Clinically Meaningful Change for Physical Performance: Perspectives of the ICFSR Task Force

Jack Guralnik, K. Bandeen-Roche, S.A.R. Bhasin, S. Eremenco, F. Landi, J. Muscedere, S. Perera, J.-Y. Reginster, L. Woodhouse, Bruno Vellas

The Journal of Frailty & Aging
Volume 9, Number 1, 2020

SPECIAL ARTICLE

CLINICALLY MEANINGFUL CHANGE FOR PHYSICAL PERFORMANCE: PERSPECTIVES OF THE ICFSR TASK FORCE

J. GURALNIK[1], K. BANDEEN-ROCHE[2], S.A.R. BHASIN[3], S. EREMENCO[4], F. LANDI[5], J. MUSCEDERE[6], S. PERERA[7], J.-Y. REGINSTER[8], L. WOODHOUSE[9], B. VELLAS[10] AND THE ICFSR TASK FORCE

1. University of Maryland School of Medicine, Baltimore, MD, USA; 2. Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 3. Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4. Critical Path Institute, Tucson, AZ, USA; 5. Fondazione Policlinico A. Gemelli, Catholic University, Rome, Italy; 6. Queen's University, Kingston, Ontario, Canada; 7. University of Pittsburgh, Pittsburgh, PA, USA; 8. University of Liege, Liege, Belgium; 9. University of Alberta, Edmonton, Alberta, Canada; 10. Gerontopole, INSERM U1027, Alzheimer's Disease Research and Clinical Center, Toulouse University Hospital, Toulouse, France.
Corresponding author: Jack Guralnik, University of Maryland School of Medicine, Baltimore, MD, USA, jguralnik@epi.umaryland.edu.

Task force members: Samuel Agus (Paris, France), Islene Araujo de Carvalho (Geneva, Switzerland), Mylène Aubertin-Leheudre (Montréal, Canada), Karen Bandeen-Roche (Baltimore, USA), Ann Belien (Hesusden-Zolder, Belgium), Shalender Bhasin (Boston, USA), Heike Bischoff-Ferrari (Zurich, Switzerland), Andreas Busch (Vevey, Switzerland), Ryne Carney (Washington, USA), Matteo Cesari (Milano, Italy), Caroline Couleur (Epalinges, Switzerland), Alfonso Cruz Jentoft (Madrid, Spain), Susanna Del Signore (London, United Kingdom), Carla Delannoy (Vevey, Switzerland), Waly Dioh (Paris, France), Sonya Eremenco (Tucson, USA), Bill Evans (Durham, USA), Toby Ferguson (Cambridge, USA), Roger Fielding (Boston, USA), Jack Guralnik (Baltimore, USA), Ludo Haazen (Hesusden-Zolder, Belgium), Joshua Hare (Miami, USA), Aaron Hinken (Collegeville, USA), Darren Hwee (South San Francisco, USA), Lori Janesko (Uniontown, USA), Kala Kaspar (Vevey, Switzerland), Francesco Landi (Roma, Italy), Valérie Legrand (Nanterre, France), Bradley Morgan (South San Francisco, USA), John Morley (St Louis, USA), John Muscedere (Kingston, Canada), David Neil (Collegeville, USA), Marco Pahor (Gainesville, USA), Marika Paul (Columbus, USA), Subashan Perera (Pittsburgh, USA), Suzette Pereira (Columbus, USA), John Rathmacher (Ames, USA), Jean Yves Reginster (Liège, Belgium), Leocadio Rodriguez Manas (Getafe (Madrid), Spain), Yves Rolland (Toulouse, France), Michelle Rossulek (Cambridge, USA), Jorge Ruiz (Miami, USA), Lisa Tarasenko (Cambridge, USA), Effie Tozzo (Cambridge, USA), Heber Varela (Miami, USA), Bruno Vellas (Toulouse, France), Jeremy Walston (Baltimore, USA), Debra Waters (Dunedin, New Zealand), Linda Woodhouse (Edmonton, Canada)

Abstract: For clinical studies of sarcopenia and frailty, clinically meaningful outcome measures are needed to monitor disease progression, evaluate efficacy of interventions, and plan clinical trials. Physical performance measures including measures of gait speed and other aspects of mobility and strength have been used in many studies, although a definition of clinically meaningful change in performance has remained unclear. The International Conference on Frailty and Sarcopenia Research Task Force (ICFSR-TF), a group of academic and industry scientists investigating frailty and sarcopenia, met in Miami Beach, Florida, USA in February 2019 to explore approaches for establishing clinical meaningfulness in a manner aligned with regulatory authorities. They concluded that clinical meaningful change is contextually dependent, and that both anchor- based and distribution-based methods of quantifying physical function are informative and should be evaluated relative to patient-reported outcomes. In addition, they identified additional research needed to enable setting criteria for clinical meaningful change in trials.

Key words: Sarcopenia, frailty, aging, disability, physical performance, clinically meaningful change, outcome measures.

J Frailty Aging 2020;9(1)9-13
Published online October 10, 2019, http://dx.doi.org/10.14283/jfa.2019.33

Introduction

Clinical research studies in older populations have shifted over the last two decades from assessing biological indicators and disease status to measuring physical function as a primary endpoint. This shift reflects the World Health Organization's (WHO's) framework for health and disability, the International Classification of Functioning, Disability and Health (ICF) (1), which provides a multidimensional framework for conceptualizing and understanding functioning and disability by integrating medical and biopsychosocial models. Using a patient-focused approach, the ICF promotes the use of language that frames disablement not in terms of disease but in how people live with their conditions across three domains: body function and structure, activity, and participation, i.e., movement in three-dimensional space, interacting with other people, and socialization (2).

With new interventions for sarcopenia on the horizon, the concept of clinical meaningfulness has emerged as an important concern for researchers, clinicians, and regulators. Thus, the International Conference on Frailty and Sarcopenia Research Task Force (ICFSR-TF), a group of academic and industry scientists investigating frailty and age-related muscle loss (sarcopenia), convened a group of experts on February 19, 2019 to explore approaches for establishing clinical meaningfulness and related regulatory considerations.

Clinical meaningfulness, as defined by the U.S. Food and Drug Administration (FDA), requires that an outcome assessment measure something that is clinically important and that substantively affects how the patient feels, functions, or survives. Thus, clinically meaningful outcome measures for sarcopenia and frailty typically assess physical function, quality of life, and survival. Muscle strength and muscle mass may also be considered as outcome measures but only if they correlate with improved function or predict other relevant health outcomes such as reduced mortality, morbidity,

Received July 16, 2019
Accepted for publication July 29, 2019

institutionalization, and disability (3-7).

Clinically meaningful measures may be used to monitor adults in clinical settings and in observational studies, to evaluate efficacy in clinical trials, and to compute sample size and power when planning a clinical trial (8). However, meaningful change in an observational study may differ from meaningful change in an intervention trial where change can occur much more rapidly in the positive direction and must have both clinical and public health relevance. Since rapid changes may be perceived as being much greater in magnitude than those that occur more gradually, objective measurement is important.

Defining a clinically meaningful change in physical performance

Meaningful change can be defined as a change that has clinical or practical importance, has an impact on an individual's self-perceived health status or quality of life, or as a fraction of the standard deviation representing a certain level of movement across the distribution of measurements in the population. Clinically meaningful change depends on the outcome on which it is based. Physical performance measures regularly used in clinical trials include various measures of gait and balance parameters and/or the Short Physical Performance Battery (SPPB), a composite measure of walking speed, standing balance, and sit-to-stand performance (9). Gait performance measures include the 4-meter gait speed test (4MGS), the 6-minute walk distance test (6MWD), the 10-meter walk test (10MWT), the timed 400-meter walk (400MW), and the 3-meter timed "Up & Go" test (TUG) (10), (11). Other possible measures such as gait variability, unipodal balance, and stair negotiation performance may also be used to assess mobility impairments (12, 13). Most evidence has been gathered for the 4MGS, which can be performed in a reasonably small space. For example, in a prospective cohort study of older adults, Perera and colleagues showed that a decline in gait speed of 0.1 m/s on the 4MGS or 1 point on the SPPB over a one-year period was associated with an increased risk of subsequent mortality (14).

Clinically meaningful changes in outcomes may be expressed as changes that exceed minimally clinically important differences (MCID), clinically meaningful differences (CMD), or minimally important changes (MIC) (15). To determine the MCID and Minimally Clinically Important Improvement (MCII), either distribution-based or anchor-based measures may be used. Distribution-based methods use statistical and psychometric properties of a measure to estimate effect size and standard error of measurement ($SEM=\sigma(1-r)^{1/2}$, where σ=standard deviation and r=reliability (16)) as functions of variability and reliability, while anchor-based methods use a change in the patient's or provider's perception to identify the corresponding magnitude of change in a selected measure (8).

Preliminary work by Perera and colleagues estimated what constitutes a meaningful change for three performance measures: gait speed, SPPB, and 6MWD using data from varying populations enrolled in both observational and clinical studies: 1) a basic training data set of a 3-month clinical trial of strength training intervention in people with mild-to-moderate limitations; 2) 1-year data of participants in the Predicting Elderly Performance (PEP) study dataset; and 3) 3-month data from the Stroke Rehabilitation (REHAB) randomized clinical trial of a therapeutic exercise program (8). Using both distribution- and anchor-based approaches, they concluded that small but meaningful changes are near to 0.05 m/s for gait speed, 0.5 points for SPPB, and 20m for 6MWD; and that substantial changes were near to 0.10 m/s for gait speed, 1.0 point for SPPB, and 50m for 6MWD.

They also found that meaningful changes are not affected by gender, race, or baseline performance in the Health ABC study. While men tended to have greater magnitudes for meaningful change in 400MWT and there were health and disease differences (e.g. substantial change estimate for SPPB for those with greater body mass index (BMI) when the anchor of walking ¼ mile was used, but not using other anchors), they did not show a consistent pattern and were limited by dropout bias in 400MWT (17).

In the Lifestyle Interventions and Independence for Elders Pilot (LIFE-P) study of exercise as an intervention, investigators examined the relationship between self-reported and performance measures and estimated the magnitude of meaningful change in 400MWT, 4MGS, and SPPB (18). They used both distribution-based and anchor-based methods, two magnitudes of change, and multiple indicators of self-perceived mobility. Relationships between self-reported and performance measures were consistent between treatment arms. Minimally significant changes were 20-30 seconds in the 400MWT, 0.03-0.05 m/s in the 4MGS, and 0.3-0.8 points in the SPPB. Substantial changes were 50-60 seconds in the 400MWT, 0.08 m/s in the 4MGS, and 0.4-1.5 points in the SPPB.

A validation approach to define meaningful change

A crucial first step in defining meaningful change is to clarify what is meant by the concept of meaningful change. A clinically important change in physical functioning should be large enough that a person perceives the change or that participation (e.g., daily roles) is affected. In clinical trials, a clinically important change indicates a treatment effect large enough to support market authorization of a drug. The analytical approach chosen should be driven by how meaningful change is defined for a particular study depending on its main purpose.

Defining meaningful change may be challenging for several reasons. First, meaningful change varies according to context, including baseline level of function as well as demographic and disease considerations. Second, when no gold standard exists with which to make a comparison, the measures by which

meaningful changes are judged may not reflect the true state.

One method for assessing the ability of a measure to discriminate individuals by their anchor status is to determine meaningful adverse change (MAC) that achieves both good sensitivity and specificity (19). The Women's Health and Aging Study (WHAS), an observational study on the characteristics and progression of disability in older, functionally limited women (20) provides an example of a validation framework for evaluating change over the course of one year using usual pace walking speed as the performance measure and self-reported walking difficulties as the anchor. Participants were dichotomized into those who worsened in any one of seven categories of walking difficulty and those who did not worsen in any category, and walking speed change was compared for those two groups. The difference in mean change between those two groups was estimated at -0.091 meters/sec (95% confidence interval [CI] of -0.128 to -0.054), with a mean change among those not worsening of 0.011 (95% CI of -0.014 to 0.035). A decline of 0.10 m/sec (substantial change), however, had a sensitivity of .41 and specificity of 0.73 for self-perceived worsening, and receiver operating characteristic (ROC) analysis of the ability to discriminate clinical change yielded an area under the curve (AUC) of only 0.59, suggesting that other considerations would be needed to adjudicate whether this is good enough for clinical practice in the community-dwelling context of the WHAS. Reanalyzing the data by evaluating empirical cumulative probability distributions of walking speed stratified by decline in 3 categories of walking difficulty all the way to improving in 3 categories of walking difficulty yielded overlapping curves (except when contrasting perception changes transitioning across multiple categories), indicating that either the anchor is inappropriate or a more sensitive performance measure is needed. In such a context, building performance indices combining multiple measures simultaneously may prove useful for enhancing precision.

Combining performance and patient reported outcome measures

Patient reported outcome measures (PROMs) have been advocated by regulatory agencies because they provide information about what is meaningful to patients. For example, fear of falling is one possible patient-reported measure that might correlate well with balance, strength, and other mobility-related functions. Many studies combine PROMs with performance measures since they provide complementary information (21). In a prospective cohort study, Perera and colleagues showed that performance change and self-reported change were both independently associated with 5-year survival (14).

Studies comparing self-reported versus activity-based performance measures of function such as self-paced walk, TUG, and stair tests have shown moderate correlations (22-25), suggesting that the measures provide complementary information. Moreover, these studies show that the selection of measures is condition specific. For example, in these studies the TUG was shown to be the most sensitive measure to change in patients who have undergone total hip replacement, while in patients undergoing knee arthroplasty the stair measure was more responsive to change.

Case study: Determining meaningful change in physical function in testosterone trials in older men (TOM)

The Testosterone in Older Men with Mobility Limitations (TOM) trial was designed to determine the effect of testosterone administration on physical function and lower extremity strength in older men with mobility limitations and low serum levels of testosterone. The trial was terminated early as a result of an increase in adverse cardiovascular events among participants in the treatment group (26). The trial included both a self-reported measure, the Late-Life Function and Disability Instrument (LLFDI), and several performance-based measures including handgrip strength, bilateral leg and chest press (a measure of strength and power), 12-step stair climb, the 40-meter walk test, and the SPPB. The LLFDI assesses participants' ability to complete discrete actions or activity and their performance of socially-defined tasks (activity and participation in the ICF framework).

Both anchor-based and distribution-based methods were used to determine the MCID for physical function. To assess anchor-based responsiveness, participants were grouped according to self-reported global rating of improvement (better versus no change or worse). The distribution-based responsiveness analysis provided an estimate of effect size, minimal detectable change based on a 90% CI (MDC_{90}), and the percentage of participants exceeding MDC_{90} by group.

These analyses demonstrated that loaded walk and stair climb were the most sensitive, with anchor and distribution-based measures being similar. The SPPB balance assessment was the least sensitive measure. Handgrip strength and LLFDI were not responsive to change while both the Foundation of the National Institutes of Health (FNIH) and European guidelines advocate using handgrip strength to identify participants for sarcopenia trials (27, 28). These results suggest that this measure may be less useful to measure responsiveness to an intervention.

Regulatory considerations of clinically meaningful change

Regulators prefer hard clinical endpoints to surrogate endpoints (e.g. biomarkers) when making decisions about market authorization. For example, in osteoporosis trials, a statistically significant difference in fracture rates – a hard clinical endpoint – is considered meaningful (29), whereas a surrogate endpoint such as bone mineral density would not in and of itself be considered meaningful, although it may be used to bridge studies for extension of indications.

The European Medicines Agency (EMA) guideline on clinical investigation of medicinal products used pain and function as co-primary endpoints in the treatment of osteoarthritis (30). The expert consensus committee that developed the guidelines suggested the threshold for minimal perceptible clinical improvement in pain as a 10 mm improvement on a 100 mm visual analog pain scale for drugs intended to improve symptoms or at least a 5 mm mean difference between placebo and active groups (31). These criteria were applied in a trial of chondroitin sulfate compared to placebo and the non-steroidal anti-inflammatory drug (NSAID) celecoxib, which showed that both drugs produced a statistically significant and clinically relevant improvement, yet whether the magnitude of the effect is sufficient to justify granting market approval remained an unanswered question (32).

A PROM, the SarQoL, has been developed to assess quality-of-life in sarcopenia patients (33). While it has demonstrated the ability to detect statistically significant change, the MIC has not yet been determined; thus, the clinical significance is not clear.

Whether to use continuous or dichotomous variables may also be discussed with regulators. For example, the FRActure in postmenopausal woMen with ostEoporosis (FRAME) study of the bone-forming agent romosozumab assessed percent change in BMD from baseline, demonstrating that the treatment results in a rapid increase in BMD in comparison to bone loss in the placebo group and at the same time reduces fracture risk (34). When using percent change the clinical significance of the observed absolute change must also be considered.

In addition to data on clinically meaningful change used to support marketing authorization for a treatment, payers and policy makers may require additional real-world data and cost-effectiveness studies to support reimbursement (35). For example, validation of the FRAX risk assessment tool was achieved by the Screening for Osteoporosis in Older Women for the Prevention of Fracture (SCOOP) study in the United Kingdom, which showed that screening with FRAX resulted in a reduced risk of hip fracture, i.e., that the tool is medically relevant (36). Another real-world study conducted by the French Ministry of Health – the Pharmaco-Epidemiology of GonArthroSis and coxarthrosis (PEGASus) study -- assessed the ability of multiple symptomatic slow-acting drugs for osteoarthritis to reduce the consumption of NSAIDs, which are associated with substantial adverse reactions. Only glucosamine sulfate showed a significant reduction in consumption of NSAIDs.

The FDA has a somewhat different perspective on meaningful change in that they focus on within-patient anchor-based change. Distribution-based approaches may be used as supportive or supplementary information. Moreover, they require changes to be meaningful to the patient, using terms to which patients can relate. This has led them to incorporate patient preferences into their deliberations and selection of outcome measures.

The Aging in Motion (AIM) coalition has been working with FDA for several years on a project to qualify gait speed alone and the SPPB as acceptable and endorsed measures of function. However, the agency has stressed the need for both an objective measure such as SPPB and a self-report approach, which has increased the complexity of the qualification process.

PROMs present many potential challenges for sponsors. The correlation between PROMs and objective performance measures is modest, and the FDA suggests using them together as joint outcomes. Multiple primary outcomes may increase trial sample size requirements. PROMs are also subject to differences in beliefs and behaviors, making them more difficult to compare across diverse populations. One suggested approach would be to use a goal attainment scale in which the patient sets goals as well as metrics for success.

PROMs, including QOL measures, also must be very specific to the indication. For sarcopenia, this means that PROMs should relate to how low muscle mass affects how patients feel, function, and survive. Used in combination with performance measures, they could provide a powerful way of demonstrating efficacy. While there has been a reluctance of pharmaceutical companies to embed context-specific PROMs in Phase 2 and 3 studies, doing so would produce an enormous body of data that could help establish relevant anchors to estimate MCID and validate other measures.

Moving Forward

One problem for research into how the ICF guidelines are interpreted is that structure and function are typically assessed with clinical measures applied in a controlled environment while assessment of activity and participation require capturing the patient perspective, which is heavily influenced by the environment, adaptation mechanisms, resilience, and coping. Moreover, meaningful change is context, perspective, and purpose dependent.

The Task Force identified several key areas for future research that should be considered when setting the criteria for a clinically meaningful change in a clinical trial:

- Published estimates of MCID derived from study participants who are only mildly functionally limited may have limited value for studies that enroll participants at high risk of physical disability. In substantially impaired participants, a small improvement in a performance test may translate into a large benefit in daily life and be perceived by the participant. Future work should address MCID in subsets of the population stratified by ability, with the instruments chosen being appropriate for that level of ability.
- The validation framework described above offers a paradigm for thinking carefully about the ideal definition of clinically meaningful change and then working backwards to identify how to measure and assess meaningful change.
- To measure clinically meaningful changes in real-world performance, it may be appropriate to incorporate into trials

continuous digital technologies such as accelerometers as well as novel analytical techniques to determine MCID, CMD, and MCII. Signal processing of accelerometer data may also identify additional features predictive of adverse or beneficial outcomes.

Acknowledgements: The authors thank Lisa J. Bain for assistance in the preparation of this manuscript.

Conflicts of interest: The Task Force was partially funded by one educational grant from the Aging In Motion Coalition and registration fees from industrial participants (Biogen, Biophytis, Cytokinetics, Glaxosmithkline, Longeveron, Pfizer and Rejuvenate Biomed NV). These corporations placed no restrictions on this work. S. Eremenco, F. Landi declare there are no conflicts. Dr. Guralnik reports personal fees from Pluristem , personal fees from Viking Therapeutics, personal fees from Novartis Pharma, outside the submitted work. K. Bandeen-Roche reports grants from National Institutes of Health, during the conduct of the study. S.A.R. Bhasin reports grants from AbbVie, grants from Alivegen, grants from MIB, other from FPT, other from AbbVie, outside the submitted work. J. Muscedere is Scientific Director for the Canadian Frailty Network, a non-for profit network funded by the Government of Canada. S. Perera has received Travel expenses to the International Conference on Frailty and Sarcopenia Task Force meeting in February 2019 in Miami Beach, FL paid by Alliance for Aging Research. J.Y. Reginster reports grants and personal fees from IBSA-GENEVRIER, grants and personal fees from MYLAN, grants and personal fees from RADIUS HEALTH, personal fees from PIERRE FABRE, grants from CNIEL, personal fees from DAIRY RESEARCH COUNCIL (DRC), outside the submitted work. B. Vellas reports grants from Nestle, Nutricia, Novartis outside the submitted work.

References

1. World Health Organization. The International Classification of Functioning, Disability and Health (ICF). Geneva2001.
2. Jette AM. Toward a common language for function, disability, and health. Phys Ther 2006;86:726-734.
3. Abellan van Kan G, Cameron Chumlea W, Gillette-Guyonet S, et al. Clinical trials on sarcopenia: methodological issues regarding phase 3 trials. Clin Geriatr Med 2011;27:471-482.
4. Vellas B, Fielding R, Bhasin S, et al. Sarcopenia Trials in Specific Diseases: Report by the International Conference on Frailty and Sarcopenia Research Task Force. J Frailty Aging 2016;5:194-200.
5. Vellas B, Pahor M, Manini T, et al. Designing pharmaceutical trials for sarcopenia in frail older adults: EU/US Task Force recommendations. J Nutr Health Aging 2013;17:612-618.
6. Chiles Shaffer N, Fabbri E, Ferrucci L, Shardell M, Simonsick EM, Studenski S. Muscle Quality, Strength, and Lower Extremity Physical Performance in the Baltimore Longitudinal Study of Aging. J Frailty Aging 2017;6:183-187.
7. Pahor M, Manini T, Cesari M. Sarcopenia: clinical evaluation, biological markers and other evaluation tools. J Nutr Health Aging 2009;13:724-728.
8. Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 2006;54:743-749.
9. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994;49:M85-94.
10. Podsiadlo D, Richardson S. The Timed "Up & Go": A test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142-148.
11. Cesari M, Fielding R, Benichou O, et al. Pharmacological Interventions in Frailty and Sarcopenia: Report by the International Conference on Frailty and Sarcopenia Research Task Force. J Frailty Aging 2015;4:114-120.
12. Brach JS, Perera S, Studenski S, Katz M, Hall C, Verghese J. Meaningful change in measures of gait variability in older adults. Gait Posture 2010;31:175-179.
13. Oh-Park M, Perera S, Verghese J. Clinically meaningful change in stair negotiation performance in older adults. Gait Posture 2012;36:532-536.
14. Perera S, Studenski S, Chandler JM, Guralnik JM. Magnitude and patterns of decline in health and function in 1 year affect subsequent 5-year survival. J Gerontol A Biol Sci Med Sci 2005;60:894-900.
15. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther 2014;9:726-736.
16. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995;4:293-307.
17. Perera S, Studenski S, Newman A, et al. Are estimates of meaningful decline in mobility performance consistent among clinically important subgroups? (Health ABC study). J Gerontol A Biol Sci Med Sci 2014;69:1260-1268.
18. Kwon S, Perera S, Pahor M, et al. What is a meaningful change in physical performance? Findings from a clinical trial in older adults (the LIFE-P study). J Nutr Health Aging 2009;13:538-544.
19. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.
20. Simonsick EM, Guralnik JM, Volpato S, Balfour J, Fried LP. Just get out the door! Importance of walking outside the home for maintaining mobility: findings from the women's health and aging study. J Am Geriatr Soc 2005;53:198-203.
21. Studenski S. What are the outcomes of treatment among patients with sarcopenia? J Nutr Health Aging 2009;13:733-736.
22. Finch E, Walsh M, Thomas SG, Woodhouse LJ. Functional ability perceived by individuals following total knee arthroplasty compared to age-matched individuals without knee disability. J Orthop Sports Phys Ther 1998;27:255-263.
23. Jones LW, Cohen RR, Mabe SK, et al. Assessment of physical functioning in recurrent glioma: preliminary comparison of performance status to functional capacity testing. J Neurooncol 2009;94:79-85.
24. Stratford PW, Kennedy D, Pagura SM, Gollish JD. The relationship between self-report and performance-related measures: questioning the content validity of timed tests. Arthritis Rheum 2003;49:535-540.
25. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. Phys Ther 2006;86:1489-1496.
26. Basaria S, Coviello AD, Travison TG, et al. Adverse events associated with testosterone administration. N Engl J Med 2010;363:109-122.
27. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing 2010;39:412-423.
28. Studenski SA, Peters KW, Alley DE, et al. The FNIH sarcopenia project: rationale, study description, conference recommendations, and final estimates. J Gerontol A Biol Sci Med Sci 2014;69:547-558.
29. Miller PD, Hattersley G, Riis BJ, et al. Effect of Abaloparatide vs Placebo on New Vertebral Fractures in Postmenopausal Women With Osteoporosis: A Randomized Clinical Trial. JAMA 2016;316:722-733.
30. European Medicines Agency. Guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis. In: (CHMP) CftMPfHU, ed. London2010.
31. Reginster JY, Reiter-Niesert S, Bruyere O, et al. Recommendations for an update of the 2010 European regulatory guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis and reflections about related clinically relevant outcomes: expert consensus statement. Osteoarthritis Cartilage 2015;23:2086-2093.
32. Reginster JY, Dudler J, Blicharski T, Pavelka K. Pharmaceutical-grade Chondroitin sulfate is as effective as celecoxib and superior to placebo in symptomatic knee osteoarthritis: the ChONdroitin versus CElecoxib versus Placebo Trial (CONCEPT). Ann Rheum Dis 2017;76:1537-1543.
33. Beaudart C, Biver E, Reginster JY, et al. Development of a self-administrated quality of life questionnaire for sarcopenia in elderly subjects: the SarQoL. Age Ageing 2015;44:960-966.
34. Cosman F, Crittenden DB, Ferrari S, et al. FRAME Study: The Foundation Effect of Building Bone With 1 Year of Romosozumab Leads to Continued Lower Fracture Risk After Transition to Denosumab. J Bone Miner Res 2018;33:1219-1226.
35. MacEwan JP, Gill TM, Johnson K, et al. Measuring Sarcopenia Severity in Older Adults and the Value of Effective Interventions. J Nutr Health Aging 2018;22:1253-1258.
36. McCloskey E, Johansson H, Harvey NC, et al. Management of Patients With High Baseline Hip Fracture Risk by FRAX Reduces Hip Fractures-A Post Hoc Analysis of the SCOOP Study. J Bone Miner Res 2018;33:1020-1026.