Patient-Reported Outcome (PRO) Measure Development in Sarcopenia

Bethesda, Maryland
March 28, 2014

Key Take-Away Messages

- While previous definitions of sarcopenia were based on the opinions of experts, the definition generated by efforts under the FNIH Biomarkers Consortium was based on data analysis. Through the Sarcopenia Project, evidence-based candidate criteria were established as follows:
    - For men: Grip strength (GSMAX) <26 kg and an index of <0.789 for appendicular lean mass adjusted for body mass (ALMBMI).
    - For women: GSMAX <16 kg and ALMBMI <0.512, where GSMAX = highest grip strength measured in 2 or 3 trials in the dominant hand; ALMBMI = appendicular lean mass on DXA scan divided by body mass index (wt in kg/ht in meters squared)
- Functional limitations and disability refer to different behaviors—not to different ways of measuring the same behavior. Therefore, functional limitations and physical disabilities both can be measured using either subjective or objective measures
- A number of studies have shown that performance measures and self-report work in concert to predict health outcomes.
- A PRO measure for sarcopenia will likely be multidimensional, detailed and will represent more than one domain. A good PRO measure for sarcopenia should:
    - Evaluate the impact of muscle wasting on an individual's life;
    - Represent a single impact rather than a multidimensional concept;
    - Be relevant to most people with sarcopenia most of the time, determined by frequency of concept mentions and importance ranking;
    - Be easily understood;
    - Measure a concept likely to change with successful treatment of the condition;
    - Minimize ceiling or floor effects; and
    - Be likely to have semantic equivalence with other languages.
- More multidisciplinary efforts such as this meeting, consortia of private sponsors, and multicenter development that incorporates existing measures incrementally through exploratory endpoints, could accelerate efforts toward developing a PRO measure for sarcopenia, and identifying sarcopenia-specific measures.
- Sponsors should communicate with the FDA early in the clinical trial development process. For FDA, it will be important for the community to narrow the target population by identifying a common set of symptoms or other defining features.
- There is a need to drill down to particular causes of sarcopenia. It will be important to examine men versus women as research has suggested notable gender differences in age-related muscle loss.

MEETING REPORT

I.) Welcome and Overview of the AIM Coalition

Dan Perry, President and Founder of the Alliance for Aging Research (the Alliance),welcomed meeting participants and explained that the Alliance is a nonprofit organization dedicated to accelerating the pace of scientific discoveries and their application to improve the experience of aging and health. Initiated by the Alliance, Aging in Motion (AIM) is a diverse coalition of more than 30 organizations pressing for research and innovation to develop treatments for sarcopenia and associated functional decline. Loss of physical function and independence is a major contributor to diminished quality of life. Presently, there are no U.S. Food and Drug Administration (FDA) measures to guide the development of drug therapies and other treatments for sarcopenia, and there are no pathways for approval for Medicare reimbursement for a new generation of interventions.

Cynthia Bens, the Alliance's Vice President for Pubic Policy, help set the stage for the meeting by offering three important points: First, while there is growing consensus of what sarcopenia is, there is not yet an accepted definition of the disease. Because the FDA views any discussion of measures for clinical trials through the lens of the defined disease to which they apply, there is a dire need for a clear clinical definition of the term. Second, there is already an affected population who suffer from poor quality of life, difficulty recovering from injury, and difficulty managing comorbidities, and that population continues to grow. Third, some companies are now moving into the later stages of developing treatments for sarcopenia, especially in Europe, but the road ahead for these companies has not been fully elucidated. In support of these efforts, therefore, today's meeting will attempt to gain a clearer sense of how the FDA views patient-related outcome (PRO) measures for sarcopenia and how PROs can be employed in clinical trials.

Our attitudes and expectations about old age have changed over time, said Mr. Perry. Old age used to be considered a gift from god that few people experienced. Today we expect to remain healthy and active into our later years. According to the Administration on Aging, the U.S. population age 65 and older increased to 40 million in 2010 (a 15-percent increase in one decade) and will grow to 55 million in the next 10 years (a 57-percent increase). By 2030, driven by the Baby Boomer generation, the number of Americans over age 65 will have more than doubled in a 30-year period. The U.S. population age 85 and over is also expected to increase, rising from approximately 4 million today to almost 20 million by 2050, and with that comes a dramatic rise in chronic disease. According to U.S. Undersecretary of State Robert Hormats, "The nations that learn to tap the productive potential of their aging populations will be those that dominate the century economically, socially, and politically."

In 2006 the Alliance for Aging Research organized a coalition to meet with the FDA to discuss the barriers to drug development in Alzheimer's disease (AD). This effort—called Accelerate Cure/Treatments for Alzheimer's Disease (ACT-AD)—is serving as a model for efforts in sarcopenia.

AIM members in this effort include patient organizations, medical experts in physiology and nutrition, caregivers, provider organizations, healthcare professionals, employers, healthcare industry, and international members especially in Europe. The AIM Scientific Advisory Board is chaired by Jack Guralnik of the University of Maryland. Other committee members in attendance at today's meeting

included William (Bill) Evans of GlaxoSmithKline, Stephanie Studenski of the National Institute on Aging (NIA), Ronenn Roubenoff of Novartis, and Carl Morris of Pfizer.

Perry closed by thanking attendees, Government representatives, and corporate sponsors Eli Lilly and Company, Regeneron Pharmaceuticals, Sanofi, Novartis, GlaxoSmithKline, Myos Corporation, and Nutritia Advanced Medical Nutrition for their support and involvement.

II. Patient-Reported Outcome (PRO) Measure Development in Sarcopenia

Foundation of the National Institutes of Health (FNIH) Scientific Program Manager Maria Vassileva opened her presentation by providing an overview of the FNIH and describing one of its major research partnerships—the Biomarkers Consortium—through which sarcopenia efforts are managed.

A nonprofit organization authorized by the U.S. Congress to create and manage innovative public-private biomedical partnerships that advance public health, the FNIH provides a neutral forum able to engage partners including corporations, foundations, academia, Federal agencies, and philanthropic individuals. The Biomarkers Consortium is an FNIH research partnership launched in 2007 to foster the exchange of knowledge and expertise among industry, academic, and Government leaders in the area of biomarker development and qualification. The Sarcopenia Project, launched in 2010, is housed within the Biomarkers Consortium and is managed by its Metabolic Disorders Steering Committee (MDSC).

Loss of muscle mass is common in aging and wasting conditions and is associated with weakness, poor function, and lower survival. This important clinical condition is poorly recognized, yet multiple potential interventions exist to treat or prevent muscle mass loss. The field therefore needs a clinical definition with clear and valid diagnostic criteria and outcome measures in order to fulfill regulatory demands and support investments in testing interventions.

The goal of the Sarcopenia Project is to create an evidence-based definition by generating the following: (1) Clear and valid diagnostic criteria and outcome measures acceptable to clinicians, the FDA, and health insurers; (2) opportunities to develop and test potential interventions on low muscle mass and strength to improve the health of older adults; and (3) clinical recognition and practice guidelines for screening, diagnosis, and management.

In recent years other groups have arrived at working definitions of sarcopenia. However, a critical question remains unanswered: When is low lean muscle mass empirically related to loss of strength and function?

While previous definitions of sarcopenia were based on the opinions of experts, the definition generated by efforts under the FNIH Biomarkers Consortium was based on data analysis. The 3-year Sarcopenia Project used cross-sectional and prospective data from several aging studies to evaluate criteria for sarcopenia diagnosis based on shared operational definitions of performance, strength, and body composition. Data were pooled from longitudinal clinical studies of more than 26,000 patients, and classification and regression tree analyses were used to derive grip strength and lean body mass cutpoints. The findings are generalizable because of the large, diverse, and well-characterized set of populations and because the pooled sample included both genders, racial/ethnicity diversity,

representation from multiple geographic regions, and subjects with a range of health and functional states. There was an explicit conceptual framework, and researchers did extensive sensitivity and cross-validation analyses. A range of sensitivity and supplementary analyses was possible because alternate measures of physical function, strength, and body composition were used to evaluate whether findings would differ substantially using different cutpoints and measures. Last, the chosen primary indicator—gait speed <0.8m/s—is widely accepted and is associated with reduced survival and increased disability.

Preliminary results of the data analyses were presented at a consensus meeting cosponsored by the FDA, FNIH, and the NIA in May 2012, in Baltimore, MD. Attendees included stakeholders in the Biomarkers Consortium as well as health organizations and academics from around the world, all of whom weighed in on the findings and recommended next steps.

Through the Sarcopenia Project, evidence-based candidate criteria were established as follows:

- For men: Grip strength (GSMAX) <26 kg and an index of <0.789 for appendicular lean mass adjusted for body mass (ALMBMI).
- For women: GSMAX <16 kg and ALMBMI <0.512.

A series of five manuscripts will be published in *The Journal of Gerontology: Medical Sciences* in April 2014 and will include a description of the recommendations from the 2012 consensus meeting as well as final research results. There will also be a series of public announcements, an FNIH press release, a webinar, and a conference to disseminate findings. Findings have already been presented at the American Society for Bone and Mineral Research Annual Meeting held October 2013 and as an Innovator Presentation at the Partnering for Cures Meeting, organized by Faster Cures in New York, NY, in November 2013.

Further research needs to be done to confirm the established criteria in populations with more severe mobility limitation, to consider comorbidities, and to account for differences in measurement on Hologic and Lunar duel-energy x-ray absorptiometry (DEXA) machines. Possible next steps for the Sarcopenia Project include (1) validating and confirming the predictive validity of the candidate criteria established by the first Sarcopenia Project through analysis of populations with higher levels of mobility limitation and with muscle wasting disorders and (2) establishing the reproducibility of repeated measures of body composition and demonstrating how changes in muscle mass affect function.

Dr. Vassileva closed her presentation by acknowledging the involvement of the Methods Core Team, the Sarcopenia Project Team led by Dr. Studenski, and funders.

Discussion and Comments

Dr. Evans asked if the project team had considered the contribution muscle has not just to strength but also to diabetes, obesity, and poor function. Dr. Studenski responded first by saying that, while science continues to evolve after a grant has been awarded, the hope is that the Sarcopenia Project will help build the knowledge base for further research. She emphasized that questions remain about whether the term "sarcopenia" is even appropriate. Research results from the Sarcopenia Project describe clinically significant weakness and significant low lean mass, and it is up to the scientific community

to determine what term best describes this. In direct response to Dr. Evans' question, Studenski said that there is clearly an interface among low mass, low strength, and other conditions like cancer and diabetes, and this is an important unexplored space that was identified as a priority in the project's 2012 consensus meeting. There is clearly a valid population of older people who are weak, she added, though the population may not be as vast as was initially believed.

Dr. Roubenoff asked why appendicular lean mass (ALM) was adjusted for body mass (BMI) and not for height. Dr. Studenski responded that participants in the 2012 consensus meeting were concerned about the effect of body weight and fat, and lean mass divided by height does not account for weight or fat. The project team was therefore asked to evaluate different approaches to developing cutpoints. Lean mass analyses were led by Peggy Cawthon of the San Francisco Coordinating Center, and what was striking was that there were notable gender differences. Men seemed to have the capacity to maintain or increase muscle when they gained weight; women did not.

Dr. Roubenoff said that the European Working Group on Sarcopenia in Older People proposes a flowchart of gait speed followed by grip strength and DEXA and asked whether the Sarcopenia Project was proposing something similar. Dr. Studenski acknowledged the project team's indebtedness to the European Working Group and responded that if a patient presents with mobility complaints or physical performance problems, a clinician would investigate a variety of possible causes of weakness, and one of those possible causes is low muscles mass.

Dr. Studenski added that the cutpoints for grip strength in the Sarcopenia Project were conservative: 16 kg in women and 26 kg in men. These cutpoints are lower than those used by the European Working Group (20 kg and 30 kg, respectively) out of concern that at the higher cutpoints, all older people were likely to be diagnosed with sarcopenia, which might be a concern to the FDA.

III. Progress on Performance-Based Measures for Sarcopenia and the Role of PROs

Jack Guralnik of the University of Maryland presented on two topics: (1) The relationship between measures of performance and self-report and (2) the FDA qualification process for drug development tools.

A. Performance Measures of Physical Functioning

Of the two primary models of the pathway from disease to disability, the Nagi model is the model that has proven most useful in studying this pathway. Each phase of the Nagi pathway was defined by Verbrugge and Jette (*Soc Sci Med*, 1994; 38:1–4) as follows:

- Pathology: Disease, injury, congenital/development condition
- Impairments: Dysfunction and structural abnormalities in specific body systems
- Functional limitations: Restrictions in basic physical and mental actions
- Disability: Difficulty doing activities of daily life

In a 1991 IOM report Nagi elaborated on the difference between functional limitations and other phases of the disease pathway. Impairments, he explained, affect tissues, organs, and systems while functional limitations affect the person as a whole. Further, while functional limitations refer to organistic performance, disability refers to social performance and is thus a relational concept. An

example of an impairment is muscle atrophy or low grip strength; an example of a functional limitation is slow gait speed or difficulty walking 100 yards; and an example of a disability is difficulty with usual household activities or social activities.

Functional limitation is used as an outcome in many studies to investigate the effects of disease and impairment on functional outcome and to evaluate the functional consequences of a risk factor or intervention. It is a more proximal outcome measure than disability, one can observe the effect free of environmental influences, and it has excellent psychometric properties. However, changes in functional limitations are difficult to interpret in relation to performing daily activities. Importantly, functional limitations and disability refer to different behaviors—not to different ways of measuring the same behavior. Therefore, functional limitations and physical disabilities both can be measured using either subjective or objective measures.

Dr. Guralnik described a handful of research studies that used one measure of performance—the Short Physical Performance Battery (SPPB). Developed at the NIA in the 1980s, the SPPB uses a balance test with three foot positions, a timed 4-meter walk, andtime to rise from a chair five times as performance measures.

- Guralnik, et al. (*J Gerontol Med Sci*, 1994; 49:M85–M94), compared self-reported disability to SPPB as a predictor of mortality and nursing home admission. The results showed that performance measures and self-report may complement each other in providing useful information regarding functional status.
- Ferrucci, et al. (*J Am Geriatr Soc*, 2000; 48:1102–10), examined the distribution of SPPB scores among disabled and nondisabled older persons. There was a wide distribution of SPPB scores among the nondisabled population, suggesting that there is potential for developing PRO measures that address the higher end of the functional spectrum.
- Guralnik, et al. (*J Gerontol Med Sci*, 1994; 49:M8–M94), examined mean SPPB scores by age and sex. SPPB scores decreased with age in two discrete subsets of the study population—(1) those that reported not needing help with ADLs, climbing stairs, or walking one-half mile and (2) those that also reported not needing help with ADLs or climbing stairs but reported needing help to walk one-half mile. This suggests that something was happening with these individuals that was not captured with the self-report. For measures—whether performance measures or PROs—to be useful, said Dr. Guralnik, they must be responsive to clinical events and they must predict adverse outcomes.
- Ostir, et al. (*J Clin Epidemiol*, 2002; 55:916-21), demonstrated that SPPB scores are responsive to clinical events. The study used data from the Women's Health and Aging Study to examine SPPB scores after hospitalization for a major event such as hip fracture or stroke. It showed substantial decline in SPPB scores over the 6-month period during which subjects were hospitalized.
- Guralnik, et al. (*N Engl J Med*, 1995; 332:556-561), demonstrated that SPPB scores can predict disability outcomes. Researchers followed people who were nondisabled at baseline and examined how their baseline SPPB scores predicted disability outcomes at 4 years. The results showed that even among those who reported no disability, performance measures were able to capture differences in health status that predicted future disability. Research by Penninx, et al. (*J Gerontol Med Sci*, 2000; 55:M691–697), also demonstrated that SPPB scores can reliably predict outcomes, and studies of gait speed have shown similar results.

A number of studies have shown that performance measures and self-report work in concert to predict health outcomes. Reuben, et al. (*J Gerontol Med Sci*, 2004; 59:1056–61), examined 4-year mortality rates using a combination of self-report and SPPB and found that performance measures and PROs are complementary. Perera, et al. (*J Gerontol Med Sci*, 2005; 60:894–900), examined the relative hazard of death over 5 years using gait speed, SPPB, the short form-36 (SF-36) health survey, and ADLs and while performance measures and self-report picked up different things, they all reliably predicted mortality.

Dr. Guralnik next described a novel approach to determining health status developed at Wake Forest University. Using videoclips that showed stick figures performing activities like walking and climbing stairs at different speeds, researchers asked subjects to indicate how well they could perform specific tasks. The 10-item battery—called the Mobility Assessment Test–short form (MAT-sf)—was proven to have good reliability, and there was a strong relationship between the MAT-sf and SPPB scores. MAT-sf is currently being used in the Lifestyle Interventions and Independence for Elders (LIFE) Study and the International Mobility and Aging Study as a PRO measure.

Dr. Guralnik concluded that the advantages of performance measures of physical function over self-report are that face validity is clear for the task being performed and that performance measures are reproducible, sensitive to change, may work well in persons with poor cognitive function, and reduce the impact of culture, language, and education. Disadvantages are that they take time to perform, require examiners with special training, and require adequate space; there is potential for injury; and simple tests may not reflect performance on complex tasks or adaptation to environment in daily life.

B. FDA Qualification Process

The AIM Coalition is supporting efforts to get SPPB and gait speed approved through the FDA qualification process for drug development tools. Dr. Guralnik described the process and shared his experience to inform the decision of whether to pursue qualification for PRO measures.

FDA clinical outcome assessments (COAs) are used to substantiate treatment benefit claims. There are two processes for FDA submission and review: (1) As part of a drug application and (2) under the Drug Development Tool (DDT) Qualification Program. The latter is a new regulatory process to provide publicly available drug development tools independent of the drug application process.

The process for DDT qualification begins with submission of a letter of intent followed by submission of a briefing package that includes description of the concept of interest and context of use, description of involvement of external experts, evidence of content validity, cross-sectional and longitudinal evaluation of measurement properties, longitudinal evaluation to provide guidelines for interpretation of trial results, and details regarding language translation and cultural adaptation. The next step is investigation and development of the qualification package with input from a DDT Qualification Review Team, followed by review for qualification decision, and finally dissemination for use by all drug developers to maximize the value to public health. While the qualification process is lengthy and challenging, FDA qualification of PROs would help propel clinical trials in sarcopenia.

Questions and Comments

Dr. Studenski asked if the FDA would be more likely to qualify a widely accepted measure of health and function like the SF-36 rather than a newly developed PRO measure for sarcopenia. She acknowledged that it would first be necessary to have a gold standard (such as gait speed) against which to compare the SF-36 in order to determine whether people can accurately perceive changes in their own mobility. Dr. Evans, who has had experience with developing measurement tools for cancer cachexia, responded that the measure would need to be specific to the indication and strongly anchored in physical function. Ashley Slagle of the FDA responded similarly, indicating that the context of use—clarity about exactly what is being measured and in whom—determines whether a particular assessment tool is appropriate. Ms. Bens added that it will be important to clearly define the disease and target subpopulations for labeling purposes. Measures of function—whether performance measures or PROs—are used as outcome measures for many conditions, added Dr. Guralnik, and the challenge lies in relating them to sarcopenia specifically.

Qian-Li Xue of Johns Hopkins University pointed out the problem of people self-reporting functional limitations when there are few or none. Dr. Guralnik responded that combining self-report and performance measures helps identify people who are misreporting. He added that in cases of disability in particular, the environment can impact self-report scores. For instance, in a small town in Italy, self-reports of difficulty with ADLs were traced to subjects having large bathtubs that were hard to climb into. Rezaul Khandker of GlaxoSmithKline asked if discordance between self-report and performance measures might be attest to the fact that ADLs are a continuous measure and therefore more variation is possible while SPPB is a more discrete measure. Dr. Guralnik responded that some self-report measures offer better correlation than others, but there are nevertheless people who truly are discordant—they report no disability but score poorly on performance measures and have poor outcomes.

Brock Beamer of the University of Maryland was discouraged by a statement Dr. Studenski had made earlier in the meeting about there being only a "2-percent level of agreement" between measures of sarcopenia. Dr. Studenski clarified that there are high levels of agreement on who does not have sarcopenia but low levels of agreement on who does have it, emphasizing that the degree of agreement depends on the population. Dr. Guralnik added that the correlation between self-report measures and performance measures is modest and ranges from 0.3 to 0.6, suggesting that they complement one another.

IV. Principles of PRO Measure Development That Apply to Sarcopenia

In his presentation, Donald Patrick of the University of Washington identified challenges in self-reports from patients with sarcopenia and evaluated possible approaches to incorporating PRO measures as endpoints in clinical trials.

Dr. Patrick began by defining key concepts as follows:

- *Patient-focused outcomes* are outcomes important to patients' survival, function, or feelings as identified or affirmed by patients, or their caregivers if patients cannot report themselves.

- *Sarcopenia* is the loss of muscle quality during aging characterized by a decline in muscle strength that if untreated can lead to weakness, disability, increased risk of falls, and loss of independence (Brotto, *IBMS BoneKEy*, 2012, Nov 14; 9. pii: 210).
- *Clinical outcome assessments* are measurements based on human assessment reported by a patient, clinician, or another observer using an instrument. COAs include performance measures and PROs and can serve as endpoints for evaluating treatment benefit in clinical trials.

A. The FDA Roadmap

The FDA's Roadmap to Patient-Focused Outcome Measurement in Clinical Trials identifies three phases in the process of qualifying a measurement tool: (1) Understanding the disease or condition, (2) conceptualizing treatment benefit, and (3) selecting/developing the outcome measure. The ultimate goal is to identify evidence of treatment benefit—through either direct evidence (such as performance measures or self-report) or indirect evidence (such as biomarkers). Direct evidence can be proximal or distal. Proximal evidence includes core signs, symptoms, or decrements in functioning. Distal evidence includes general psychological or physical functioning, social functioning, productivity, or health-related quality of life.

To be qualified for use in a clinical trial, a COA must be a well-defined and reliable assessment of a specified concept of interest (COI) for use in a specified context of use (COU). According to the FDA, COA qualification represents a conclusion that, within the stated COU, the COA can be relied upon to measure a specific concept and have a specific interpretation and application in drug development and regulatory decision-making and labeling. For COAs that do not provide evidence of how patients feel or how they function in daily life, qualification also includes a review of the evidence that the concept assessed is an adequate replacement for how patients feel or function in daily life.

Using the FDA Roadmap as a model, the steps toward qualifying a PRO for sarcopenia are described below.

(1) Understanding the Disease or Condition. First it is necessary to understand the natural history of the disease—its onset/duration/resolution, diagnosis, pathophysiology, and range of manifestations. A good definition of sarcopenia is important because not everyone who has relatively low muscle mass has a clinical problem. Efforts therefore continue to identify acceptable indirect indicators of muscle mass and muscle strength and thereby to achieve consensus about whether sarcopenia should be a medically recognized disease. Also in this preliminary phase of qualifying a PRO, it is necessary to identify patient subpopulations (e.g., by gender, comorbidity, severity, phenotype), describe the healthcare environment, and understand patient and caregiver perspectives.

(2) Conceptualizing the Treatment Benefit. Sarcopenia is clearly a condition of high importance to people who have it and their loved ones and providers, and therefore, there is a need to evaluate treatment benefit. This involves identifying COIs for meaningful treatment benefit, like how the patient feels and functions and whether the patient survives. This also involves defining the COU and selecting an appropriate COA. In defining the COU, Dr. Patrick proposed using an endpoint model that displays the role and hierarchy of relevant outcome concepts in clinical trials—for example, change in selected biomarkers, change in usual gait speed, change in SPPB, and PRO measures such as change in sarcopenia-related signs and symptoms and change in sarcopenia-related impacts.

(3) Selecting/Developing the Outcome Measure. This step involves first determining if there are existing COAs measuring the COI in the COU and then either modifying an existing COA or developing a new one. The greatest challenge in this step lies in the interpretation of the data, Dr. Patrick emphasized, and it is therefore essential to define and document COIs with as much precision as a chemist or other laboratory scientist would.

B. PRO Measures in Sarcopenia: Lessons From Current Research

Dr. Patrick described a 2011 study, in which he was involved, to develop a new measure of self-report of reduced muscle strength (Evans, et al., *J Am Med Dir Assoc*, 2011 Mar; 12(3):226–33). The study population was adults age 55 years and older with sarcopenia who attended open-ended concept elicitation interviews to characterize the functional effects of reduced muscle strength on their lives. The resulting qualitative data were analyzed using a qualitative analysis software program, the PRO measure was drafted, and cognitive interviews were conducted with additional sarcopenia subjects to refine the measure. Funding was provided by Amgen, and the measure remains proprietary, but the study provides assurance that PRO development is possible.

Among lessons learned in conducting this study were that a good PRO measure should:

- Evaluate the impact of muscle wasting on an individual's life;
- Represent a single impact rather than a multidimensional concept;
- Be relevant to most people with sarcopenia most of the time, determined by frequency of concept mentions and importance ranking;
- Be easily understood;
- Measure a concept likely to change with successful treatment of the condition;
- Be unlikely to be vulnerable to ceiling or floor effects; and
- Be likely to have semantic equivalence with other languages.

Among signs and symptoms elicited in this study with direct patient input were strength, energy, balance, endurance, coordination, and emotional symptoms. Impacts elicited included limitations in ADL, social limitations, and many emotional symptoms, the latter which may not be related to muscle weakness.

Another lesson to be learned from this study is the importance of posing interview questions using the language of performance rather than the language of capacity as the two can elicit very different responses. Performance-oriented questions are: *How much difficulty do you have* climbing stairs, walking a straight line, walking 100 yards? Capacity-oriented questions are: *Can you* climb stairs, walk a straight line, walk 100 yards? The former indicates what people actually do; the latter indicates what people think they can do.

This study demonstrated that it is possible to conduct qualitative research sufficient to provide evidence of content validity. However, a prominent question remains: What is the COI in development of PRO measures for sarcopenia? Is it daily experience of muscle weakness, symptoms and impact of muscle weakness, or something else? Other remaining challenges include achieving consensus on the definition of sarcopenia, incorporating both proximal and distal endpoints, developing a relationship

between endpoints in the evaluation of treatment benefit, and meeting the measurement challenges with a well-developed PRO.

C. Conclusions

There are hundreds of generic measures of physical function, said Dr. Patrick. However, there is a need to develop specific measures of physical function that capture, for instance, the difference between mobility, which means getting around in ones environment, and ambulation, which is the phyical act of walking. We should consider measures of mobility among people who are mobile but cannot walk. Condition-specific measures are more responsive to condition specific treatments that may have minimal effect on the broader population.

In addition, it is important that research include both objective and subjective measures—e.g., both an exercise test conducted in a hospital setting and a self-report measure that captures challenges in performing daily activities at home. Each measure captures something different, but the two are complementary. It is possible for objective measures alone to indicate that there is no problem when self-reports indicate compromised physical functioning and diminished quality of life.

To accelerate efforts toward developing a PRO measure for sarcopenia, Patrick suggested conducting more multidisciplinary efforts such as this meeting and including more social scientists in those efforts, gathering consortia of private sponsors through the Critical Path Institute PRO Consortium, pursuing multicenter development through NIH funding, incorporating existing measures incrementally through exploratory endpoints, and identifying sarcopenia-specific measures. He concluded by emphasizing that the science of measurement is the same for all types of COAs, that PROs augment other endpoints in the hierarchy used to evaluate treatment benefit, and that a PRO cannot be chosen or developed without a well-defined COU and targeted COI based on understanding the condition.

Questions and Comments

Dana Sue Hardin of Eli Lilly asked if Dr. Patrick could foresee development of a sarcopenia-related PRO measure with core questions applicable across multiple disease states. He replied that the core questions are already contained in the generic instruments, but the outstanding need is to determine how specific to be. The comorbidity issue is critical, he said, because so many things can contribute to loss of physical performance—e.g., diabetes, arthritis, environment. He emphasized that this is not a linear process. There is a feedback loop: What people do affects their muscle mass.

Dr. Evans commented that clinicians do not systematically measure muscle mass or function and patients cannot self-report on muscle mass, so it seems logical to use muscle weakness as an outcome measure. He emphasized, however, that context is important. For instance, patients with cancer-related cachexia generally report more muscle weakness than patients with sarcopenia because those with cancer lose muscle mass and strength rapidly whereas sarcopenia develops very slowly and patients make appropriate accommodations. Dr. Patrick observed that while his physical therapist tests his physical performance, his orthopedic surgeon makes the decisions about his treatment. The two are in separate offices, he said. They do not use the same charting system, and they do not communicate in any systematic way. He suggested that moving functional testing into the clinical practice would likely

improve patient outcomes. A key consideration, however, is reimbursement: Clinicians currently do not get reimbursed for testing muscle mass or function.

Basil Eldadah of the NIA asked whether patients are able to accurately attribute functional deficits to loss of muscle strength, noting that patients with cognitive impairment may not accurately self-report. Dr. Patrick responded that while patients may not accurately attribute functional deficits to loss of muscle strength, it is still important to attempt to gather data via self-report. Patients with cognitive impairment in particular over-report their ability to perform tasks, he said. Dr. Cawthon added that attribution is particularly challenging in sarcopenia because loss of strength and, in particular, decline in walking speed are part of the aging process. It is therefore difficult for patients to disentangle accelerated loss of strength from growing older. To address this issue, she suggested that PROs for sarcopenia include language that captures changes in strength over time. Dr. Patrick agreed, adding that response shift is a significant problem with self-report, and for that reason it is critical to use not only PROs but also performance measures—which are not subject to response shift—to assess clinical outcomes.

Dr. Eldadah also inquired about the PROMIS Project, in which Dr. Patrick has been involved. Patrick said that all the measures conducted in the PROMIS Project use the language of capacity rather than the language of performance. This may not make a difference on the population level, he said, but it makes a difference to individual patients. He therefore suggested that the battery of PROMIS measures may not have content validity for sarcopenia.

Dr. Evans said there is real concern that everyone over 70 will want a prescription for a medication that treats muscle loss as has occurred with medications for testosterone and erectile dysfunction. He therefore emphasized the importance of clearly defining the disease and the target populations. Dr. Patrick agreed that "we really need to be specific."

V. Status of PRO Development and the Role of PROs for Sarcopenia Treatment Trials

Dana Sue Hardin of Eli Lilly moderated a panel discussion on the status of PRO development and the role of PROs for sarcopenia treatment trials. Panel participants included William Dale of the University of Chicago, Jack Guralnik of the University of Maryland, Bill Evans of GlaxoSmithKline, Ashley Slagle of the FDA, and Donald Patrick of the University of Washington.

Hardin: Of the PROs that have already been qualified by the FDA, what facilitated their approval and what we can learn from that process? Also, we know the FDA has outlined a qualification process, do you foresee any changes with that since you now are searching for a new director?

Dr. Slagle responded that under the DDT Qualification Program, managed by Study Endpoints and Labeling Development (SEALD) staff, the FDA has qualified one instrument: the EXAcerbation of Chronic Pulmonary Disease Tool (EXACT) PRO. The process was lengthy, but the FDA learned a lot from it and is improving efficiency. Among lessons learned is that sponsors need to identify the COI and the target population early in the qualification process. This is typically the most difficult part of the process, Dr. Slagle emphasized. She added that the FDA expects drug development tools to be exceptionally sensitive and specific because ultimately the FDA will be qualifying these tools for use

across multiple drug development programs. Currently, SEALD is searching for a new director, but no major changes are anticipated with the change in management.

Hardin: Considering that there is no agreed-upon definition of sarcopenia, what are your recommendations for moving forward?

Dr. Slagle said there are many diseases and conditions that are not well defined and yet there is a need to move forward in these areas. First and foremost, she encouraged sponsors to communicate with the FDA early in the qualification or instrument development process. In addition, Dr. Slagle emphasized the importance of narrowing the target population by identifying a common set of symptoms or other defining features, adding that the FDA can expand qualification later if the instrument is shown to be appropriate for broader populations.

Dr. Dale described efforts to qualify a PRO measuring strength loss in a very specific population—men with prostate cancer who were being treated with androgen-deprivation medications. He asked if DDT qualification would be facilitated if sponsors working in different specific areas like cancer and diabetes were to join forces in a single application package. Dr. Evans added that this particular PRO measure for men with prostate cancer was part of a larger effort by Amgen, and the mechanism of action of the drug was clear, the population was well-defined, and the PRO was specific to the population. Drs. Evans, Dale, and Guralnik asked Dr. Slagle whether it would be advantageous or even worthwhile to pursue qualification in cases such as this—where the COU is so narrowly defined. Dr. Slagle responded that an instrument may be qualified for a very specific COU, but once qualified it is publicly available and theoretically can be used in a different context. Those using an instrument outside of its approved COU, however, are taking on additional risk and are therefore encouraged to speak with the FDA. The benefit of pursuing qualification for an instrument with a narrowly defined COU, she said, is that it that as more people use it, the body of knowledge is enriched.

Hardin: Has adequate patient input been incorporated into the concept selection for existing PRO measures for sarcopenia? And is there adequate representation across races and socioeconomic groups?

Dr. Patrick responded that the target population for sarcopenia, in the broadest sense, is all older adults. Nevertheless, there is a need to drill down to particular causes of sarcopenia, such as androgen deprivation among men with prostate cancer as described above. Other populations to examine are men versus women as research has suggested notable gender differences in age-related muscle loss. Dr. Patrick made two additional points: First, qualification was originally intended to reduce the number of instruments used in a given domain, but this is problematic considering the variety of target populations and COUs possible for any single instrument. Second, Dr. Patrick pointed out that it is necessary to clarify whether the ultimate goal of treatment is to prevent muscle deterioration or to improve muscle strength because the instrument will look quite different depending on the goal. This was an issue with the EXACT PRO, he said, and is an issue as well in current efforts to qualify an instrument for cystic fibrosis.

Dr. Evans said that Amgen had initiated qualification for three PROs for muscle wasting based on specific patient populations recommended by the FDA, but they found the process to be too costly. He therefore reiterated a question Dr. Studenski had posed earlier in the meeting about whether it might

be more appropriate to investigate use of existing, validated PROs that include questions about physical function. It is rare to find a geriatric patient who is sarcopenic and does not also have hypertension, diabetes, or another condition, he said, and so to be too specific in defining a sarcopenic population is perhaps an exercise in futility.

Hardin: From whom/where should we solicit expertise on subgroups such as ethnic minorities?

Dr. Dale, who is involved in a nationally representative survey of older adults, said that that in large longitudinal clinical surveys, questions always arise about which subgroups are represented and whether those subgroups adequately represent the larger population. To get cross-sectional data, he said, there must be representative samples so that substudies can take advantage of that data rather than repeating the baseline comparison for their particular subgroup. In other words, the problem of representation is solved scientifically by picking the sample correctly in the first place. Dr. Slagle acknowledged that it is a considerable challenge to identify a study population that is specific enough for use in clinical trials and yet is also representative of the broader population.

Dr. Patrick made a comment on the broader discussion of mobility limitation in sarcopenia, emphasizing the importance of distinguishing between "mobility" and "ambulation." People in wheelchairs cannot ambulate but are nonetheless mobile and would benefit from interventions that enhance muscle strength and function. With this in mind, Dr. Patrick advised considering what activities besides walking and climbing stairs are related to muscle weakness—e.g., turning over in bed. Dr. Studenski pointed out that mobility is one of the most hierarchical of behavioral phenomena, and there is extensive literature in this area, so a PRO for sarcopenia can be robust beyond ambulation. It was therefore suggested that research focus on a range of activities within the broader spectrum of mobility. The challenge with sarcopenia, said Dr. Evans, lies in relating mobility to muscle mass because muscle is difficult to measure.

Hardin: Are there existing PROs that could show treatment benefit with appropriate sensitivity and specificity?

Dr. Guralnik responded that it depends on the specificity of the measure, but there are general outcomes that are quite responsive to interventions. He added that it may be necessary to move beyond demanding absolute specificity because almost nobody has a strength problem without some comorbidity. Dr. Patrick cautioned against looking for "a magic bullet that comes in 10 items." A PRO for sarcopenia will likely be multidimensional and detailed and will represent more than one domain.

Hardin: Would any PRO that we designed today be durable, or what do we need to do to ensure durability?

For a PRO to be long-lasting, replied Dr. Cawthon, it needs to be less specific; however, a PRO that is less specific is less likely to be approved. Dr. Guralnik provided an example of how trends may affect durability. He said while it appears that there has been a decline in physical disability since the 80s and 90s, it is possible that existing PROs simply failed to capture the growing use of assistive devices among the aging. A current study—the National Health and Aging Trends Study—is examining this further.

Hardin: Payers may not be as interested in PROs as clinicians and the FDA. Should a PRO therefore be relegated to a secondary outcome measure?

Dr. Dale added to Hardin's question, asking if patient-reported experience might capture something different than performance but nevertheless be linked to performance. He said that "patients seem to know something about themselves that is more important than other measures." Dr. Patrick responded that self-report and physical performance are absolutely linked; we just need to investigate the linkage. While performance is the "anchor," the whole cascade of endpoints contributes to the evaluation of treatment benefit.

Dr. Hardin next opened the floor to questions and comments from the audience.

Dr. Studenski asked if the group should be exploring common ground in Europe and other developed nations addressing similar issues. Dr. Evans responded that the European Union's Innovative Medicines Initiative, through co-funding with industry, has committed 24 million euro toward an intervention trial to investigate the extent to which SPPB scores can be improved. The goal is to reach consensus about what "frailty" means, so there are plans to identify biomarkers for frailty and determine how they change with intervention. Though the project does not include PRO measures, it could nonetheless inform understanding of sarcopenia.

Ms. Bens commented that sarcopenia is only part of the definition of frailty, and she emphasized the importance of taking time to clearly and carefully define sarcopenia early in the process. A challenge arose in AD research, she said, when it became clear that AD begins to develop much earlier than first believed. As a result, researchers are now faced with having not only to rework the definition of AD and to redefine subpopulations but also to rework the instrumentation, all of which is costly.

Dr. Slagle said that the instrument development and DDT qualification process is perhaps not as arduous as she originally suggested. The patient reported outcome (PRO) assessment for one product for enlarged spleen—Jakafi®—was developed in less than 3 years, and applications that have used existing measurement instruments will move through the process even more quickly. Dr. Patrick suggested that using an existing measurement instrument does not necessarily guarantee quick and easy qualification. As examples, he cited efforts to develop measures for irritable bowel syndrome, lung cancer, and depression.

In response to a comment from Dr. Xue, who suggested that defining sarcopenia may be easier than defining frailty, Dr. Evans responded that the challenges in defining sarcopenia are significant: First, there is currently no clinical tool to measure muscle mass, so any definition of sarcopenia that includes low muscle mass is inherently problematic. Second, while we could choose to define sarcopenia in terms of physical performance, poor performance can be caused by a variety of things besides low muscle mass—e.g., cancer, diabetes, arthritis.

Dr. Xue also inquired about improving specificity. He referred to earlier discussion about how measuring change at the individual level can help improve the specificity of a measure. Dr. Guralnik provided what he described as a "dissenting viewpoint" on specificity, saying that just because multiple risk factors or interventions track to an outcome, it not necessarily a poor outcome. Further,

some outcomes like “I have more energy,” “I am more active,” or “I am happier,” are not specific but are incredibly important.

Charles Benson of Eli Lilly observed that without an indication, it is difficult to arrive at a definition of sarcopenia. There was a similar challenge with osteoporosis, for which an indication—fracture prevention—was eventually identified and treatment interventions followed. Eli Lilly, said Dr. Benson, has a number of potential drugs for sarcopenia and has identified the top 30 diseases associated with muscle loss but has yet to find an indication, and thus, progress has been stalled. Dr. Studenski responded that the difficulty lies in establishing the link between the intervention and the mechanism of action—namely, increasing muscle mass. In osteoporosis, DEXA scan is used to measure changes in bone mineral density even though it is an imperfect measure. Dr. Studenski suggested comparing DEXA to current methods for measuring muscle mass like anthropometry and measures of protein turnover to see if they have similar reliability. Dr. Guralnik commented that while indications are important, it is easier to identify outcomes—which can be broader and can capture multiple interventions. Dr. Evans added that ideas about indications for use of a drug to treat sarcopenia are vast and vary between pharmaceutical companies.

Dr. Dale asked if anyone in attendance knew of an actigraphy-based measure of muscle function. Dr. Studenski responded that actigraphy at first seemed promising, but there are too many psychological and environmental factors that affect activity. Dr. Slagle added that actigraphy-based measures are problematic because it is difficult to identify exactly what is being measured, which presents challenges with understanding treatment benefit and describing it in labeling.

Dr. Slagle expressed concern that there appear to be drugs that might have utility for sarcopenia but that efforts to move forward are stalled because there is no disease definition and no agreed-upon outcome assessment. She said the FDA might need to consider how to work with sponsors earlier in the development process to provide input as they consider these critical questions.

To conclude the panel discussion, attendees shared ideas about how to advance drug development efforts for sarcopenia. Dr. Dale observed that there has not been robust patient involvement and advocacy in sarcopenia, as there was in HIV/AIDS, and he suggested soliciting the involvement of stakeholders who could advance political will in this area. Dr. Evans suggested the focus be narrowed to specific populations of elderly people—for instance, people with hip fractures—and he encouraged use of the SPPB in every geriatric clinic as a standard measure for which clinicians are reimbursed. Dr. Studenski noted that physical performance measures like SPPB are increasingly being used in cardiology, oncology, and other disciplines. Heart surgeons, for instance, have found that performance measures predict post-operative complication rates better than anesthesia risk scores, and a pulmonologist in London has shown that the SPPB is a useful functional outcome measure in chronic obstructive pulmonary disease. In other words, performance measures are clearly on a path to broader acceptance.

Concluding Remarks

Mr. Perry concluded the meeting by identifying some of the key concepts that arose in presentations and discussion, including the importance of tapping into the patient experience, identifying the linkage between clinical measures and patient reports, building patient self-report into submissions to the

FDA, and engaging in discussions with the FDA early in the qualification process. It was also clear from the discussion that arriving at consensus about the definition of sarcopenia and identifying target subpopulations is critical to moving forward. The journal article authored by Studenski. et al., due out in April 2014, will help advance these efforts and dovetails with efforts led by Dr. Guralnik to inform the qualification process for PROs.

Most of the health challenges that we will face in Unites States, Europe, and Asia are chronic, age-associated conditions, said Mr. Perry, and they do not lend themselves to the HIV model of the young and empowered. What has worked with AD and will likely work with sarcopenia as well is bringing together the common interests of women's health organizations, seniors' groups, men's health groups, organizations addressing dystrophic diseases, and people working in osteoporosis. There is a societal stake, he emphasized in closing, in keeping people as healthy, functional, and engaged as possible in their later years.