

weakness, the impacts of muscle weakness. And we have to focus on the right concept of interest. And what those domains are, I think, remains to be seen.

So the challenges, I think, the main challenges of using patient report is we do need the consensus on disease definition, because we have to know who we're talking to and what those patients are. And without that context of use answered, it makes, I think, all of our work difficult. I think that's what you're experiencing with the qualifications process. And I think that there's great progress reported at this meeting on that.

I think focusing on both the proximal and the distal is probably a challenge and that we need to do both. We need to be thinking of the whole trajectory and the cascade of endpoints from the actual biomarkers of muscle weakness down to the reports of daily experience in terms of what people do.

The bottom line, of course, is a very broad question and this is a capacity question, are you able to do what you want to do? And I can tell you in every piece of qualitative work we do that's what we hear. The bottom concept is are you able to do what you really want to do?

And then meeting the measurement challenges that have already been identified in the guidance and the PRO guidance remains to be the light here for all types of clinical outcomes assessments.

Now, the question came up about generic versus specific. And I think I did want to say something about that.

Yes, we have physical function measures that are generic. Not only do we have ClinROs like the Barthel Index. We have – oh, I'm saying hundreds of physical function measures.

The SF-36 is called the PF-10, Physical Function 10. And then it focuses on primarily ambulation-related.

Well, early on in my career, I was fortunate enough to work with the Sickness Impact Profile and Marilyn Bergner at the University of Washington when I was still in New York, and very quickly learned the difference between walking and mobility. Walking is not mobility. If we're talking about mobility, mobility is moving from place to place and you can do that without walking.

And so we are actually talking about gait speed, that's walking. We are going to have this issue of what do we do about muscle mass with people who do not walk? And I think there's a huge proportion of the population that's going to be using aids and having the ability to be mobile, but not necessarily walking.

I personally believe we need to be specific, and this is for several reasons. Condition-specific measures have been far more responsive to treatments that may be minimally effective. Okay?

So if you have got a blockbuster treatment and you've got a huge effect, you're going to be able to detect that with something like the SF-36, but that's generally not our case. We're generally trying to decide whether the treatment benefit in a cascade of outcomes is significant enough to think that we can label the product as improving that.

And so, when we do that, we would really like to be able to say it improves the symptoms or it reduces the impact of muscle weakness. And that's not necessarily going to be done with a generic instrument.

That doesn't mean we shouldn't analyze the generic instruments. And this is still an open topic. But I would say cutting the pie is best done with specific measures.

Also, the question has come up this morning a lot about objective and subjective. And I'm very fond of the difference between the exercise test in the hospital and the day-to-day activity at home. Those are two different environments. And I think we are talking about correlation, but we're actually talking about measures of two different things. This is measurement of physical performance in a tested situation and it's measurement of physical function in daily life.

And I think we want both. It's not one or the other. And we can see how they are correlated. I'm not disturbed by a 0.40 correlation. I think that's okay. That's actually about right, between 0.40 and 0.60. Above that you wonder why you need one or the other. You only use the test.

Separately is any kind of a testing can be perfectly negative and the patient still feels horrible. And I think this may also characterize sarcopenia. You may improve the muscle weakness, but they're still having a really rough time. So we need to be careful about that in terms of self-report.

As far as well-defined and reliable, it's all of these things in relation to the concept of measurement or concept of interest within the context of use. And I think we can address every one of these without too much difficulty in this field.

So what is my opinion of the way forward? Well, more multidisciplinary efforts like the one today. And I think the inclusion of more social scientists and PRO persons like myself in your discussions would be very, very helpful. We're not so dumb we can't follow your biomarker work. And it would be useful to have the entire panoply of the outcome measures represented when we're talking about the endpoint model and not focusing only on the clinical. A possible consortia of private sponsors, I would much prefer that to a single company, at least a group of companies that wanted to do this and do it through the qualifications process. After which it's very, very important to point out that once it goes through the qualifications process, it's in the public domain. Okay? That's a very important thing.

I also, secondarily, want to say that just getting qualified doesn't mean you're going to be a success. The product has to have two randomized, well-defined trials showing that you have treatment benefits. So you can have a qualified outcome, but at least it can be in the hierarchy with greater confidence of have a chance of seeing something.

My preferred way, and this may not be possible, would be multicenter development through NIH funding. I'm involved in so many of these at NIDDK and other institutes. Can we not take this as an important part of what's gone on before and find the funds to do that?

There is a way through incremental incorporation of existing measures through exploratory endpoints. And you might have the chance of doing that. It's a very high hurdle to get a generic measure into the label. And we remain with the need for what we can say is specific to sarcopenia whether it was developed for a generic use or for a specific use.

So I conclude that the science of measurement is the same for all types of clinical outcomes assessments and the patient-reported outcome will augment the endpoints in the hierarchy used to evaluate treatment benefit in this condition. A PRO cannot be chosen or developed without this well-defined context of use and

targeted concept of measurement based on understanding of the condition. Thank you very much.

[Applause]

Dan Perry: I want to thank Dr. Patrick for a very interesting and telling real overview and summary of the challenges and a look ahead at possible use of multi-company consortia and multicenter efforts. And, boy, are you punctual. Right on the button. That's incredible.

We have about 15 minutes for questions and answers, so please come forward.

Dana Hardin: Hi. I'm Dr. Dana Hardin from Lilly. Thank you, it was a great talk.

I just wondered, as we look toward sarcopenia, muscle wasting in general, could you foresee, perhaps, a PRO where there are core questions that are applicable across multiple disease states, but then, perhaps, individual disease states have additional things that need to be asked? Is that something that you see?

Donald Patrick: Yeah. I think we already have the evidence that that's the case. The core concepts are already contained in the generic instruments. The question is how specific do we need to get to this? And we focus on one indication. And people come in packages of coexisting conditions. So I really enjoyed the conclusion that we really have to pay attention to this comorbidity issue, because there's so many things that can contribute to a loss of physical performance in daily life, not only the environment, but I just told you about my tendons, arthritis, diabetes.

And, by the way, there was one point that I – I tried to get all the points that I wanted to make down here on my list, but there's one I missed. In 1970, I wrote a commentary to a paper on how self-report can affect HbA1c. We need to remember that it's not a linear process and that there's a feedback loop. And, of course, what we do affects our muscle mass. So the self-report feeds back into the earlier endpoints and this is not a strictly straight, linear process.

Bill?

Bill Evans: Yeah. Very nice talk. And just a couple of things to say about when we thought about this PRO.

Number one, sarcopenia has typically been described as an amount of muscle that predicts some other thing. The problem is is that no clinician ever measures muscle. So going into a clinic and even looking at all of the patient records, nobody measures muscle. And, secondarily, the geriatricians in our department of geriatrics don't measure function in a systematic way, because they don't get reimbursed for it.

So you have patients coming in that we know are frail or weak, but there's no systematic way to evaluate them. And it seems to me that that's kind of the first step.

So as we were developing this PRO, we started off with the questions, considering your low muscle mass – well, no patient knows what their muscle mass is. They don't know how much they've changed. So considering your muscle weakness is a much more important question.

And, as you say, the context is so important. And I was struck by just one statement, although the data weren't published. Looking at patients who are weak, because of cancer, they had lost their muscle mass and strength rapidly. And so their PROs, I guess, their outcomes were really different, because sarcopenia develops very slowly and older people make the appropriate accommodation.

So, for example, there was one question that came up frequently among the weak cancer patients, I have trouble standing in a line for a long time. That never came up for older people, because they don't, so it's not one of the considerations.

So even the context of how rapidly the change has occurred may give a different context to consideration. So, again, that's also comes to the specificity of the disease or problem and how you got to be weak.

Donald Patrick: Right. Boy, I have so many images. I mean I'm in Ellis Island and I'm standing in line looking at those steps. And how formidable that would have been to think about I got to walk up there, because I'm sure that they knew that there was somebody at the top looking at them, and what I can do.

I'm also struck by the imagery. And I drew it out, actually, of the bigger person. Fortunately, in one way, we don't shrink. And what that body mass is appended to. And the idea of that little old,

85-year-old woman with spindly legs really strikes me, because that would be the gender difference and that there would be more strength in the thighs of many men compared to what I've seen in the nursing home. And I just observe this.

And I'm interested, Bill, we don't systematically do this, but the clinicians have got this body of comparative evidence of people that they've treated. And they probably are paying quite a bit of attention to this in a non-systematic way.

And so I guess we need to demonstrate, and, of course, you picked the key issue up, am I going to get reimbursed for doing that?

And I'm struck how the real measurement in my own condition is done by the PT person not by the knee surgeon. All the decisions of treatment are made by the knee surgeon, but all the functional testing and the performance testing is all done by the PT. And they are at separate offices. They're not on the same chart system. They don't share Epic. It's like I have to carry the notes back and forth myself to get them to talk to each other.

So trying to get that performance testing or functional testing within the clinical practice, I think is a very interesting idea and something that really might make a big difference.

Basil Eldadah: Hi. Basil Eldadah from NIA. I'm going to ask you two questions.

The first is following up on the previous discussion. So the instrument that you have is very intriguing. And I just wanted to ask if you could comment a bit more on this idea of attribution? Each question is preceded by asking the subject to consider their loss of muscle strength. And I was just curious if you could comment on how well in your observations were subjects able to attribute these various functional deficits to loss of muscle strength versus something else? Or had you not asked that question at all? Would you have gotten different results?

And if you could also comment on how cognitive impairment might affect whether one is able to attribute their functional impairments to a loss of muscle strength?

Donald Patrick: Huge amount of issue on both of those.

Attribution is a very imprecise piece of information. We know very little about this. It's there, because if we don't try then everything else can creep in. So, remember with the

patient-reported outcome, and this is one of the biggest problems with any kind of a question answered, is that we are assuming that every person interpreted it in exactly the same way. That is, of course, not borne out. That's why we do cognitive interviewing.

So attribution is probably a very poor way of trying to control for that. But if we're trying to get specific, then the context of that becomes important.

And cognitive impairment, the work I do in stroke, oh, my goodness. Physical performance testing in stroke is extremely important, because what we get in self-report is a well-known phenomenon that goes on of their over reporting of their ability to do things.

Basil Eldadah: And if I could just ask you a second question, which is you've been involved in the PROMIS project for a while, and I was curious to know whether there might be an opportunity for something like this to be part of PROMIS or whether PROMIS has anything to offer here?

Donald Patrick: Well, you might have had a PROMIS speaker here. But let me tell you that nine years ago I lost a very major battle, very major. And I tried for months on this. The PROMIS measure is all done in capacity wording. Jim Fries says it makes no difference. It may not make any difference on a population level, but it makes a huge difference with an individual patient.

So all their physical function measures are, are you able to do this not do you have difficulty doing it. So I have a lot of problems with that. We tried this with our patients with HIV and gave it up. We're going to use something else.

But there's a huge battery of PROMIS measures that are domain-specific. Again, you would need to look at the content validity within the muscle weakness population. Just because PROMIS did 2,000 people in an Internet sample doesn't mean they have covered people with sarcopenia.

In fact, with HIV, we discovered at least in the field of depression, at least, severely depressed was not represented in the PROMIS development and suicide was not even part of the instrument development.

So, again, we have this question of how specific do we have to be?

PROMIS has done a great job with modern psychometric theory of coming through great domain-specific measures. And they will find their way and possibly will find their way up the hierarchy in the same way as the generic measures. It still begs the question of whether they are content-valid for sarcopenia.

Basil Eldadah: Thank you.

Peggy Cawthon: Hi. I have a comment. And maybe you could address it and then we could discuss it in the later session.

I think the problem for attribution is particularly problematic for sarcopenia, because loss of strength and in particular slow walking speed or decline in walking speed is a hallmark of the aging process. You cannot find someone who is 90 years old, who walks the same as they did when they were younger. So it's very difficult for people to disentangle loss of strength from getting older and from walking slower from getting older.

So I wonder if the results from the development of the questionnaire would have been much difference if instead of saying loss of strength, you could have used a phrase such as, as you aged, and if that would be much different.

And cancer cachexia, it's different, because there's a defined and unique definition. So I think that's something that we can consider.

Donald Patrick: Thank you. I think that's great. And following on Bill's comment and that comment, we need a lot of work in this area. I think that's very clear. And I think one of the ideas that I have sitting up here is that we really can't use point in time. We need to look at change. Just like we look at gait speed, we would look at – and how do you do that? Change in your reported physical difficulties over a period of time and it may be rather slow.

Nobody has brought up response shift, a whole big topic here, but it is a very big topic in this area, which is we get used to what we do and we get used to very, very quickly and we don't recognize that it's a problem. That's why we will never do without the performance measures. The performance measures are not subject to response shift. So, that's, I think, important there.

I think as far as attribution, we would need to explore this both with cognitive work and possibly with empirical work. Good survey methodology has been done in survey labs with testing this

thing on small samples. It may not be done in the PRO field, because we tend to work in much larger areas, but survey labs could do that work.

Sir?

Brock Beamer: Yes. Two, first just the comment that, although I'm sure a lot of people would measure more if they got reimbursed, the big reason, what we teach the house staff and students is if you have an action that will change. So if we ever do get a drug, people will measure whether they get reimbursed or not.

Donald Patrick: But I have to come back –

Brock Beamer: I agree that money helps.

Donald Patrick: I have to come back on that one.

Brock Beamer: Yeah.

Donald Patrick: I really have to come back on that one. I happen to be involved in a whole exome project, okay, which means that we have, what, 3,000 possible and incidental findings to feed back. And I want to the committee and they said we're not going to feed back Alzheimer's risk, because there's nothing you can do about it. I am sorry. That is probably the single most important thing to older adults. So whether or not we're reimbursed, it's important to the patient and we pay attention to what's important to the patient or they'll leave your practice.

Brock Beamer: Right. Right. I guess I was just trying – right.

Donald Patrick: Okay.

Brock Beamer: The point being that reimbursement, money is nice, but there are other motivators.

Donald Patrick: Right.

Brock Beamer: Not many, but others.

I'm going to try to rephrase the question that I bungled earlier. If we have a PRO, and we wanted to show that, yes, there's a meaningful change and I recognize we need to do that, can it be in a spectrum of older people? Or does it have to be in people who have met the definition of sarcopenia?

Donald Patrick: Okay. I –

Brock Beamer: And you will get reimbursed if you answer.

Donald Patrick: Yeah. No. I think yes and yes. You see, I think if I had the dataset that I'd like to analyze, I'd like to analyze change in physical performance as related to change in self-report, and that we shouldn't be doing these cross-sectional looks nearly as much as we should be doing longitudinal. It's probably being done at NIA. I can't imagine that we don't have the possibility of actually doing that analysis. To have some idea of the correlation. And, again, we're not trying to correlate. We're trying to correlate two measures of different things. It's not trying to find out whether they're the same thing. Okay?

I was thinking about whether I wanted to do Bland-Altman plots on physical performance and self-report. And I thought no, because we're not talking about two measures of the same thing. We're talking about two measures of different things and highly related they are.

So you could possibly find the cut points in change in physical performance to where you get a cut point in self-reported physical function. And we're not going to be able to interpret physical function with all the environmental possibilities that affect it without physical performance measures.

So I would say the endpoint model in this condition requires every single one of those things in order to have a really good evaluation of the treatment.

Did I get another rise, Bill?

Bill Evans: No. I think that there is real concern about sarcopenia. If we have a medicine that treats a specific patient population and the outcomes are good, there's real concern that now everybody that's over 70 is going to want to use it. And that's a genuine concern –

Donald Patrick: Oh, oh, oh, just like EPO. Well, boy, I would like EPO every day.

Bill Evans: No, just like testosterone.

Donald Patrick: Oh, testosterone. Yes.

Bill Evans: But testosterone is approved for men who have low T. And so now you see the ads, are you tired? You might have low T. And so what's happening now, interestingly, is a number of physicians are writing prescriptions for T without even checking their patients' T levels.

And so there is a real concern, and I think that's getting at the question, do you want to treat all old people? No. We want to treat a specific problem that's well-defined and to the extent that we can manage who ultimately gets treated.

But, again, the other kind of example of that are medicines for erectile dysfunction. At first, remember Bob Dole was the spokesperson. And it was targeted specifically for men with ED, and that's not the case anymore. And it's reimbursed, maybe because men are all on the reimbursement boards. But –

Donald Patrick: I do want to point out to you that that experience – I've worked for 15 years in premature ejaculation, which is equally an issue as erectile dysfunction. There's no approved product for good reason, because it's not specific. And everybody is waiting for it to be used by everybody in the population.

So I really appreciate that last comment, because I think that would tell us we really need to be specific.

Dan Perry: Any more questions for Dr. Patrick?

[Applause]

Dan Perry: Very nice. Let me say one quick thing on reimbursement.

The AIM Coalition simultaneously is in touch with officials at CDC and is working on the ICD-10 coding to explore reimbursement for that first generation of interventions targeted at sarcopenia. So we're doing both things and recognizing that that's the way that you really improve health. So that's probably another conference like this. We won't go into that today, but I wanted to tee that up.

We're going to take a ten-minute break. We have some refreshments in the hall. Please come back promptly in ten minutes. We have a terrific panel. We have a half dozen speakers – some you've heard; some you have not yet. It will be headed by Dr. Hardin. So please come back in ten minutes. Thank you all.

[Break in the conference]

Cynthia Bens: Hi, everybody. I'm going to ask you all to please take your seats.

Dan Perry: If everyone will take their seats, we have some information on taxicabs that you will not want to miss. So, Cynthia,

Cynthia Bens: I am making the important taxi announcement and it is very important. If you signed up for a taxi with our registration desk, you need to actually go to the concierge desk at the hotel and speak with them, because the taxis that were preordered have the destination they're going to. And so if you get in the wrong taxi, they will not ask you where you were going and you may find yourself at the wrong airport. So, please go to the concierge desk and let them know what meeting you were at and your name and they'll make sure that you get to the right place. So, thank you.

Dan Perry: We now are going to hear from our panel headed by Dr. Dana Hardin. Dr. Hardin is a physician in Global Regulatory Affairs at Eli Lilly, and prior to that was a full professor and section chief in pediatric endocrinology and metabolism at Ohio State University School of Medicine. So now she's making the leap from pediatric metabolism to the other end of the lifespan, and we're glad to hear that.

So I will let her introduce the rest of the panel. But for the next hour and some, we will hear a rigorous discussion on some of the key questions that have arisen from all of your work and from our discussion so far today. So let me turn it over to Dr. Hardin.

Dana Hardin: Hi. This is exciting. This is going to be, hopefully, very dynamic. Before I introduce our speakers, and we have thought of a few questions for our esteemed panel, but, please, we didn't think of everything and we certainly welcome you all to chime in, ask questions, answer questions. It's just a forum to open this up for dynamics.

I do hope that we get everything on this list and more answered. So if I find that we're being real circular in a discussion and we've already hit the main points, I'll probably move us on. I will try to do that very respectfully, though.

At any rate, we have already had the opportunity once to hear from Ashley Slagle from the FDA. We're excited to have her. She is part of the SEALD group and that's very invaluable to us today.

You've met Donald. He did an awesome job. Jack, awesome job.

Two new people that have asked questions, but not necessarily been on the podium, William Dale from California and William Evans. William Dale is from the University of Chicago and Bill Evans is from GSK.

So, with that, I'm going to put our FDA specialist on the hot seat for the first two questions. They're kind of linked together and that will kind of be a leap frog. So I'm sorry, Ashley, that you're on the hot seat, anyway.

First of all, can you tell us about some, perhaps, PROs that are already approved by the agency? What you felt facilitated their approval? What this group can learn about that approval process from them?

This is the second half of that question, but we know you all have an approval process. It's been outlined. Do you foresee any changes with that since you now are searching for a new director? Yeah.

Ashley Slagle: Okay. Those are a lot of questions. First, I need to remind everyone, thanks, Cynthia, that I'm speaking for myself and I don't necessarily represent official positions of the agency.

So when we talk about approved outcome assessments, there are two pathways that we can advise on outcome assessments and endpoints. And we heard about this earlier with Dr. Guralnik's talk.

There is the typical individual drug development program that comes in under an IND and then an NDA is submitted. And we can advise on outcome selection, outcome assessment development, endpoints. And SEALD is consulted through the division to help with some of those questions. So that's one pathway.

The other pathway is through our formal Drug Development Tool Qualification Process that's managed by SEALD for the clinical outcome assessments and that's outside of the process of an individual drug development program.

And so through that qualification process, we've actually only qualified one instrument so far. It's a relatively new process. And that instrument was the EXACT-PRO, which was qualified in

January of this year. It's for acute exacerbations of bacterial chronic bronchitis and COPD. And I will tell you that that was a fairly long qualification process. I think the person who leads that, Nancy Kline Leidy, said that it took her longer to do this than to finish her doctoral degree. But we learned a lot going through that process.

So I say that not to scare everyone, but to say that we learned a lot about the process and working with multiple groups who are now going through the process, we've identified ways that we can improve the efficiency of developing tools.

One of the key things that we've discovered that needs to be determined early on is what are you measuring or the concept of interest and in what population?

And I think that with most of our projects that has been more difficult than actually developing the outcome assessments. Because developing the outcome assessments, especially for PROs that we've thought so much about the methods for that, that's fairly straightforward. There are scientific questions that have to be answered, but that process is not difficult. It's determining who your population is and making sure that you're developing an assessment that's appropriately targeted to that population.

Because we're developing this outside of a particular drug development program and, ultimately, FDA would be qualifying this instrument for use across multiple programs, we do expect that these instruments should be a bit sensitive, more specific.

And so in an individual drug development program, you might use a very generic measure. Your drug has a huge treatment benefit. It's detected by this generic measure. We have to figure out how to label it, but we know something good happened.

But when we're developing a new instrument through the qualification program, we don't know if all of the treatments are going to have huge treatment benefits, so we need something that's specific that can pick up even small benefits that we can then weigh against the risks and make approval decisions.

Dana Hardin: Changes with new management.

Ashley Slagle: Changes with new management. I don't think we're going to see a lot of changes with new management. So Laurie Burke, who has been a leader in this field for many years, retired as the director of

SEALD late last year. And we're still searching for a replacement for Laurie, although no one can really replace Laurie. Right now, our acting director is Sandy Kweder, who is the deputy director of the Office of New Drugs, and she was Laurie's boss, so there's some continuity. I mean Sandy has been engaged with SEALD for many years. So, right now, we're just moving forward as nothing has really changed for us internally. We do continue to seek a replacement director, so if you know of anyone that might be good, send their resume. But, no, I don't foresee any major changes in the way SEALD operates or the principles of good instrument development that we are proponents for.

Dana Hardin: In a space like this where we don't have a pure definition yet of the condition like sarcopenia, what are your best thoughts on moving forward for this group as we would like to get somewhat of the drug development pathway instruments?

Ashley Slagle: I think there are a couple of things you can do. Within an individual drug development program, talking to the division, the medical review division at the agency very early is critical. Unfortunately, a lot of times we see that, for some reason, sponsors don't come and share a lot of their plans with the agency early enough for us to provide advice and help move the path – move forward.

So, if you're thinking about doing work in this area, even in the Pre-IND stage, if possible, start talking about who the population is and what you think you would like to measure.

Through the qualification program, a lot of conditions are not very well defined and we have to find a way to move forward in these areas. One of the things that you can do is narrow down the population by a common set of either symptoms or onset of the condition or whatever is relevant to make a fairly homogenous population and develop an instrument for that population. And then that doesn't mean that we can't later expand the qualification when we find that the instrument is appropriate for other groups as well. But starting with a very narrow population can sometimes help get the instrument developed and we can learn more about the instrument and then possibly expand its use.

Dana Hardin: Anyone else on this panel have specific questions regarding what we just talked about?

William Dale: I have one. So my setting is we care for cancer patients and this issue of comorbidities has come up a couple of times. Even the

PRO that I was fortunate enough to work on with Professor Patrick and Bill Evans back in the day, we had a population which was men with prostate cancer treated with androgen deprivation. They clearly experienced these strength losses related to their treatments. It's very well defined. And it looked like a promising population for a potential drug in part because it was something that happened on a shorter timeframe than say sarcopenia typically occurs.

So a long of way saying is a better strategy to join those of us in different specific areas to the larger effort that the larger group would be helpful? Or should we separate ourselves, pursue a very specific application for a particular indication, because it's easily defined, the population is clearer? So I guess that's my question.

It never went anywhere after our initial efforts for a variety of reasons, but I'm curious what the FDA's view is or your view.

Ashley Slagle: Unfortunately, I'm not familiar with the effort that you're talking about, so I can't comment on that specifically. I guess I don't have a great answer for this. Depending on what your goal is, if your goal is to quickly develop a measure that's targeted to a specific population, then narrowly defining your population might help facilitate instrument development. But then you have to weigh that against, well, is that really going to be generalizable to a broader population? And should we have started with a broader population to begin with? May make the instrument development effort a bit more challenging, but I can't answer what's a better process or pathway. You just have to weigh the pros and cons.

Bill Evans: Just maybe to put a little bit more context, I don't know if it'll help you answer the question. The PRO that was published was part of an overall effort on the part of Amgen. They had a specific drug, an anti-myostatin, monoclonal. It was a peptibody to treat muscle wasting.

And so they had identified a couple of different patient populations, but the first patient population that they were interested in was just what Bill talked about, which was men with prostate cancer treated with testosterone ablation therapy. So it was very, very specific. They experience muscle wasting as a result of the testosterone ablation, probably not because of their cancer.

So I guess my question to you is we have a mechanism of action of the drug, we have a specific patient population that's defined, and a PRO that was developed with that patient population, is that kind

of what you would be looking for in the development of a PRO? Maybe that's the question.

Jack Guralnik: Can I just follow that question? To add to that, is it worthwhile going through the qualification process for something this specific when you can just do the clinical trial and kind of do it the old way? The qualification process, as you say, is long and arduous. And you're not guaranteed that you can go beyond that context of use if you get qualified for something that specific.

Ashley Slagle: Well, if it's qualified for something specific, that doesn't mean that it can't be used in a different context of use. It's just not qualified for that context of use. So, in that case, the instrument would still be publicly available and we would suggest that you speak with the agency about using that instrument in a different context of use.

So I think there are still benefits to going through the qualification process with a very narrowly defined context of use so that the instrument becomes publicly available, we can learn more about it. People may use it outside of the context of use with a little bit of additional risk by them, because we haven't actually reviewed the data and we don't know that it actually works outside of the context of use. But getting more people using it I think can only add to the body of knowledge.

Dana Hardin: Any from the audience on this particular topic or expanding on that discussion? If not, I'm going to move onto another question.

And this one is – there's some good expertise in this room on PROs. And I guess we have the example provided for us, and that was very helpful. But do you feel like there have been – and this is open for all of you to answer – have been adequate patient input to the PROs proposed for sarcopenia or used for specifically like the self-rating of health? Have we solicited enough from the patients?

And then the thing that came to my mind – it's kind of, again, a two-part question – was, as I looked at the PRO that was circulated to us, it was pretty clear that the demographics of that population were fairly similar. And I wondered what the effect of a different race or socioeconomic group or education group might have had on those PROs that we've seen so far? So I'm opening that big old can of worms for anyone who wants to speak up first.

Donald Patrick: Was that a hint, Ashley?

Ashley Slagle: I was told to hand it to you.

Donald Patrick: Thank you. This population, as Bill just told you, was a very specific population. And I would think that we are having all sorts of problems at defining what is our target population in this discussion, because I see it as all older adults in its largest definition. Then we get down to a particular cause, an androgen deprivation was the context within prostate cancer. So that's not only one cancer. It's one treatment-related issue, which is often very common.

So I would say that we have to be very clear and that given what we've heard today about men and women that would be a huge issue.

I'm not so sure it's worthwhile going out and talking about self-rated health. I mean there's enough information about that in every population. I doubt whether it's going to be thought of any differently, but that's not the issue.

The issue is what is specific to the experience with this particular slice of the target population.

Let me also give you my opinion about qualifications. Qualification is a really interesting idea and its original notion was to cut the number of instruments that are used in the area. But when you think of the different target populations and the different contexts of use, you're likely to be in a little bit of a position of having to do some more work even if you have qualified instrument. It doesn't mean that it's going to be totally transferrable to your population. And I think EXACT was a good example of that.

And I want to raise another issue that was here this morning and that is treatments that want to prevent deterioration versus treatments that treat or try to improve. Now, in both cases, we're probably trying to slow the decline. And I don't know what's possible in terms of improvement. I mean if you can make people move from the far end of those videos to the left side, that would be wonderful, but I suspect we're trying to prevent them from getting too slow at that. That's a very different use of an instrument.

So that came up with EXACT. Are we trying to prevent exacerbations? Or are we trying to have a measure that evaluates improvement with exacerbation?

And I'm going through the qualifications process now in cystic fibrosis and the same thing comes up. So I'll just give you a little bit of a hint. I was asked was the qualitative research different for those who were in exacerbation, for those who were in a stable state? That's very interesting, because, in fact, it was exactly the same concepts. There wasn't anything new in exacerbation. It was just a different range.

And that may be true here as well. If our target population is very narrow, particularly on a severe group, and a good temple-induced learning is you start with the severe group and try to improve them. That group is quite different than aging population in general.

And I'm also struck, and I'd love to hear from that side of the table, how long is the period of observation necessary to see the benefits for different types of treatments?

Bill Evans: Maybe just a couple of things to say about this. Obviously, it depends on what the intervention is to see the effects. But I was brand new to this process when Amgen began to think about developing PROs. And the development of three specific PROs for muscle wasting was initiated on the advice of Laurie Burke and the FDA and based on their recommendations for specific patient populations that experienced muscle weakness.

Now, maybe, qualifications have changed, but that's a very difficult, expensive, time-consuming process and then to validate and then qualify the PRO is even longer. And I'm wondering, to get at maybe what Stephanie asked, there are a number of extremely well-validated PROs that have in them a number of questions related to function, which is ultimately what we're trying to deal with.

And I guess part of my question is, using already validated PROs, can we look at subsets of those questions to see if they discriminate among different conditions? Because when we're dealing with geriatric patients, every patient is different. Every condition is different. It's really hard to think of a geriatric patient that is sarcopenic that doesn't have diabetes or doesn't have hypertension or doesn't have some other condition. So to be really specific about a sarcopenic population is an exercise in futility I think.

Dana Hardin: I appreciate those comments as well. I think one of the things that if we were to, again, with a goal of going for at least one type of approval that could be put across sarcopenia, I want to hear what

you all think about the need for us to be more specific or get more input on ethnicity and gender, for example. What is the expertise here? Stephanie, please chime in. Anybody in the audience, who's had this experience, as we move forward with this goal, which is a goal of AIM, what do you think about that? And where should we get that expertise? From patients of what demography? What do you want to do?

William Dale: I have a question and a comment.

So another part of my life is to work on a nationally representative survey of older adults. So there's a difference between a large, longitudinal, clinical survey that we often see represented in the largest studies. And then this question of which subgroups are represented and how much do they represent anybody else always comes up as one of the challenges.

It seems to me if we want to get cross-sectional data, you would want something like representative samples in which you had the right questions in there so that sub-studies would take advantage of those representative samples with the questions. You say it's different in this way for this group or for that group rather than every study having to redo the baseline comparison for their particular group. I think you'd still need for the change that you're going to need – gait speed changes, it turns out to be really important. SPPB changes turns out to be really important. But you don't need a super big sample if you define it right to get the change information, right?

So, my own example with the SPPB, we're doing randomized study of an exercise intervention using the SPPB change as the outcome. And you need a half point to a point outcome change, basically, to show clinically relevant differences. I don't need a big sample. And it turns out I can say strength training matters more than endurance training in our study with a small sample.

But it depends on having these baseline samples. And I always feel like we mix big clinical samples with representativeness of samples in this discussion. I have no idea if the FDA and the drug approval process cares, but scientifically it's always struck me as slightly strange that we don't emphasize more the representative issue. We just emphasize the bigness issue.

Dana Hardin: What do you think, Ashley? Any comments on that? Boy, we're really making you work today.

Ashley Slagle: Hmm. I don't have anything profound to say. I think that it's always a challenge and it depends on the condition, the patient population, so many different things. We are seeking to understand how this works in the real world and we want our populations to be as representative as possible. But we know that to be able to detect treatment benefits in a very short period of time, that having a very representative sample with all the heterogeneity makes that more challenging. And so it's a constant struggle between having a very narrowly defined population for clinical trials that are sufficiently representative of the broader population.

William Dale: Just to respond to that a little bit. It strikes me that the comment about aging, for example, that all older adults, as Don was saying, are affected by sarcopenia. You're talking about all the older people. In a representative sample, I can say this person represents a group, because they were selected in a specific way. So the questions of representativeness are solved scientifically by picking the sample the right way in the first place.

Our studies, we shouldn't – necessarily is our study, but any of these studies going in the field every year, you just put in the right measure, if you had the measure, and you're going to get a representativeness. That person represents more than themselves. They represent a huge group of people.

So I just think for something like sarcopenia, we're trapped by thinking of this as a drug indication problem when this is slightly different and would benefit from these social science surveys.

Bill Evans: It's usually easier to think backwards as to what it is you're actually trying to treat. And I think unlike cancer cachexia where we're really ideally trying to prolong life, in sarcopenia, I think what we're really trying to do is or at least we focused on preventing mobility disability.

If we take that as the construct, then we don't need all the old people. We actually just need the ones who are at risk for mobility disability. Or we could take falls and say we want the ones who are at risk of falling over the next year.

We have a very hard endpoint there that we can actually make into an indication and at the same time we don't have to have all the payors faint, because we now have 80 million people we want to treat.

So I think that if we use that kind of construct, we use the definitions that we've heard about as an entry point, we have the eventual outcome of preventing mobility disability.

What we're really asking of the PRO is to be an intervening variable that will allow us to measure how a treatment affects the patient while the physical function measure tells us something more objective about how well they're responding.

But I think we're trying to respond to that FDA requirement to both make people feel better and function better. A six-minute walk or whatever gait speed will tell us about function. It won't tell us if they feel better. We got to ask them.

Donald Patrick: Would you mind defining mobility disability for me?

Bill Evans: Well, the simplest thing is the CDC definition, right? Can you walk 400 meters?

Donald Patrick: Okay. So, again, I have to – please call that ambulation.

Bill Evans: Okay. Fair enough.

Donald Patrick: It's walking. It is not mobility. We need to be very, very clear on that.

The other thing that has struck me is I'm not so sure that we have deconstructed the actually body movements sufficiently affected by muscle weakness. So these would be almost universal across conditions meaning – I'm going to give an example that it probably doesn't relate to, but it's one that interests me – and that would be turning over in bed. That's something you can't do easily on a performance measure, but is extremely important to the patient.

What of those, if I might call them, biological body movements or the movement of the actual body parts are related to muscle weakness and must be examined by a PRO?

And I would like to see that proximal analysis done before we put in the problems of the environment, because one of the problems that the agency has is that they can't label this drug as improves your work in the environment. I mean it cannot say something – that's why quality of life doesn't work, because the drug is indirectly improving that, but it is not the direct benefit that would be in the package insert regardless of the evening news ads, which

I agree that's a promotional issue separate from what the actual benefit of the treatment is.

And I think we may be moving from something that's very proximal to something that's very distal without that territory in between.

Stephanie Studenski: I just wanted to briefly mention. In the context of mobility has been known for 20 or 30 years that is probably one of the most hierarchical of behavioral phenomenon. And these things like Rasch analysis and use of item response theory are very robust, because of the hierarchical nature. So you can – I mean I've been interested in this, because in aging, for example, among non-ambulatory people, the amount of care somebody needs, it matters a lot whether they can turn over in bed.

Donald Patrick: Yes.

Stephanie Studenski: It matters a lot whether they transfer. But there are very strong hierarchical properties. And there's extensive literature on this.

So the PRO can be robust beyond ambulation to the lowest to some very high things. And I think Jack's mention of the videos, there were hierarchical concepts –

Donald Patrick: Yeah. They used IRT to do that.

Stephanie Studenski: So from the point of view of the discussion we were having, I don't think the mobility measures have ever been thought of as only walking.

Bill Evans: Just to follow up. To get at your point, one of the most frequent statements among those weak is that I have trouble standing up from a toilet, right? So, that, most geriatricians know that. That is maybe not totally and completely related to muscle strength, but it's a big component of that. And so one of the questions is I have difficulty standing up from a chair, which gets at the same sort of issue. So I think that, obviously, it's great to hear from weak, old people.

Part of the problem with a diagnosis of sarcopenia will always be, I think, trying to relate that to muscle in some way, because it's hard to measure how much muscle you have. It's hard to measure how much muscle you have in your legs.

And I think that we've thought about sarcopenia using methods to measure muscle mass that don't actually measure the tissue that we're trying to measure. And we're making a lot of conclusions based on, probably, at least, partially inaccurate data.

And so, maybe, we need to move much more towards measurement of strength or function with the assumption that some of these measurements, like standing up out of a chair, are pretty closely related to how strong you are.

Ronnen Rubenoff: And I think just as you're getting your microphone, I think this also gets to the ranges that we're trying to measure. So people who can't turn over in bed, obviously, usually can't walk either. So we have two issues here. What's the outcome we're trying to fix? And the other one is what is the range in which we're trying to measure for the PRO?

We work a lot with muscular dystrophies and other kinds of degenerative muscle disorders where getting out of bed and getting out of our chair are the overwhelming things. Walking is not even on the range.

So I think we are going to have to pick a range of interests for any given program or any given outcome.

Dana Hardin: So kind of along those same lines, do we think that any of the existing PROs could actually show a treatment benefit with some degree of sensitivity specificity in this group of sarcopenia, whatever that really means? And, if so, what would those be? And, if not, where do you want to start with making a new one?

Jack Guralnik: Well, they certainly can show a benefit. I mean from the observational work that we've done, we see that both performance and self-report can improve after say an exercise intervention. There are lots of examples where self-report improves. And, again, it depends on the specificity of the measure, but there are kind of general outcomes that are quite responsive to interventions.

So I don't think it's a problem of whether self-report is responsive. I think it's how you construct this measure that somehow relates to the sarcopenia indication. And it's difficult there. This idea, kind of moving along the pathway from where Bill was talking about, actually measuring strength is then in PROs asking the attribution question, due to your poor strength, and you did that in your scale, but it's very, very difficult to do, because almost nobody, as you say, just has a pure strength problem.

Stephanie Studenski: Can't hear you.

Jack Guralnik: It's very hard to – can you hear now? It's very hard to do that, because almost nobody has a pure strength problem. If you've got pain and weakness, you can't sort out how much the pain is having an effect on your function and much the weakness is having an effect. So we may have to get beyond demanding absolute specificity, because older people just aren't that simple. And we'll never see kind of pure cases or rarely see pure cases of sarcopenia without any comorbidity.

Donald Patrick: I would agree with that. I also think this question of standing up from a chair, that has to be done in a performance measure, because I can stand up from any toilet that's in a disability room, but I can't in a modern European thing that's nearly to the floor. And I've been wondering for a long time, what do you do in India or where you don't have a toilet and you have to squat? And do they have the same problem?

But I want to say a couple of things about the PRO question you asked. I think one of our big issues is that we're looking for some sort of magic bullet that comes in ten items. And I'm not so sure that's the case. And that a good examination of a more multidimensional and detailed –

So I mentioned the Sickness Impact Profile. It doesn't have one category of mobility. It has a mobility and that's defined as moving from place to place. It has an ambulation category and it has a body movement category. Those three things have been mentioned this morning. It's not one domain.

So if we can figure out the domain and we can figure out the hierarchical aspects of that – and I think we have to separate what I call mobility from anything else, because you can be very mobile and be very incapacitated – that then we would be able to look at that hierarchy and we do have the techniques and we've had them since Leon Guttman developed this in World War II of looking at hierarchical assessment, and that's simply what these modern test techniques are, if you can walk, what's the probability that you can stand or what is the probability that you can get up from a chair, and examine where in the continuum.

But the domain definition to me remains elusive. And the work that we did in cancer gives me that impression as well. And that we need to drill down to the precise kinds of physical actions that

are going on somewhere in between the performance measure and the broader PRO.

Dana Hardin: That was good discussion. And thank you so much on that.

I was struck by something Stephanie said. And it was that they started their study in the '90s and then they designed it with the relevant ideas then. And then, now, we're in a different space. We've learned more.

So here is a question that we need to think about. Would any PRO that we design today be lasting and durable? And, if so, what are the things that we need to do to make that happen? And I'd love anybody's thoughts on that topic.

When I taught Sunday school, I threw candy at kids if they answered a question. I have some candy here, so –

Peggy Cawthon*:* I think it goes back to the problem about how specific the underlying population is and the mechanism of action for any intervention. So it's theoretically possible that you could improve function without improving strength if you say improve reaction time, which might somehow fit into this domain of sarcopenia.

So then if you ask people how does your loss of strength affect some other outcome, they might say I don't have any loss of strength, but I'm faster. And that might improve their daily function, but the questionnaire then wouldn't be sensitive.

So to have it be long-lasting, you almost by definition need it be less specific. But then if it's less specific, then it's less likely to be approved by the FDA as a global instrument potentially.

So I think there's a tension between those two concepts.

Dana Hardin: That's a good point. Jack?

Jack Guralnik: One interesting way of thinking about this, there's been a decline in disability through the '80s and '90s and it may be turning around a little bit. But people ask the question, is that because people inherently are healthier and functioning better? Or the way we ask the questions, it's possible what we could miss is that they're using assistive devices and therefore they're not disabled and so the trends could be explained that way.

And there's now a terrific study called, "The National Health and Aging Trend Study" that is really trying to sort that out and look at whether assistive devices are explaining some of the trends.

But as long as we take that into account, these kinds of measures can be durable. And we feel like that we can truly study trends. And, hopefully, into the future, while we know assistive device use will become more common, we can pick up on that.

Donald Patrick: I just quickly have to make a note from the perspective of Baltimore that as disability has gone down, requests for disability compensation has gone up.

Dana Hardin: In the future, and we all know as physicians that patient reports to us of how they feel with anything we prescribe is real important for us wanting to re-prescribe the same regimen. If we look at the importance to the agency, we get a good feel that you all are interested in patient-reported outcomes. I'm not so much sure about payors. And I ask that, because, yes, he's coughing, it's true. They're not so interested. Should knowing all of that, should a PRO be a primary outcome measure, a co-primary outcome measure, should it always be relegated to a secondary outcome measure? What are the thoughts on that?

William Dale: I'm just going to add to the question, because it fits with what I was curious about.

So, it seems to me that we're trying to get from sarcopenia or muscle loss or muscle strength loss to these functional issues. And as Stephanie laid it out and the logic of we tried to go from what does the clinician think about when they're thinking about these patients and getting from some very specific physiologic process to that.

The PROs, I still am unclear. Are the PROs a different kind of outcome? So there's the specific outcome of SPPB or gait speed. And there's another phenomenon, a different measure that's an outcome that's the patient-reported experience. Or is the patient-reported experience – it strikes me as capturing something different, mediating something between the muscle and the functional issues that they encounter in the environment, even setting aside the environment issue.

I actually don't know the answer, but I am struck how often self-reported health is an example, somebody tells us something. The only way to measure it is to ask them and it's way better than

all the other measures we have that there's something about what they know about themselves that's way more important above these other measures that makes me think they're separate but linked. I don't think they're unlinked.

So I'll throw it to Don, because I'm still unclear about that from the FDA perspective.

Donald Patrick: Oh, boy, I really love that. It was worth coming to Washington just to hear that.

They are absolutely linked. And we just need to investigate the linkage. And it's the PRO that is the highly important report that comes from the patient. I mean the clinician doesn't ask how you doing when they come into the room for a stupid reason. It's basically to get a patient report precisely of their global evaluation of how within all of their life and their environment they're being affected by the condition. So there's a huge number of intervening variables.

But I would say we look at this cascade of endpoints and we think of them as all contributing different pieces of information to the evaluation of treatment benefit, and that we need to look at the links, theoretically. That doesn't mean that we have to have a link. I cannot imagine that changes in physical performance will not be linked to self-report. If that's the case, then the environment and the compensation is so great that somehow we're having a problem, and that's why we have to drill down the questions to be absolutely precise about what they actually do do in that.

We have all sorts of problems in the PRO field of having an anchor when we do construct validation. And the anchor in this particular area would be the performance measure.

So I think we have to keep at these as being very linked and think of how important it is to have that self-reported change in mobility disability – I'll give you that, I guess – and how it's related to all these other things.

And I love the slide on – I think it came from Stephanie's study on how the global – self-rating of health. It was called global on that slide. How the self-rating of health tracked the other measures. And that's the knowledge base that we need from epidemiology and from treatment trials.

Dana Hardin: We've gotten through the questions that we all thought of ahead of time, but I know that we only thought of some. And so I'd like to open it up. We have five minutes, is that right, maybe? We have no minutes. Twenty, well, wonderful.

Dan Perry: We have box lunches waiting for you. We can break now and you could go out and help yourself or you can – we're going to be wrapping up after this. You could take one with you.

Dana Hardin: I'd like to know other questions that people might have thought of and open up. I really would love a dynamic discussion within the room, because we're all just trying to learn this field, so, anything.

Stephanie Studenski: I think everyone would agree that we want to have the patients' perspective as part of benefit of treatment.

And so I'm interested in opportunities to think bigger than just United States. I know that there's a great deal of interest in Europe and I think in other developed parts of the world to be addressing similar issues. There are regulatory agencies that have some similar ground. So, for example, I'm aware of a big initiative in Europe that's going to be trying to explore some of this space.

Are we going to do anything with PRO development in IMI? And should this group be informing ourselves about common ground and where we can do some things here that parallel PRO development there?

Bill Evans: That's a really question. I don't know how many people here are familiar with the IMI initiative, the Innovative Medicines Initiative. Through co-funding by industry, the European Union has committed 24 million Euro, that's Euro, towards an intervention trial to see how much we can move – it's actually based on life to see to what extent we can move SPPB scores. And, in particular, the goal there –

The EU has a slightly different take on this or the EMA, European Medicines Association. They, as you probably know, published a paper in the *New England Journal of Medicine* about a year and a half ago saying that they consider the concept of frailty as important, treatable, with potential important societal effects.

And their goal there is to come to some consensus about what that is, what frailty is, so that that could be used as a basis upon which

to diagnose this problem that we all kind of are dancing around, which is weakness and limited mobility.

So, next week, we'll finalize what the intervention will be and what the measurements will be. And part of that process is to determine appropriate biomarkers for frailty and how those biomarkers might change with the intervention.

But as far as I know, and I'm pretty sure, there is no part of that that is designed to use a PRO.

Stephanie Studenski: That's a tragedy. You're going to have this wonderful intervention. You'll be measuring body composition. You're going to be measuring strength.

Bill Evans: I'll make that recommendation.

Cynthia Bens: The only thing that I would add, sarcopenia is only part of, as I understand it, the frailty definition and so it's much broader. And I would suspect that if they added any sort of PRO development to that that it would capture a much broader spectrum of things that you would then need to draw from to maybe develop a composite tool that was appropriate for use in a sarcopenic population. I guess that's how you would extrapolate that.

But my comment was actually not related to EMA, but rather to the comment I think that, Bill, you made about the expense of going through the process of developing these measures and how challenging it is. And I just look back on the work that Dan referenced that we've been doing in Alzheimer's disease and there's been a lot of money thrown into Alzheimer's drug development based on what we all believed we knew about the disease. And I think that at this point, now, we all realize that the disease is much earlier and everyone wants to intervene earlier. And we're sort of faced with a disease definition problem and what the subpopulations are and also an instrumentation problem, because the instruments that newer instruments are being developed off of are all for a much later population.

And so I think that this discussion that we're having now is really timely and important and we work through some of these issues now rather than companies going and spending a lot of money on these trials and then ultimately having a definition and an instrumentation problem. So that's just my caution.

Bill Evans: Yeah. And the other thing. I have made a recommendation that we try to use this instrument at least instead of developing a new one to see, because this instrument was developed but never tested for an intervention. It would be interesting just to see if these questions actually moved to any extent with an exercise intervention.

Ashley Slagle: I want to make a couple of comments. One, I'm afraid I left you with a very bad impression about the length of time for qualification. So the EXACT did a take a while.

But there have been other examples, not through the qualification process, but using the same development process looking at the Jakafi product that was looking at symptoms of enlarged spleen. That was developed, I think, in less than three years and supported labeling claims. There are other examples.

We're working on over 30 projects in our qualification process right now. And they will be done in much faster than 7 years, maybe 3 years.

In other cases, they'll be done faster, because they started with an instrument that was already developed and there's some gaps in the knowledge that we need to have to be able to qualify this, but it doesn't take as long as starting from scratch.

So I think there are some opportunities there. And it's not a horribly arduous process.

The other thing is that the benefit of the qualification process is that we do collaborate with the EMA. So if you're qualifying an instrument through both the EMA process and the FDA process, that gives us a pathway to talk with the EMA openly and it helps to – we might end up with separate opinions or advice, but at least we're able to harmonize to the best extent possible to make sure that we're not at least giving conflicting advice. So I think that's another benefit to think about with the qualification process.

Donald Patrick: I think it's also worth mentioning that qualifications both comes from a lot of different sources. So I don't know how many projects there are in the Critical Path or the PRO Consortium that are groups who have formed already, but seven of them. Okay? So it's instructive to look at those seven.

It started out with IBS, which is having all sorts of trouble, which won't surprise you, because they can't define the condition. And

so it is in the same sort of problematic with functional bowel disease.

There's lung cancer for which we have a lot of preexisting instruments.

But the one that blows my mind is that there's one on depression. Well, there must be how many different depression instruments? A thousand almost out there and they're creating a new one for qualification.

So I don't think that we can easily think that the medical product evaluation context takes instruments developed for other reasons and applies them easily within this framework. I may be wrong.

Dana Hardin: One thing I wanted to ask about qualification as I was thinking about it is, is there an advantage or disadvantage given what we'd like it to be as a complementary – if it's exactly the same as a PBM, we don't need a patient-reported outcome, right? So it needs to be complementary. But in the validation process, is there a role for real-time validation? So the patient does that four-minute walk and immediately afterward takes a tool that asks them how they performed. Is that something that would move the field along a little bit?

Ashley Slagle: It might. I don't understand the question about it's the link to qualification, because I think that the same information and support for instruments and what's being evaluated and whether it's clinical meaningful and how to interpret change are all things that we have to think about whether we're going through the individual drug development process or through the qualification process. So a lot of these good ideas to accumulate evidence could be used in both qualification or in an individual drug development program.

Dana Hardin: So what tips the balance again?

She's been such a good sport. And then she also has to pass the microphone. Have you noticed? I don't think I'd want my name on the FDA right now, right? It's uh.

But I wonder is there a gold standard if you think about validation? I know our company has tried a validation trial and they're difficult. And so I guess we're just picking your brain about what makes it seem more valid to you, more robust, more certifiably good besides reproducibility and those things?

Ashley Slagle: I think one of the biggest challenges people face is content validity. So what are we measuring? What are the domains that are being measured? Are we really measuring those things? Is this clinically meaningful to patients, to clinicians? How do we interpret meaningful change?

And reproducibility is pretty straightforward, I think. It's some of the other issues that are much more challenging.

And I think Donald mentioned this that we don't quite yet understand what domains we should be measuring fully in this very broad patient population or narrow patient population, however we define it. I think that's the real challenge here.

Qian-Li Xue: Hi. I just have a couple of comments to make. So, I heard a lot about the sort of the importance of specificity in selecting of PROs or any performance measures. So we pay a lot of attention to the comparison between different measures. But also I think that maybe it's useful to also consider how we can improve specificity within a particular measure. So, in that sense, I'm really glad to hear a lot of talking about importance of measuring change at the individual level and how that individual level change can help improve the specificity of the measure, which is typically used cross-sectionally and also – so that's one comment.

And the other is, so, if we agree that measuring individual level change is important, maybe, it will be useful to look at how a PRO is related to that change measure. So that's what the – I mean, I guess the comment about the change in performance versus change in self-report. I think it will be a very interesting analysis to do.

And the other comment that I wanted to make is the frailty versus sarcopenia. I think there are similarities in terms of challenges in defining sarcopenia versus frailty. I mean I think that – one of the similarities is the definition of what, for example, frailty is. And it's interesting that I would think that sarcopenia would be an easier task to tackle compared to frailty, because at least for frailty we still we haven't sort of drawn a distinction between the construct of frailty versus the manifestations at a clinical level that we now can measure. And so there's a disconnect between the two. I mean I think that that actually goes back to an earlier comment about what exactly are you trying to measure? I mean so we have a lot of measures of frailty currently, but we don't know, at least in my opinion, clearly what we are actually measuring, so –

Bill Evans: I don't want to get into a discussion of frailty, because I think that that could last all day, tomorrow and the next day. However, what I will say is there is no clinical way to measure how much muscle you have. So if we're going to base the definition of sarcopenia on an amount of muscle, there is no tool that you can use to do that, no approved tool.

So I think it's a problem that we have right now. We know a lot about strength and its ability to predict outcomes. We don't know as much about mass.

So if you're going to say that it's easy to come up to a definition of sarcopenia, I would say as long as we can measure muscle accurately and directly that can be done in the clinic, yeah, but we don't have that now.

Dana Hardin: I guess to play the devil's advocate, though. I don't really care how much muscle someone has as much as what they can do with what they have.

Bill Evans: Well, if that's the definition of sarcopenia, it's a problem.

Dana Hardin: Oh, I'm with you. That's the big problem.

Bill Evans: If sarcopenia is only based on performance, that's fine. But as Stephanie, and we all realized, performance, deficit is caused by lots of other things besides low muscle mass.

Donald Patrick: Boy, I really think that's important comment. I would really concentrate on the definition of the condition as much as possible. And the presentation on the biomarker this morning is right where we're at. And if we can find the indicators –

And, also, the IMI study, even if we don't have PROs, we got to know more about body composition and distribution of muscle in upper body and lower body and how this relates to the different tasks that we do, because I think we're talking about the performance of physical tasks and that's somewhere in between your performance test and your daily function.

Jack Guralnik: Let me address something that Qian-Li brought up about specificity. And we've heard about specificity. And it's kind a knee-jerk reaction, oh, yes, we've got to have more specificity. But I want to give a dissenting viewpoint for that.

I mean even in looking at classic interventions for disease, you may have heart attack and stroke as the outcome when you're studying hypertension, but it's also the outcome when you're studying cholesterol or you're studying other risk factors. So that's not terribly specific.

And with all the comorbidity we see in older people, you're not going to get specificity necessarily. And some of the kind of the biggest, most important outcomes, I have more energy, I'm more active, I'm happier. These are not specific outcomes. But when you're talking about patient-reported outcomes, they're the most powerful things that we would like to see.

So just because multiple risk factors or interventions track to an outcome doesn't mean that that's a bad outcome, because it's not specific for those interventions. And that's another way of looking at specificity issues.

Chuck Benson: Yeah. Chuck Benson with Eli Lilly as well. I think that most of the day we've kind of gone around and around that you try and have a definition, but, yet, we don't have an indication. And without an indication and an outcome, it's really hard to decide what the definition is.

So we like to use osteoporosis as sort of our guide here. And osteoporosis didn't really have much in the way of a good definition until somebody had this indication of fracture prevention and then you could go through and say, well, here's the level in which we would like to treat. And you use a DXA scanner and decide whether that patient is indicated.

So it seemed like when we started this morning that everyone was talking about an indication, and maybe we still are, treatment of sarcopenia in the elderly and which the problem with that is, of course, being old isn't a disease. And we think that the likelihood of the FDA saying here your indication is that when you're old, we can treat you with this and you can climb stairs faster or have an improvement in your PRO. We think that that likelihood is vanishingly small.

And so, at Lilly, we actually have a number of drugs, which we think will work for sarcopenia, but we are extremely challenged in finding an indication, because of a number of reasons. But, for the most part, finding the outcome on any of these is very, very difficult.

So I'd like to hear the panel's comments on the indications they would recommend.

Just in the last few weeks, we went through this exercise again for the third time in the last 10 years of all the diseases which are associated with sarcopenia, and there's at least 136. I think that's what we started with. And we've now narrowed it down to the top 30 again and looking at each of these and what we think we might be able to intervene with a drug which improves muscle, but, again, all of these, the outcome is difficult and the regulatory path is very difficult as well.

Brock Beamer: Can I add onto that? Just it kind of hits exactly what I was running through my head that we talked about something that – so this ability to rise from the toilet is much bigger than rising from the toilet. It's the question about the payors and CMS and the reason that people go into nursing home is they can't off the toilet or they can't get out of a wheelchair, they can't get off of a bed.

So the simple question can you independently transfer from wheelchair to bed is both a PRO and a performance measure and a very big cost and huge social implications. Why can't that be definition outcome? We take people who can't rise out of a wheelchair. We give them a pill. They can rise out of a wheelchair. We have an indication. I mean I know it's – well, maybe, not the simple. Maybe it is that simple.

But why are we spending as much effort on something that's less directly applicable to – I mean it's not called a disease yet, but it costs us a lot of money. Thanks.

Dana Hardin: So we have like no time, right? Two minutes? Two minutes? Yeah.

Dan Perry: I'd like to take the privilege of the chair for just a second to make a quick statement, because the last two speakers have raised some very provocative issues and I do want to hear from the panel. Don't worry about the time.

Dana Hardin: Okay. Thank you.

Dan Perry: Respond to that.

But this meeting would not have taken place without the hard work and diligence of Cynthia Bens, our vice president of public policy. She needs to leave.

[Applause]

Dan Perry: Cynthia needs to leave now, because her brother, who has just returned from military duty in Afghanistan, is receiving a commendation at the Pentagon.

[Applause]

Dan Perry: We want her to be with her family. So, Cynthia, you are dismissed. Enormous thanks –

Cynthia Bens: Thanks everybody.

Dan Perry: – for all you do.

Dana Hardin: We appreciate it.

Cynthia Bens: Thank you.

Dan Perry: Thank you.

[Applause]

Dana Hardin: So let's –

Dana Hardin: And those are very provocative questions and both are equally weighty, but let's just go in order then. I think that's the easiest way to handle this conundrum. And I'd love to hear from anyone, anyone, even in the audience just to answer those.

Stephanie Studenski: I've been hearing from advice at FDA that the constraint is that there has to be a link between the intervention and the mechanism of action, and that's where we get caught on trying to do these more generalizable things.

And that's where with Bill I got hung up on, if the mechanism of action of a lot of our treatments is to grow muscle, we got to measure muscle somehow. And maybe DXA is imperfect, but many measures are imperfect and I wouldn't want to weigh how much more or less is it imperfect compared to anthropometry and labeling creatinine turnover, various other things.

Bill Evans: I guess we just don't know the answer to that question yet. I mean we don't know really how responsive – we know that exercise

interventions clearly result in changes in DXA. We know that inactivity results in changes in DXA.

Part of the problem with DXA is that it's still expensive. It's not clinically approved as an outcome measure or any sort of clinical measure at all. And it's not typically controlled by the people that are treating old people.

So I mean our experience in Arkansas was our chair of geriatrics wanted to buy a DXA instrument to measure bone density in every old lady that came through the doors. And the endocrine division screamed bloody murder and we were never able to use it. And there still remained a two-month waiting list to get a DXA.

So I think that there are – DXA is a perfectly good measurement. I think that for – it's not approved yet and it's not a clinical measurement. And maybe it will be, but I think that it represents a tremendous amount of challenges.

And I think that every one of us in pharma thinks about different indications. And it probably doesn't serve us any good to think about how we think about it, how you think about it. Everybody thinks about it differently. We're all thinking about what's the initial indication for a muscle-acting drug and is it going to make any difference in the patient population that were of interest? And we change every week as well in thinking about which one is going to be the most appropriate.

Dana Hardin: So back to this whole thing –

William Dale: I'm just curious from the mobility experts, is there any actigraphy-based measure that anybody is comfortable with as a potential measure, more steps, more activity, more something? At least it's an objective measure of what we think is closer to what we care about, which is the ability to move around appropriately, however we define it. Stephanie?

Stephanie Studenski: We thought that was the answer. And the only thing I can tell you is that there are so many psychological and environmental factors influencing activity that we find many people where we're clearly making them stronger, making them walk faster, making them get out of chairs better, aren't changing the way they spend their day. So we had great hopes and, at least, we haven't been seeing it so far. I don't how anybody else is. It sounds good.

Ashley Slagle: We also have difficulty with knowing exactly what's being measured. And when we think about labeling in a way that's not misleading – we can't just say I had more steps on my Fitbit, because we don't know what that means. It's sort of step equivalents or – and if someone is not stepping, but they're making this motion or – so it seems like an objective measure. But when we really think about what it's measuring, we still don't know exactly what it's measuring, so actigraphy, some of the activity monitors are very challenging for us to label.

Dana Hardin: So back to that definition, I think Chuck pointed out that we've been a bit circular. And it sounds like we still need a bit narrower definition to get our PROs in the future for sarcopenia or frailty or any of these conditions.

But I guess I'd like to say that I don't think we should give up and I think you have to start somewhere, right? Stephanie started somewhere and we got something from it. And Europe started somewhere.

So if you had to name – boy, I'm really being mean to you today. I'm so sorry. Anybody on the panel, but where we should start? Three things, just give me the top three for sarcopenia, because we don't have a definition. But where we are today, sarcopenia, frailty to get to a PRO that's good and useful.

Ashley Slagle: Well, I have to think about – I want to provide advice that fits into the regulatory framework. And we only have certain ways that we're able to provide advice. And so through the qualification process is a way to get both SEALD and the appropriate divisions together to look at the proposal and be able to work with you on developing an assessment or selecting an assessment. So that's one way. The other, again, is to talk to the division within an individual drug development program.

What's concerning to me is that I hear there are compounds out there that you think might have some utility. And because there's no disease definition and there's no outcome assessment settled on yet, they're not moving forward. And so, until you have an IND opened or you're in Pre-IND stage, it's difficult to talk with us.

So I think one of my takeaways from this is that we need to figure out how we can work with individual sponsors even earlier in the process to talk about some of these things. And, unfortunately, we don't have a mechanism for that yet, but I've been to several of these meetings and that's something that I'm hearing a lot. So

we're moving a big ship sometimes at the agency, but we'll see what we can figure out.

Dana Hardin: Any other thoughts on that definition question that came up?

Bill Evans: I think that most of us are thinking less about a broad definition, although maybe that would be great someday, but thinking about much more specific populations of old people that may have low muscle mass and low function and then trying to treat those specifically. So I think that that probably is a strategy. Hearing kind of what advice the agency has provided in the past and kind of considering how we develop things is that there are patient populations of elderly people with low muscle mass that are much more specific than a broader definition of old people with low mobility and low muscle mass.

Hip fracture is one, clearly, patient population that we know the natural history. We know that they have muscle wasting. We know that they have really, really poor functional capacity six months after fracture. That could be a very specific intervention. And there are a number of others that are possible.

Jack Guralnik: Yeah. I wish I could have heard this discussion and your drug company going through all 130 or however many conditions, because I think we are all struggling with this. And it's easier to come up with the outcome than to figure out what the indication is. And it's a little bit funny. It's traditionally gone in the other direction, but it's important.

And I guess you need to just chip away and come up with something very specific even though the outcomes, as I say, can be a little broader, then they can capture multiple interventions. A specific indication like hip fracture, like short-term immobilization, where there's a lot of mobility loss, other things like that rather than kind of long-term loss of muscle that is so common, but will be much more difficult to get an indication for.

William Dale: I have one. So it occurs to me that one issue is there's not enough of a patient movement for these agents. I think of the HIV experience, at least as I experienced it, which is a very active, very aggressive patient group said, we're not going to wait around forever for the approval process of these medications. Now, their motivation was pretty high of life-threatening illness that was rampaging through the community.

One way to solve some of the political strategic problems of this is we need the right stakeholders and I'm not sure that this is always the right stakeholder group. We need some patient group, who is going to push for, fine, we don't exactly know who we are, but I think I would benefit. Can we please try it? We're happy to volunteer to do so. That's one thought for the political solution.

Dana Hardin: It's a good thought. Yes.

Najwa Mostafa: I was just looking at this article by Dr. William Dale. And I saw this wonderful questionnaire at the back of it. So I was wondering, they've given scores – like the patient would give scores from one to ten. Is it possible to just add up the scores and like come up with a number? And say if the patient has less than X number in the scoring system then they go ahead and go for the next step, such as doing a DXA scan to measure the muscle strength and be treated more like a symptom complex. I know I heard the discussions about environment influencing these things. But if you add up say about 10 or 20 different symptoms, like getting up from the chair or transferring – treat each of them as a symptom and then add up the score and then go the next step of investigating if the score is less than this number and then go about treating it accordingly. Is there a possibility of doing that?

William Dale: I'll defer to my more experienced colleagues, but we have this in quite a few of our cancer patients at this point assuming it's a publicly available measure. There's a lot of work to do to define how the scale works. So that's probably the biggest problem right now is it hasn't done a lot of that work since it was originally developed to be honest. But, yeah, I mean it's 130-point scale that we need to get information about if there's a chance that it has more use. And we'll probably use it, but there's still, unfortunately, quite a bit of work to do between this and all the things you mentioned about scale development.

Bill Evans: I might just say that if geriatricians could get reimbursed, however small the reimbursement would be, just to do an SPPB measurement, that way if you call up your friend, your colleague and say, look, I have a patient with an SPPB score of six. Right now, they have no idea what that means. If they did, then they could say, oh, let me get that person in.

So I think that until there is some sort of process by which we can get that standard measurement in every geriatric clinic, I think that we're – we have a simple tool right now that would be great if we could all use.

Stephanie Studenski: One of the most heartening things I've seen is that these groups that don't care at all about geriatrics and don't pay any attention, there's physical performance measures are popping up in all kinds of important groups that are totally separate from our world.

And there's this whole bunch of surgeons. It started with heart surgeons, but they found out that physical performance predicts post-op complication rates. And they care about that, because there's a lot of CMS influence on their reimbursements. Much better than anesthesia risk scores and their complicated stuff. The cardiologists are discovering all this stuff that it helps them with their agendas. I think it's emerging in oncology.

So that I think it is on a path to broader acceptance than – reimbursement of geriatricians for SPPBs is tiny compared to whether the surgeons and the other important guys, who drive care –

Bill Evans: Well, what's interesting there's a muscle biologist, pulmonologist, Michael Polkey in London, who now uses the SPPB regularly and has shown that the SPPB is a better predictor of outcomes in patients with COPD than FEV1.

Dana Hardin: That makes sense.

Stephanie Studenski: Or the Texas group, who is showing these things predict length of the stay in the hospital. Talk about another big driver.

Dana Hardin: So any last comments and, if not, I'm going to thank the panel and, particularly, our FDA representative who has been very gracious today for being put on the hot seat. When she volunteered to come to a meeting, she probably didn't quite know what she was getting herself into. I'll turn it back over to you.

Dan Perry: Let's thank the panel.

[Applause]

Dan Perry: Thank you, Dana, for chairing it.

I have two pieces of good news. One is that there is, courtesy of a not-for-profit organization, a sumptuous lunch for you in cardboard boxes out there. I didn't want to break the flow of this by having everyone go out and get a box, but they are there. Take a couple.

The even better part of the good news is that while the agenda says I'm going to make closing remarks for 30 minutes now *[laughter]*, I'm going to make just a very, very few observations in closing.

We really did seem to come kind of full circle. We started off talking about the need for a good definition and a good focus on the right patient population. And sure enough, after a very robust and intellectually stimulating discussion from a number of perspectives on this condition, we find ourselves right back there again. So I think we know where the alpha and omega of this is.

I think it's terrific news that Stephanie's paper, which will be out very soon, a matter of weeks, no more than a month, really dovetails into the efforts that are being led by Jack Guralnik and the other members of our Science Advisory Board in the qualification process. So, Stephanie, your work and that of your colleagues is arming our effort to try to get some stake in the ground to begin to define the disease and identify those subpopulations.

And while that's not where we want to be ten years from now, it's a pretty darn good place to start. And then other things start to fall into place: identification of patients; awareness of physicians; better testing.

And the case that Jack and the Science Advisory Board are making come down to usual gait speed and SPPB. And that is going to be more of a story in this condition and, perhaps, some other uses as well.

And then what we heard from Dr. Patrick was that even the best consensus among academicians or regulators or industry is not sufficient unless you are really tapping into the patient experience with the condition. And we also heard that Dr. Patrick's doctor thinks that 70 is young – I'm all for that – and it's time to work on a new knee for the decades that lie ahead. But, seriously, he gave us a terrific sense of the really essential central role of the patient experience and ways to tap into that. And exactly what the linkage is and how we track the linkage from the clinical measures to the patient reports, again, is one of the sweet spots that I think have been identified and where we're going to need a bit more work.

Well, let me just jump down to the value of building some of these patient observations into submissions to the FDA. And Ashley

told us that she's eager to start talking to sponsors as early in the process as possible if we can find the mechanisms to do that.

The one real great value I see is in bringing PROs into the process as we work through qualification. What ends up at the end is in the public domain. It's not going to be a proprietary use. It's something that everyone can use and will really help lift all boats.

And then I'll finally say that the comments about where is the patient group for this? As I told you, we've been through the experience with Alzheimer's and with working with FDA to try to get past some of the barriers there. The fact of the matter is most of the health challenges that we will face in this country and in Europe and in Asia over the next hundred years are going to be chronic, age-associated, let's say, geriatric conditions.

They don't lend themselves to the HIV model of people, young and empowered and feeling like their life is at stake, and they're out chaining themselves to a fence outside FDA.

But what we have found has worked in the Alzheimer's area, and will work, I believe, in sarcopenia, is bringing in the common interests of women's health organizations, seniors' organizations and men's health groups, groups that are involved with muscles and dystrophic diseases, osteoporosis. And we are bringing all of those into this coalition and recognizing the societal, medical stake that we all have in keeping people as healthy and functional and participatory as possible in their last years.

For seven years, we've been meeting with the FDA under the umbrella of that coalition. And I have seen remarkable progress. We're now seeing the disease needed to be identified decades earlier. We started out with simple cognitive measures. We've now moved through a whole host of biomarkers. We've examined what is going wrong in Phase II with these studies holding up in Phase III.

That kind of growth and experience is now available to us in this, which I consider only after Alzheimer's disease, the biggest clinical and quality of life hurdle that all of us are going to face in the later years and for our family members.

So you're all pioneers. We've been meeting once a year now for seven years with the Alzheimer's Coalition. This is the first of the AIM Allies. I hope you continue to hear what's going on. We'll be reaching out to you again for the next round.

And, again, enormous thanks to all of our sponsors, all of our speakers and to all of you and all of the government people. That whole back row is FDA, so I thank you for that. Thank you all and safe travels.

[Applause]

[End of Audio]