

Outcome measures in clinical trials

Juergen M. Bauer, *Department of Geriatric Medicine, Friedrich-Alexander University of Erlangen-Nuremberg, Germany*

In order to identify new treatments that can reduce the burden of physical decline among older people, it is important to establish guidelines on how these treatments should be developed and tested.¹ Janssen et al found that reduced skeletal muscle mass was significantly and independently associated with functional impairment and disability,² while Clark et al did not find an association between muscle mass and muscle strength.³ Although debate continues regarding the link between muscle mass, strength, and function, the current presentation will review ways to assess muscle quantity and quality, as well as physical function, in studies of people with sarcopenia.

Parameters of muscle mass

Muscle mass has been evaluated in older people by estimating total muscle mass, appendicular muscle mass, or fat-free mass. Testing methods include anthropometric measures (skin thickness or limb circumference), bioimpedance analysis (BIA), dual beam X-ray analysis (DEXA), computerized tomography (CT), and magnetic resonance imaging (MRI). DEXA is a reliable method to measure muscle mass (total or appendicular SM or estimation of fat-free mass),⁴ while CT and MRI may be preferred with regard to the measurement of body fat (total or visceral).

Parameters of muscle strength

Muscle strength can be measured as maximum muscle strength or mean muscle strength. Methods used for measuring strength of specific muscles or muscle groups include handgrip strength, biceps curl, leg press, and knee extension. In practice, there was a linear relationship between baseline handgrip strength and incident disability for activities of daily living (ADL)⁵ and similarly between low knee-extension strength and risk for mortality.⁶

Parameters of muscle function

Functional performance can be measured using tests such as 6-minute walking distance, stair climb, usual gait speed, and Short Physical Performance Battery (SPPB). Performance tests of lower extremity function (gait speed and SPPB) have been shown to predict disability across diverse populations.⁷ Gait speed has been widely used as a measure of function; Abellan van Kan Gabor and colleagues have provided an overview on cutpoints that predict or reflect functional status (Table 1).

Table 1. Risk cutpoints based on usual gait speed (2009, Abellan van Kan Gabor et al with permission)

Gait speed, m/sec	Status/risk
> 1.3	Extremely fit
> 1.0	Healthy older population
< 1.05	Cognitive decline within 5 yrs
< 1.0	Hospitalization within 1 yr
< 0.8	ADL and mobility disabilities within 2 years
< 0.7	Death, hospitalization, institutionalization, falls
< 0.65	Death within 6 mo in severe coronary artery disease patients
< 0.6	Functional or cognitive decline, institutionalization, mortality
< 0.42	Functional dependence and severe walking disability
< 0.2	Extremely frail
< 0.15	Institutionalization; highly dependent older person

Other clinically relevant outcomes

It is also important to consider outcomes that are very important to the study population—activities of daily living, quality of life (QOL), falls, fractures, institutionalization, hospitalization and mortality. For example, QOL measures in people with mobility impairment revealed that power training methods were more favorable than conventional strength training.⁹ Further, functional changes have been indicative of institutionalization and hospitalization.

Take-home messages

- In order to identify new treatments that can reduce the burden of physical decline among older people, it is important to establish guidelines on how these treatments should be evaluated.
- Outcome measures relevant to age-related sarcopenia and its treatments include muscle mass, muscle strength, and physical function; a wide range of techniques have been used for assessment.
- In 2009, the best supported assessment techniques for measuring function are usual gait speed or Short Physical Performance Battery; DEXA techniques are appropriate for measurement of total or appendicular skeletal muscle mass. Another strategy is to use Short Physical Performance Battery plus DEXA.

References

1. Working Group on Functional Outcome Measures for Clinical Trials. Indications, labeling, and outcome assessment for drugs aimed at improving functional status in older persons: a conversation between aging-researchers and FDA regulators. *J Gerontol A Biol Sci Med Sci.* 2009;64A(4):487-491.
2. Janssen I, Heymsfield SB, Ross R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. *J Am Geriatr Soc.* May 2002;50(5):889-896.
3. Clark BC, Manini TM. Sarcopenia \neq dynapenia. *J Gerontol A Biol Sci Med Sci.* Aug 2008;63(8):829-834.
4. Nass R, Pezzoli S, Oliveri M, et al. Effects of an oral ghrelin mimetic on body composition and clinical outcomes in healthy older adults: a randomized trial. *Ann Intern Med.* 2008;149:601-611.
5. Al Snih S, Markides K, Ottenbacher K, Raji M. Hand grip strength and incident ADL disability in elderly Mexican Americans over a seven-year period. *Aging Clin Exp Res.* 2004;16:481-486.
6. Rantanen T, Sakari-Rantala R, Heikkinen E. Muscle strength before and mortality after a bone fracture in older people. *Scand J Med Sci Sports.* Oct 2002;12(5):296-300.
7. Henwood TR, Riek S, Taaffe DR. Strength versus muscle power-specific resistance training in community-dwelling older adults. *J Gerontol A Biol Sci Med Sci.* Jan 2008;63(1):83-91.
8. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci.* Apr 2000;55(4):M221-231.
9. Katula JA, Rejeski WJ, Marsh AP. Enhancing quality of life in older adults: a comparison of muscular strength and power training. *Health Qual Life Outcomes.* 2008;6:45.

Discussion

Abellan van Kan Gabor: I like gait speed as a physical performance measure, but Professor Visser spoke yesterday about hand grip—the relationship between low handgrip and mortality and onset of disability. When we do a physical performance measure, should it be a composite? In a trial we really don't know what will change, so we might include more than one physical performance measure, even if we take just one as a main outcome. I also recognize problems with many of these test methods. The problem with hand grip is that we need to use the same handgrip machine at all centers. For usual gait speed, a single measure could be dangerously unreliable. Perhaps DEXA is the best biomarker of changes, even if we don't know exactly what's going on with muscle mass.

Juergen Bauer: In a new protocol, I would include more functional tests. Calculation of study sample size must be based on some outcome. At present, the best options are gait speed and Short Physical Performance Battery because we have the most information on those two parameters, and we have some information about meaningful changes. With regard to hand grip, we lack this same level of understanding. In my perspective, hand grip may be appropriate for screening, but it is limited as an outcome measure.

Catherine Johnson: Would it be appropriate to have co-primary variables—for example SPPB and either hand grip or knee extensor strength? And then use functional tests or DEXA as secondary variables?

Juergen Bauer: It is difficult to have co-primary outcomes. If you want to have a clear-cut protocol, you need to have a composite score then, such as for hand grip and SPPB, or hand grip and gait speed. I'm not a statistician, but I am reluctant to accept the idea of co-primary outcome measures.

Alfonso Cruz: Having 2 co-primary measures is acceptable in many areas of medicine, for example in cardiology and neurology. You don't need to compound them, but if you see an effect on 2 or 3 of the variables, that is important. However, I would use multiple parameters for different outcomes (ie, muscle strength, muscle function, muscle mass)—not multiple parameters for a single outcome.

Marjolein Visser: Your suggested endpoints are all objective endpoints, and you refer to disability outcomes as further down the line and not feasible to use for a study. I think that if you ask people about limitations they experience, for instance the strength-training people, they self-report dramatic functional effects. It seems very relevant to know how older people experience their improvements. How do you view self-reported items as outcomes for clinical trials?

Juergen Bauer: I think they are very important. However, I don't have the experience with regulatory authorities. How much can you base a study application on self-assessment outcomes? Such assessments are really relevant, but I would rather start with objective measurements that are related to function.

Alfonso Cruz: There are at least 3 parameters that can be used as primary outcomes, either alone or in combination—muscle function, muscle mass, and muscle strength.

For muscle function, it is clear that SPPB and gait speed are the two most relevant measures; they are well validated, recommended by organizations, and have sensitivity to detect change after intervention.

For muscle mass and strength, it is not so clear. Muscle mass can be measured by DEXA or CT scans; both are widely available, and results from these methods are correlated. However, many DEXA machines measure only bone mass and do not have the software to measure lean body mass. For muscle strength, we can measure hand grip, but why not use a good measure of lower-extremity strength. Or instead of mass and strength, why not measure power?

Juergen Bauer: You are opening it all up, that's really good. Is there a measure of lower extremity strength that can be done by a practitioner? What about power measurement?

Jeffrey Stout: Isokinetic testing equipment is available at physical therapy locations. This equipment can simultaneously measure both power and strength with a simple leg extension and leg flexion test, which takes 5 minutes. When selecting variables, each study group must independently establish values for reliability. It is not correct to rely on published literature.

Carolyn Greig: I suggest the use of the Nottingham power rig, which is designed to measure lower limb power output in very frail individuals. It is used widely in the UK and in studies in the United States.

Abellan van Kan Gabor: Do you have a notion of the time frame for the trial? How do we balance long-term impact with study dropouts? A 6- to 12-month study might be ideal.

Juergen Bauer: You are right, with regard to value of longer follow-up periods and gait speed. Study duration depends on the protocol, which varies for different interventions. With new drug interventions that may have risks associated with them, we need to begin with shorter-term studies, in the range of 3 to 6 months.