Meaningful Change and Responsiveness in Common Physical Performance Measures in Older Adults

Subashan Perera, PhD,[*†] *Samir H. Mody, PharmD, MBA,*[§] *Richard C. Woodman, MD,*[§] *and Stephanie A. Studenski, MD, MPH*[*‡]

OBJECTIVES: To estimate the magnitude of small meaningful and substantial individual change in physical performance measures and evaluate their responsiveness.

DESIGN: Secondary data analyses using distribution- and anchor-based methods to determine meaningful change.

SETTING: Secondary analysis of data from an observational study and clinical trials of community-dwelling older people and subacute stroke survivors.

PARTICIPANTS: Older adults with mobility disabilities in a strength training trial (n = 100), subacute stroke survivors in an intervention trial (n = 100), and a prospective cohort of community-dwelling older people (n = 492).

MEASUREMENTS: Gait speed, Short Physical Performance Battery (SPPB), 6-minute-walk distance (6MWD), and self-reported mobility.

RESULTS: Most small meaningful change estimates ranged from 0.04 to 0.06 m/s for gait speed, 0.27 to 0.55 points for SPPB, and 19 to 22 m for 6MWD. Most substantial change estimates ranged from 0.08 to 0.14 m/s for gait speed, 0.99 to 1.34 points for SPPB, and 47 to 49 m for 6MWD. Based on responsiveness indices, per-group sample sizes for clinical trials ranged from 13 to 42 for substantial change and 71 to 161 for small meaningful change.

CONCLUSION: Best initial estimates of small meaningful change are near 0.05 m/s for gait speed, 0.5 points for SPPB, and 20 m for 6MWD and of substantial change are near 0.10 m/s for gait speed, 1.0 point for SPPB, and 50 m for 6MWD. For clinical use, substantial change in these measures and small change in gait speed and 6MWD, but not SPPB, are detectable. For research use, these measures yield feasible sample sizes for detecting meaningful change. **J Am Geriatr Soc 54:743–749, 2006.**

Key words: meaningful change; responsiveness; physical performance measures; gait speed; SPPB; 6-minute walk

From the [*]Division of Geriatric Medicine and [†]Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania; [‡]Geriatric Research, Education and Clinical Center, Department of Veterans Affairs Hospital, Pittsburgh, Pennsylvania; and [§]Clinical Affairs, Ortho Biotech, LLC, Bridgewater, New Jersey.

Address correspondence to Subashan Perera, PhD, Division of Geriatric Medicine, University of Pittsburgh, 3471 Fifth Avenue, Suite 500, Pittsburgh, PA 15213. E-mail: pereras@dom.pitt.edu

DOI: 10.1111/j.1532-5415.2006.00701.x

Physical performance measures have great value when used as baseline factors to discriminate future health and function in populations of older adults.[1,2] Although performance change over time has been reported as an outcome in observational studies and clinical trials,[3,4] little is known about the amount of change that could be considered clinically important or meaningful. An estimate of the amount of change in performance measures that is clinically meaningful can support clinical and research needs in geriatrics. In the clinical setting, criteria for meaningful change in physical performance could help providers incorporate performance measures into clinical practice by giving guidance about how to determine whether a meaningful change has occurred. In research, criteria for meaningful change could help plan, evaluate, and compare the effectiveness of interventions that use performance measures as outcomes. For example, for an interventional study, a definition of meaningful change can be used to calculate sample size, report proportions that benefit from an intervention, and estimate the number needed to treat.

Current state-of-the-art methods for determining meaningful change belong to two broad categories: the distribution- and anchor-based approaches.[5] Distribution-based methods rely only on the statistical and psychometric properties of a measure in a population. They use the empirical distribution of a measure and its psychometric characteristics, such as variability and reliability, to estimate the effect size[6,7] or the standard error of measurement (SEM).[8,9] Anchor-based methods define a clinical standard for comparison, using the patient's or provider's perception of change as an external anchor to determine the corresponding magnitude of change in the measure of interest.[10] A combination of distribution- and anchor-based methods has been used to estimate meaningful change in other measures, such as in the Functional Assessment of Cancer Therapy.[11] An estimate of the magnitude of meaningful change in a measure can be used to calculate its responsiveness or the ability of a measure to detect clinical change.[12] The Responsiveness Index (RI) is based on the magnitude of meaningful change and the distribution of change in a population[13] and is an important factor in predicting sample size for a clinical trial.

JAGS 54:743–749, 2006

0002-8614/06/$15.00

The purpose of this study was to estimate the magnitude of meaningful change in three commonly used physical performance measures: gait speed, Short Physical Performance Battery (SPPB), and 6-minute-walk distance (6MWD).[1,2,14–16] The goal was to create initial estimates of small meaningful change and substantial change by looking for concordance of estimates between populations, designs, and analytical approaches. Multiple studies were used as data sources to represent varying populations of older adults and observational and clinical trial designs. Furthermore, distribution- and anchor-based methods were used.[5] Finally, sample-size requirements were estimated using the RI for clinical trials that plan to use performance measures as outcomes.[13]

METHODS

Data Sources

Three data sets were used (described below). The University of Pittsburgh institutional review board approved this study.

Basic Training Data Set

The basic training (BT) data set was a two-arm randomized, controlled clinical trial of a 3-month home-based strength training intervention in 100 older persons with mild to moderate mobility limitations.[14] Data were from baseline and the 3-month postintervention assessments.

Predicting Elderly Performance Study Dataset

The Predicting Elderly Performance (PEP) Study data set was a prospective observational cohort study of community-dwelling older adults.[2] Participants were assessed in their homes every 3 months for 1 year and every 6 months for 2 years there after. Only first-year data were used. There were 492 participants at study entry and 457 after 1 year.

Stroke Rehabilitation Trial Data Set

The Stroke Rehabilitation (REHAB) Trial data set was a two-arm randomized trial of a 3-month program of therapeutic exercise in 100 stroke survivors.[16]

Measures

Performance Measures

Gait speed was available in all studies and was calculated as distance in meters divided by time in seconds. Distances varied from 10 feet to 10 m.[2,14,16]

The SPPB score was performed and calculated as recommended.[1] A summary score of 0 to 12 (higher score indicating better function) is based on performance on three tasks: gait speed, chair rise, and three standing positions. Data were available from the PEP Study.

6MWD was performed as recommended.[15] Data were available from the REHAB and BT studies.

Anchor Measures

Short Form-36 Mobility Items: Two items from the self-reported Short Form-36 (SF-36) physical function scale were used: ability to walk one block and ability to climb one flight of stairs. Rating options were limited a lot, limited a little, or not limited at all.[17,18] Data were available from all three studies.

Global Mobility Change Rating: A single question, "Since your last quarterly visit, has there been any change in your mobility?" was used The response was made on a 15-point self-reported Likert scale based on recommendations for global measures of change by Guyatt[19] as follows: -7 = a very great deal worse; -6 = a great deal worse; -5 = a good deal worse; -4 = moderately worse; -3 = somewhat worse; -2 = a little worse; -1 = almost the same, hardly any worse at all; 0 = no change; 1 = almost the same, hardly any better at all; 2 = a little better; 3 = somewhat better; 4 = moderately better; 5 = a good deal better; 6 = a great deal better; 7 = a very great deal better. These data were available for the PEP Study. Small decline was defined as -2 to -3 and substantial decline as -4 to -7.

Data Analysis

Overview

Two distribution-based (effect size and SEM) and one anchor-based (means comparison using several clinical anchors) approaches were used to determine magnitudes of meaningful change in performance measures. Small meaningful change was defined based on the literature recommendations for minimally significant change as effect size of 0.2[7,20] scores of -2 to -3 on the Guyatt 15-point scale, or a 1-level change on the two SF-36 mobility items. A larger standard was also applied for a first estimate of substantial change, which was defined empirically as effect size of 0.5, scores of -4 or worse on the Guyatt 15-point scale, and a two-level change in the SF-36 mobility items. Distribution-based methods do not account for the direction of change, whereas the anchor-based methods used here were based on report of decline.

Subsequently, calculated magnitudes for the performance measures were inspected for consistency across data sets, approaches, and anchors. The best initial estimates for small and substantial change were based on nearby rounded numbers within the range of calculated estimates. Finally, the initial estimates were entered into the RI to calculate expected sample sizes.

Effect Size Calculation

This is a distribution-based method of quantifying a difference between two means on a unitless standard scale where effect size $\delta = (\mu_1 - \mu_2)/\sigma_1$, where μ_1 and μ_2 are the means at baseline and follow-up, respectively, and σ_1 is the standard deviation at baseline.[6] Guidelines for interpreting an effect size are 0.2 for small, 0.5 to 0.6 for moderate, and 0.8 to 1.0 for large changes.[7,20] Effect size–based estimate for a small change was computed as $0.2 \times \sigma_1$, and that for a substantial change was computed as $0.5 \times \sigma_1$.

Standard Error of Measurement

SEM, defined as $\sigma_1\sqrt{(1-r)}$, where r is the test-retest reliability, is a distribution-based method of estimating a small but meaningful individual-level change in a measure.[13,21] Reliability estimates for gait speed and SPPB measures were calculated from subsamples of subjects from the studies, whereas reliability for 6MWD data came from a published source.[22]

Comparison of Means

This anchor-based approach is based on the difference in mean change between groups with and without the anchor standard of change defined above. Those who could not change, because they were at the floor or ceiling, were excluded from the relevant analysis to reduce bias. A mixed model implemented in SAS MIXED procedure (SAS Institute, Inc., Cary, NC), which includes subject as a random effect, was used to obtain *P*-values, which take into consideration the nonindependence among intervals caused by many intervals of change contributed by the same subject.

Responsiveness Index

RI is defined as the ratio of clinically important difference to the between-subject variability of individual changes and is computed using $RI = \Delta/\sqrt{(2 \times S)}$, where Δ is a difference deemed clinically important and S is the mean square error obtained from an analysis of variance model examining repeated test observations in the nontreated or reference subjects, assuming independence.[13] This is to be interpreted in terms of the number of subjects needed to detect the clinically important difference, where a clinical trial using outcomes with high RIs requires only a small sample size, based on the standard sample-size formula $N = 2((Z_{\alpha/2}+Z_{\beta})\sigma/\Delta)$, where $\sigma = \sqrt{(2 \times S)}$ under the conservative independence of repeated-measures assumption. RI was evaluated for each of the performance measures using the final recommendation based on the integration of all findings as described above of Δ for small meaningful and substantial difference. Sigma (σ) was estimated using the change in control group in the BT and REHAB datasets for the standard deviation of the change score, rather than as $\sigma = \sqrt{(2 \times S)}$, more in line with statistical theory.

RESULTS

Detailed subject characterizations of the three studies appear elsewhere[2,14,16] in detail and are summarized in Table 1. In general, the participants in the REHAB Trial had the greatest deficits in performance, and the participants in the PEP Study were the most diverse in performance. For the BT strength training trial in mobility-limited older people (n = 100), the mean baseline age was 77.6 ± 7.6, gait speed was 0.78 ± 0.26 m/s, and 6MWD was 233 ± 94 m. For the PEP Study data set (n = 492), baseline age was 74.1 ± 5.7, gait speed was 0.88 ± 0.24 m/s, and SPPB score was 8.3 ± 2.7 points. For the REHAB Trial (n = 100), baseline age was 69.8 ± 10.3, gait speed 0.65 ± 0.28 m/s, and 6MWD was 228 ± 98 meters. Retention rates were 87% in the BT Study, 92% in the REHAB Study, and 93% in the PEP Study.

Meaningful Change Using Distribution-Based Approaches

Effect Size Analysis

Table 2 shows estimates of meaningful change based on effect size using the three data sets. Substantial change was operationally defined as moderate effect size. Most gait speed change for small effects ranged from 0.04 to 0.06 m/s and for moderate effects from 0.10 to 0.14 m/s. SPPB score changes corresponding to small and moderate effect sizes were 0.54 and 1.34 points, respectively. 6MWD changes for small effects were 19 to 20 m and for moderate effects were 47 to 49 m.

SEM Analysis

Table 2 also shows estimates based on SEMs. These results essentially correct the effect size estimates for reliability and are recommended only for calculating criteria for small meaningful change. The SEMs for small change in gait speed ranged from 0.04 to 0.06 m/s for 4-m and 10-m walks and 0.10 m/s for the 10-foot walk. The latter larger estimate may be due to the lower reliability of the shorter walk. For 6MWD, the SEM was 21 to 22 m. The SEM for the SPPB

Table 1. Subject Characteristics and Relevant Baseline Variables from Three Data Sets

Characteristic	Basic Training Data Set (n = 100)	Predicting Elderly Performance Study Data Set (n = 492)	Stroke Rehabilitation Data Set (n = 100)
Age, mean ± SD	77.6 ± 7.6	74.1 ± 5.7	69.8 ± 10.3
Female, n (%)	50 (50.0)	213 (43.7)	44 (44.0)
10-foot gait speed, m/s, mean ± SD	0.63 ± 0.21	—	—
4-meter gait speed, m/s, mean ± SD	—	0.88 ± 0.24	—
10-meter gait speed, m/s, mean ± SD	0.78 ± 0.26	—	0.65 ± 0.28
6-minute walk distance, m, mean ± SD	233 ± 94	—	228 ± 98
Short Physical Performance Battery score, mean ± SD	—	8.3 ± 2.7	—
Short Form-36 mobility questions, n (%)			
Climbing one flight of stairs			
Limited a lot	21 (21.7)	68 (13.7)	17 (17.0)
Limited a little	43 (44.3)	107 (21.6)	33 (33.0)
Not limited at all	33 (34.0)	320 (64.7)	50 (50.0)
Walking one block			
Limited a lot	17 (17.2)	62 (12.5)	10 (10.0)
Limited a little	31 (31.3)	65 (13.1)	35 (35.0)
Not limited at all	51 (51.5)	368 (74.3)	55 (55.0)

SD = standard deviation.

Table 2. Distribution-Based Meaningful Differences

Effect Size, SEM, and Nature of Change	Gait Speed (m/s) 10-Foot (BT)	Gait Speed (m/s) 10-Meter (BT)	Gait Speed (m/s) 10-Meter (REHAB)	Gait Speed (m/s) 4-Meter (PEP)	Short Physical Performance Battery (Points) (PEP)	6-Minute Walk Distance (m) (BT)	6-Minute Walk Distance (m) (REHAB)
0.2 (small) small meaningful change	0.04	0.05	0.06	0.05	0.54	19	20
0.5 (moderate) substantial meaningful change	0.10	0.13	0.14	0.12	1.34	47	49
SEM small meaningful change*	0.10	0.06	0.04	0.06	1.42	21	22

* Standard error of measurement (SEM) is used only for estimating minimally meaningful change, not substantial change.
BT = Basic Training strength training clinical trial; REHAB = Stroke Rehabilitation Trial; PEP = Predicting Elderly Performance Study.

was 1.42. This larger value was due to the somewhat lower reliability of SPPB than of gait speed and 6MWD.

Meaningful Change Using Anchor-Based Methods

Anchor-based methods can be applied based on direction of change. The results presented here are based on self-reported decline. The number of subjects and subject intervals in which subjects were at floor, had a small decline, and had a substantial decline are summarized in Table 3 for the three data sets. Those at floor did not have the opportunity to self-report small or substantial decline and were excluded from subsequent anchor-based analyses. No change in self-reported mobility was seen in the vast majority of quarterly intervals in community-dwelling older people, with some episodes of small decline and a few of substantial decline. The difference between mean performance change in groups with and without self-reported decline, using three items (ability to walk a block, ability to climb one flight of stairs, and global assessment of mobility change) are presented in Table 4. Definitions of small and substantial self-reported decline are described in the Methods section. Overall, the mean difference in gait speed between persons who reported a small decline versus those who did not was 0.00 to 0.10 m/s, whereas the gait speed difference for substantial decline was 0.08 to 0.11 m/s. For SPPB, small decline ranged from 0.27 to 0.55 points, whereas substantial decline ranged from 0.99 to 1.88 points. For 6MWD, a small decline ranged from 21 to 54 m, but substantial decline could not be estimated in the REHAB Study, because few participants had substantial self-reported decline on the two SF-36 items, and the global mobility item was not collected here.

Responsiveness and Sample Size Estimates for Clinical Trials

Table 5 provides initial overall estimates for small and substantial change for the performance measures considered. Because there is no known best statistic or mathematical formula to allow for the combination of these estimates, the recommendations were based on the general consistency and tendencies found in the distribution- and anchor-based results reported above, and a preference for rounded numbers was applied for ease of practical use. The RI using the best estimates and the expected distributions of change was then applied to estimate sample sizes for small and substantial change (Table 5). All estimates assumed a clinical trial with an independent comparison of two groups at a significance level of .05 and 80% power. For detecting small meaningful effects, expected sample sizes ranged from 71 to 161 per group. For detecting substantial effects, expected sample sizes ranged from 13 to 42 per group.

DISCUSSION

Initial estimates of the magnitude of clinically meaningful change in physical performance measures can contribute to the needs of clinical geriatrics and research in aging. Estimates of small meaningful and substantial change in three commonly used performance measures were provided, based on state-of-the-art analytic techniques using data

Table 3. Number of Intervals and Unique Subjects with No, Small, and Substantial Change with Respect to Each Anchor

Self-Report Anchor	Predicting Elderly Performance Study: At Floor	No Change	Small Meaningful Decline	Substantial Decline	Stroke Rehabilitation Study: At Floor	No Change	Small Meaningful Decline	Substantial Decline
	Intervals, n (Unique Subjects, n)				Subjects			
Short Form-36								
Climbing one flight of stairs	113 (52)	1,222 (402)	181 (151)	17 (17)	17	45	13	2
Walking a block	132 (56)	1,306 (399)	144 (120)	17 (17)	9	57	6	2
Global mobility change	—	1,437 (448)	106 (85)	108 (78)	—	—	—	—

Table 4. Anchor-Based Meaningful Decline Estimates

Self-Report Anchor	4-Meter Gait Speed Decline Difference (m/s) (PEP)	10-Meter Gait Speed Decline Difference (m/s) (REHAB)	Short Physical Performance Battery Score Decline Difference (points) (PEP)	6-Minute Walk Distance Change Difference (m) (REHAB)
Small meaningful change				
SF-36 climbing one flight of stairs (no change vs decline of one level)	0.00	0.01	0.27*	21
SF-36 walking a block (no change vs decline of one level)	0.04†	0.10	0.55†	54
Global mobility change (no change (-1 to $+1$) vs small decline (-2 to -3))	0.04†	NA	0.48†	NA
Substantial meaningful change				
SF-36 climbing one flight of stairs (no change vs decline of two levels)	0.11†	NA	1.88†	NA
SF-36 walking a block (no change vs decline of two levels)	0.08†	NA	0.99†	NA
Global mobility change (no change (-1 to $+1$) vs substantial decline (-4 to -7))	0.09†	NA	0.99†	NA

* $.01 < P < .05$; † $P < .01$ with a subject random effect to account for several intervals from the same subject, if any.
SF-36 = Short Form 36; REHAB = Stroke Rehabilitation Trial; PEP = Predicting Elderly Performance Study; NA = anchor not available in data set or sample size too small.

from diverse groups of older participants from observational and intervention studies.

As performance measures become part of usual care of older adults, clinicians will need easy access to such criteria to determine whether a change in performance in an individual patient is meaningful. In research, such criteria are useful in evaluating the clinical significance of an intervention. Sample size heavily influences the statistical significance of an improvement in performance in a clinical trial. Clinically interpretable trial effects on performance measures can be examined based on standards of meaningful change by comparing the proportion of treatment and

Table 5. Overall Recommendations for Criteria for Meaningful Change and Sample Size Estimates Based on Responsiveness Index

Performance Measure	Recommended Criterion for Meaningful Change (Δ)	Estimated Standard Deviation of Change for Stable Subjects (σ)	Number Needed per Group for 80% Power in a Between-Group Comparison (n)
10-foot gait speed			
Small meaningful change	0.05 m/s	0.11 m/s	77
Substantial meaningful change	0.10 m/s	0.11 m/s	21
10-m gait speed			
Small meaningful change	0.05 m/s	0.15–0.16 m/s	142–161
Substantial meaningful change	0.10 m/s	0.15–0.16 m/s	37–42
4-m gait speed			
Small meaningful change	0.05 m/s	0.12 m/s	90
Substantial meaningful change	0.10 m/s	0.12 m/s	23
Short Physical Performance Battery score			
Small meaningful change	0.5 points	1.48 points	138
Substantial meaningful change	1 point	1.48 points	35
6-minute distance			
Small meaningful change	20 m	45–53 m	71–115
Substantial meaningful change	50 m	45–53 m	13–20

control groups who achieve change and calculating the number needed to treat.[23] Finally, sample size estimates are needed in the planning stage of research studies and should be based on ability to detect an important level of change. Sample size estimates in turn affect important trial aspects, including cost, time, and recruitment plans.

A desirable characteristic of estimates of meaningful change is whether that magnitude of improvement can be realistically achieved. Although the anchor-based estimates were derived from self-reported decline, distribution-based estimates apply to improvement and decline, and the meaningful change criteria appear achievable, because they are comparable with the magnitudes of improvement reported in recent research studies. In a study of home exercise training, with and without additional group training, gait speed increased 0.06 to 0.07 m/s.[24] For SPPB change, controlled exercise trials of mobility-limited older people and community-dwelling elderly women yielded gains of as much as 2.7 points.[25] Six-minute walk gains in two strength and endurance training regimens in older subjects with chronic obstructive pulmonary disease were 26 m and 34 m[26] and was 30 m in a home-based exercise program.[27]

This study had several strengths. Several data sets that represent a range of performance in older adults and included observational and intervention studies were used. Because baseline variability heavily influences distribution-based measures of change, it was possible to test assumptions about diversity on the estimates and look for consistency between data sets and populations. Finally, several different clinical anchors were used to evaluate consistency of estimates. For anchors, indicators of change in current status at two time points and transition measures such as self-rated magnitudes of decline over a period of time were employed. The gait speed estimates were the most robust, because it was possible to use all data sets.

There is little information on the magnitude of meaningful change in performance measures. One study estimated the magnitude of small meaningful change in 6MWD as 54 m in a sample of 112 patients with chronic obstructive pulmonary disease.[28] Ad hoc estimates of performance change against 5-year survival have previously been examined.[29]

These effect estimates may not reflect populations that are different from the one in the current study. They are most relevant to populations of older adults with mild to moderate mobility deficits. Such populations are frequently seen in clinical geriatrics and are appropriate targets for interventions to improve mobility. This study also was unable to estimate a small meaningful effect for the SPPB, because its integer-valued scoring precludes individual measurement of 0.5 points. Also, SPPB was available only in the PEP study, and only the SF-36-based anchors were available for estimates of change in 6MWD. The estimates of change should be considered preliminary evidence and will require further confirmation using other data and using similar as well as additional analytic techniques. Finally, Guyatt's RI must be interpreted with care. It depends heavily on Δ, which represents the effectiveness of an intervention in a clinical trial setting. The greater the effectiveness of an intervention, the greater the responsiveness of a particular measure. Therefore, it may be feasible to use a particular measure to detect a large effect with statistical significance, but it may be impractical because of large sample size requirements when attempting to detect a small effect. The RI also heavily depends on σ, a measure of between-subject variability in the population of interest. Guyatt recommends estimating σ with "stable" subjects, which can be interpreted as a suitable group of comparison subjects who have not received the intervention of interest but not necessarily stable across time. In a clinical trial setting, the most feasible group of reference subjects may be found in the control, nontreated, placebo group or the standard treatment group.

In summary, as physical performance measures become integrated into geriatric clinical care and research, familiarity with the meaning of these measures is essential. Performance measures can now discriminate between older adults in current and future health, function, and utilization. This study has shown that they are capable of reflecting important degrees of change over time and that thresholds for meaningful change can be estimated. Like vital signs such as body weight or blood pressure, performance measures may offer a powerful mechanism to understand and act on the healthcare needs of older adults.

ACKNOWLEDGMENTS

This study was supported in part by Ortho Biotech Clinical Affairs; Merck Research Laboratories; Pittsburgh Department of Veterans Affairs Geriatric Research, Education, and Clinical Center; Kansas and Pittsburgh Claude D. Pepper Older Americans Independence Centers (NIA P60 AG14635 and P30 AG024827); and other NIH grants (K07 AG023641) and appeared as an abstract in the *Journal of the American Geriatrics Society* 2005;53:S116.

Financial Disclosures: Perera: funding from Eli Lilly and Co. and Ortho Biotech, LLC, to do observational research and past consultant to Teva Neuroscience, Inc. Mody and Woodman: employed by Ortho Biotech, LLC. Studenski: funding from Eli Lilly and Co and Ortho Biotech, LLC, to do observational research.

Author Contributions: Perera: study concept and design, analysis of data, interpretation of data, manuscript preparation. Mody and Woodman: Interpretation of data, manuscript preparation. Studenski: study concept and design, acquisition of subjects and data, interpretation of data, manuscript preparation.

Sponsor's Role: None. Authors retained complete independence in scientific investigation and reporting.

REFERENCES

1. 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 placement. J Gerontol 1994;49:M85–M94.
2. Studenski S, Perera S, Wallace D et al. Physical performance measures in the clinical setting. J Am Geriatr Soc 2003;51:314–322.
3. Penninx BW, Guralnik JM, Onder G et al. Anemia and decline in physical performance among older persons. Am J Med 2003;115:104–110.
4. Salbach NM, Mayo NE, Robichaud-Ekstrand S et al. The effect of a task-oriented walking intervention on improving balance self-efficacy poststroke: A randomized, controlled trial. J Am Geriatr Soc 2005;53:576–582.
5. Norman GR, Sridhar FG, Guyatt GH et al. Relation of distribution- and anchor-based approaches in interpretation of changes in health-related quality of life. Med Care 2001;39:1039–1047.
6. Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178–S189.

7. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.
8. Wyrwich KW, Nienaber NA, Tierney WM et al. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469–478.
9. McHorney CA, Tarlov AR. Individual patient monitoring in clinical practice: Are available health status surveys adequate? Qual Life Res 1995; 4:293.
10. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimally important difference. Control Clin Trials 1989;10:407–415.
11. Cella D, Eton DT, Lai J-S et al. Combining anchor and distribution-based methods to derive minimal clinically important differences on the functional assessment of cancer therapy (FACT) anemia and fatigue scales. J Pain Symptom Manage 2002;24:547–561.
12. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol 1997;50:239–246.
13. Guyatt G, Walter S, Norman G. Measuring change over time assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171–178.
14. Chandler JM, Duncan PW, Kochersberger G et al. Is lower extremity strength gain associated with improvement in physical performance and disability in frail, community-dwelling elders? Arch Phys Med Rehabil 1998;79: 24–30.
15. Troosters T, Gosselink R, Decramer M et al. Six-minute walk test: A valuable test, when properly standardized. Phys Ther 2002;82:826–827.
16. Duncan PW, Studenski S, Richards L et al. Randomized clinical trial of therapeutic exercise in subacute stroke. Stroke 2003;34:2173–2180.
17. Ware JF, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): Conceptual framework and item selection. Med Care 1992;30: 473–483.
18. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Med Care 1988;26:724–735.
19. Guyatt GH, Townsend M, Berman LB et al. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987;40: 1129–1133.
20. Testa M. Interpreting quality of life clinical trial data for use in the clinical practice of antihypertensive therapy. J Hypertens 1987;5:S9–S13.
21. McHorney C, Tarlov A. Individual-patient monitoring in clinical practice: Are available health status surveys adequate? Qual Life Res 1995;4:293–307.
22. Harada ND, Chiu V, Stewart AL. Mobility-related function in older adults: Assessment with a six-minute walk test. Arch Phys Med Rehabil 1999;80: 837–841.
23. Guyatt GH, Osoba D, Wu AW et al. Methods to explain the clinical significance of health status measures. Mayo Clin Proc 2002;77:371–383.
24. Helbostad JL, Sletvold O, Moe-Nilssen R. Home training with and without additional group training in physically frail old people living at home. Effect on health-related quality of life and ambulation. Clin Rehabil 2004;18:498–508.
25. Bean JF, Herman S, Kiely DK et al. Increased Velocity Exercise Specific to Task (InVest) training: A pilot study exploring effects on leg power, balance, and mobility in community-dwelling older women. J Am Geriatr Soc 2004;52: 799–804.
26. Mador MJ, Bozkanat E, Aggrawal A et al. Endurance and strength training in patients with COPD. Chest 2004;125:2036–2045.
27. Nelson ME, Layne JE, Bernstein MJ et al. The effects of multidimensional home-based exercise on functional performance in elderly people. J Gerontol A Biol Sci Med Sci 2004;59A:M154–M160.
28. Redelmeier DA, Bayoumi AM, Goldstein RS et al. Interpreting small differences in functional status: The six-minute walk test in chronic lung disease patients. Am J Respir Crit Care Med 1997;155:1278–1282.
29. Perera S, Studenski S, Chandler JM et al. Magnitude and patterns of decline over one year in health and function affect survival over five years. J Gerontol A Biol Sci Med Sci 2005;60A:M894–M900.