© International Journal of Clinical and Health Psychology
ISSN 1697-2600
2010, Vol. 10, Nº 2, pp. 359-378

Evaluation of walking speed tests as a measurement of functional limitations in elderly people: A structured review

Carmen L. Muñoz-Mendoza[1], Julio Cabrero-García, Abilio Reig-Ferrer, and María José Cabañero-Martínez (*Universidad de Alicante, Spain*)

ABSTRACT. The aim of this theoretical study was to evaluate the conceptual model, burden, interpretability, floor or ceiling effects, reliability, validity, and responsiveness of walking speed tests as a measurement of functional limitations in elderly people. A systematic search was conducted in Medline, AgeLine, Embase, CINAHL, and IME, manual searches and references searches. Standardised criteria were applied to assess the quality of the measurement properties. 102 studies were analysed, identifying 18 tests. The most used tests were: 2.44 meters, 4 meters, and 6 meters, carried out at usual gait speed. Most of the findings focused on predictive validity and test-retest reliability; in the latter case, the coefficient values were higher than the quality standards recommended. Scant metric evidence was provided for the attributes burden, interpretability, floor or ceiling effects, and responsiveness. In epidemiological studies, the evidence available supports the use of walking speed tests as predictors of adverse results related with health in elderly people. However, further studies are required to support their viability and applicability in clinical practice, for both screening purposes and to monitor, and evaluate change.

KEYWORDS. Walking speed test. Elderly people. Functional limitations. Theoretical study.

RESUMEN. El objetivo de este estudio teórico fue valorar el modelo conceptual, la carga, la interpretabilidad, los efectos suelo y techo, la fiabilidad, la validez y la

[1] Correspondence: Departamento de Enfermería. Universidad de Alicante. 03690 Alicante (España). E-mail: carmen.munoz@ua.es

sensibilidad al cambio de los tests de velocidad de andar como medida de limitaciones funcionales en personas mayores. Se realizaron búsquedas electrónicas en *Medline*, *AgeLine*, *Emba*se, CINAHL e IME, búsquedas manuales y por referencias. La calidad de las propiedades de medición fue evaluada mediante criterios estandarizados. Se analizaron 102 estudios que identificaron 18 tests. Los tests más utilizados fueron: 2.44 metros, 4 metros y 6 metros, realizados a velocidad del paso habitual. Las evidencias se centraron en la validez predictiva y en la fiabilidad test-retest; en este caso, los valores de los coeficientes fueron superiores a los estándares de calidad recomendados. La interpretabilidad, los efectos suelo y techo y la sensibilidad al cambio son los atributos con menos evidencias. En estudios epidemiológicos, la evidencia disponible apoya el uso de los tests de andar como predictores de resultados adversos relacionados con la salud en personas mayores. Se precisa de estudios que apoyen su viabilidad y aplicabilidad en la práctica clínica, ya sea con la finalidad de *screening* como para la monitorización y evaluación del cambio.

PALABRAS CLAVE. Tests de velocidad de andar. Personas mayores. Limitaciones funcionales. Estudio teórico.

In research about ageing, measurements of functional limitations, self-reporting, and objective physical performance tests, are used to indicate the impact of disease, impairments, and other risk factors on physical function. However, it is not always easy to differentiate between measurements of impairment and functional limitation (Guralnik and Ferrucci, 2003). A useful conceptual structure to understand both concepts is the disablement process proposed by Nagi (1991). According to this framework, impairment involves "anatomical, physiological, mental, or emotional abnormalities or loss" such as balance, visual acuity, and maximum oxygen consumption, for example. Functional limitations, on the other hand, refer to "limitation in performance at the level of the whole organism or person". Walking speed tests are the most frequently used objective physical performance tests to evaluate functional limitations of the lower limbs (Buchner, Guralnik, and Cress, 1995; Guralnik and Ferrucci, 2003; Imms and Edholm, 1981). Speed is calculated over a relatively short distance and its measurement does not influence the resistance factor (Steffen, Hacker, and Mollinger, 2002).

In longitudinal epidemiological studies, speed tests have demonstrated their capacity to predict important adverse results such as: hospitalization, dependence, and mortality (Cesari *et al*., 2005; Cesari *et al.,* 2009; Onder *et al*., 2005; Ostir, Kuo, Berges, Markides, and Ottenbacher, 2007). Furthermore, numerous advantages have been reported in terms of test-retest reliability, sensitivity to change, applicability to different population groups, etc. (Bohannon, 2009; Guralnik, Branch, Cummings, and Curb, 1989; Guralnik, *et al*., 1994). In the area of clinical research, although studies are still scarce, the findings point in the same direction. In a study carried out on a sample of primary care patients, Studenski *et al.* (2003) found that walking speed, calculated over a distance of 4 meters, was a predictor of hospitalization, health deterioration, and physical function. Furthermore, Cavazzini *et al.* (2004) performed a study to see whether a simple test based on physical performance could be incorporated into routine clinical practice. The results supported its viability and efficacy.

However, so far no reviews have analysed the use of this test as a measurement of functional limitations in elderly people in a critical and integrated way. Solway, Brooks, Lacasse, and Thomas (2001) performed a systematic review of walking tests used in patients with cardiac or respiratory pathologies; in this case, the walking tests used (6-min or 9-min walking tests) were aimed chiefly at evaluating resistance capacity and tolerance to exercise, and to obtain estimations for a series of cardiorespiratory parameters (maximum oxygen consumption, maximum exercise capacity, etc.), which correspond to measurements of impairment. Coman and Richardson (2006) performed another systematic review of the relationship between self-reporting measurements and objective measurements of physical performance in elderly people. Their results do not present an individual analysis for walking speed tests, and in some cases, the type of measurement used to establish this analysis is unclear.

The aim of this theoretical study is to evaluate the conceptual model, burden, interpretability, floor or ceiling effects, reliability, validity, and responsiveness of walking speed tests as a measurement of functional limitations in elderly people, using standardised attributes and quality criteria (Carretero-Dios and Pérez, 2007; Fernández-Ríos and Buela-Casal, 2009; Montero and León, 2007; Scientific Advisory Committee of the Medical Outcomes Trust, 2002; Terwee *et al*., 2007), thereby providing researchers and clinicians with a foundation so they can choose which walking speed test is most useful for clinical practice or in the field of research.

Methods

Search strategy

To obtain original documents, electronic searches were carried out on the following international databases: Medline, AgeLine, Embase (1980-2006), and CINAHL (1982-2006), as well as the Spanish database IME (1980-2006).

The search focused on keywords, both in the title and abstract, related with objective measurements of physical performance: comfortable gait speed OR fast gait speed OR gait speed OR gait speed test OR gait test OR gait velocity OR lower extremity test OR mobility test OR performance-based instrument* OR performance-based measure* OR performance-based method* OR physical performance battery OR physical performance measure* OR physical performance test* OR short physical performance battery OR timed walk test* OR walk* speed OR walking speed test OR walk* test OR objective test*.

Searches were also performed on Medline (1997-2006) by names of tests identified and authors. Furthermore, manual searches were performed in journals considered relevant in the field of study and in the area of ageing, both in Spain (*Atención Primaria*, *Gaceta Sanitaria, Enfermería Clínica, Revista Española de Geriatría* y *Gerontología, Medicina Clínica, Geriátrika and Revista Multidisciplinar de Gerontología*) and at an international level (Journal of the American Geriatrics Society, Journal of Gerontology. Series A: Biological Sciences and Medical Sciences, Archives of Physical Medicine and Rehabilitation, Physical Therapy, and Age and Ageing) (1997-2006). Finally, the references used in each study analysed were reviewed to obtain additional articles.

Inclusion criteria

Original studies were included in Spanish and English, performed on samples of elderly people, with a mean age of ≥ 65 and ≥ 55 in the lower range, using objective physical performance tests related with walking speed as a measure of functional limitations, either as individual tests or measurements included in multidimensional batteries, providing an individual analysis had been performed of the test. The studies had to evaluate at least one of the attributes considered in this review (*e.g.,* conceptual model, reliability, validity, responsiveness, etc.).

Exclusion criteria

Studies were excluded from this review if they made use of objective walking speed tests but did not contribute information about their performance in any of the aspects indicated in the data extraction section. Studies were also excluded if they used walking tests as a measurement of impairment. Furthermore, studies that did not present a description of the test and those which determined walking speed using special equipment were also excluded.

Study selection proces

Two researchers independently performed an individual assessment of each of the studies, reviewing the abstracts and, if necessary, the complete study. Unclear cases were resolved by consensus, and when no agreement could be reached, the final decision was made following a discussion with a third independent researcher.

Data extraction

Data extraction was carried out by the same researchers who selected the studies, following prior training, independently and resolving any disagreements by consensus with a third person. The information extracted was divided into two sections: information about the study and information about the tests. The information related to the study included the characteristics of the study and sample. Since there are no standardised criteria to evaluate the quality of objective measurements of physical performance (Terwee, Mokkink, Steultjens, and Dekker, 2006), the attributes proposed by the Scientific Advisory Committee of the Medical Outcomes Trust (2002) and the quality criteria recommended by Terwee *et al.* (2007) for instruments that measure health status and health-related quality of life (HRQOL) were adapted, selecting those properties relevant for walking speed tests.

Attributes and quality criteria selected

Appendix 1 shows a summary table of the attributes and their respective criteria, utilised to evaluate the findings of the walking speed tests. Each attribute, with the exception of predictive validity which includes two positive scores, could be classified as positive (+), indeterminate (?), negative (-) or no information available (0). Following the recommendations of Terwee *et al.* (2007) for all the attributes evaluated, a sample size of at least 50 people was considered. This table was used to create a summary table of the findings from all the walking speed tests identified.

The seven attributes considered were: 1) conceptual model: basis for the use of the test as a measurement of functional limitations; 2) burden: administration time and demands of the task for the participant and examiner; 3) interpretability: drawing up normative data in representative samples of the general population and minimal important change (MIC) based on the distribution of anchoring measures; 4) floor or ceiling effects (> 15% of the participants achieved the lowest or highest score); 5) reliability (test-retest and interrater): estimation of intraclass correlation coefficient (ICC), Kappa's coefficient or Pearson's or Spearman's correlation coefficient; 6) validity (predictive validity and construct validity: associations between measurements from the same construct and different constructs); and 7) responsiveness: determination of an effect size statistic.

Results

Studies and tests identified

375 studies were identified in the electronic searches as potentially relevant. Of this total figure, 69 met the eligibility criteria. The manual review contributed 12 additional studies and the searches performed using bibliographical references contributed a further 21 studies. In total, 102 studies were analysed, leading to the identification of 18 walking speed tests. The most frequently used test distances were: 2.44 meters (8 feet), 4 meters (13.12 feet), and 6 meters (19.69 feet), used in 12, 13, 14 studies, respectively. Usual gait speed was used with greater frequency than fast gait speed. The least frequently used tests included those that covered distances over 7.5 meters (24.61 feet) with a total of 37 studies; of these, 9 did not indicate the gait speed at which the test should be performed. Owing to the low number of studies identified in this last group of tests (fewer than 4 studies per test), which makes it difficult to perform an individual analysis on each of them, this review will focus solely on the most frequently used tests in studies on elderly people.

Conceptual model and description of the tests

The tests performed over 2.44 meters and 4 meters at usual gait speed provide the best foundation for measurements of functional limitations of the lower limbs, using Nagi's disablement process (Guralnik *et al*., 2000; Hoeymans, Feskens, van den Bos, and Kromhout, 1996) as a conceptual base. They also describe the procedures for their correct administration and scoring in greater detail. However, differences are observed in certain technical aspects, revealing a lack of standardisation, for example start and end of timing: in certain tests, the timer is started when the participant starts walking, whereas in others, varying numbers of additional meters are included to control the effects of acceleration and deceleration; method for obtaining the score: the shortest time of two journeys or the mean of the two tests, etc.

Burden

In general, evidence for this attribute is scarce in most of the studies reviewed. Only two indicate the administration time for 4-meter test (< 2 minutes and 3 minutes, respectively) (Simonsick *et al*., 1997; Studenski *et al*., 2003). The need for training is mainly reported in tests over 2.44, 4 and 6 meters, but most of them do not specify what this training would involve.

Interpretability and floor or ceiling effects

The interpretability data and floor or ceiling effects of the tests are practically non-existent. Only the test over 4 meters indicates interpretability data through normative data (Guralnik Simonsick, *et al*., 2000) and the calculation of MIC; in this case, the global estimations for a small meaningful change and substantial change were .05 m/s and .10 m/s, respectively (Perera, Mody, Woodman, and Studenski, 2006). Mean speed is most frequently reported in tests over 4, 5, and 6 meters, performed at usual gait speed; the values (comparing two studies with similar characteristics) vary between 0.88 m/s for the 4 meters test (Studenski *et al*., 2003) and 1.17 m/s for the 6 meters test (Cesari *et al*., 2005).

Reliability

Test-retest reliability is reported in all tests, with the exception of the 5 meters test carried out at usual gait speed (Table 1). The ICC values for the tests over short distances (2.44 to 4 meters), both at usual and maximum gait speed, were very close to .90 for a time interval of one week. ICC values over .90 were also recorded for the 6 meters test.

Interrater reliability studies are scarce (4 studies). The ICC values recorded were .52 for 2.44 meters test at usual gait speed (Ostchega *et al*., 2000), and 3 meters test at maximum gait speed (Sharpe *et al*., 1997), and .99 for the 6 meters test carried out at usual and maximum gait speed (Rehm-Gelin, Light, and Freund, 1997) (Table 1).

TABLE 1. Reliability of the walking speed tests.

Tests	*Reliability*	
	Test-retest	*Interrater*
	Usual gait speed	
2.44 meters (8 feet)	ICC = .72 (*n* = 136; TI= 3 weeks) (Ostchega *et al*., 2000); ICC = .79 (*n* = 105; TI = median 14 days) (Jette, Jette, Ng, Plotkin, and Bach, 1999); r_p = .90 (*n* = 104; TI= 2 weeks) (Hoeymans, Wouters, Feskens, van den Bos, and Kromhout, 1997)	ICC = .52 (*n* = 256) (Ostchega *et al*., 2000)
3 meters (9.84 feet)	ICC = .97 (*n* = 81; TI = same day) and MDC = 36% decrese between first and second measurements (Simpson, Valentine, and Worsfold, 2002); ICC = .88 (*n* = 16; TI: not reported) (Thapa, Gideon, Fought, Kormicki, and Ray, 1994)	ICC = .84 (*n* = not reported) (Sharpe *et al*., 1997) ICC = .91 (*n* = 23) (Fox, Felsenthal, Hebel, Zimmerman, and Magaziner, 1996).
4 meters (13.12 feet)	ICC =.84 (*n* = not reported; 1 week) (Studenski *et al*., 2003); ICC = .86; .80 and .89 (*n* = 91, 90 and 84; TI = three pairs of weeks), respectively (Ostir, Volpato, Fried, Chaves, and Guralnik, 2002); ICC = .88; .89; .88; .87 (*n* = 102; four measurements with one week apart during 24 weeks), respectively (Guralnik *et al*., 1999); Walking speed remain reliable over an extended period time (*n* = 99; TI = measurements weekly over a 6-month period) (Ferrucci *et al*., 1996)[a]	Not reported

TABLE 1. Reliability of the walking speed tests. (Cont.)

Tests	*Reliability*	
	Test-retest	*Interrater*
	Usual gait speed	
5 meters (16.40 feet)	Not reported	Not reported
6 meters (19.69 feet)	ICC = .97 (*n* = 96; TI = same day) (Steffen *et al.,* 2002); ICC .92 (*n* = 20; TI = 48 hours) (Rehm-Gelin *et al.*, 1997); r_p = .95 (n = 30; TI = 1 week) (Curcio, Gómez, and Galeano, 2000); ICC =.92 (*n* = 10; TI = 1 week) (Thomas and Hageman, 2002); ICC = .90 (*n* = 23; TI = same day) (Brusse, Zimdars, Zalewski, and Steffen, 2005)[c]; ICC = .97 (*n* = 30; TI= one day[d] and one week[e]) (Sherrington and Lord, 2005)	ICC = .99 (n = 20) (Rehm-Gelin *et al.*, 1997)
	Fast gait speed	
3 meters (9.84 feet)	r = .80 (*n* = not reported; TI = 2 weeks) (Seeman *et al.*, 1994); ICC =.78 (*n* = 199 ; TI = 48 hours) (Tager, Swanson, and Satariano, 1998)	ICC = .52 (*n* = not reported) (Sharpe *et al.*, 1997) ICC =.92 (*n* = 23) (Fox *et al.*, 1996)
5 meters (16.40 feet)	r = 0.93 and 0.92 (n = 1077 and 18; TI = same day and 1 year), respectively (Nagasaki *et al.*, 1996)	Not reported
6 meters (19.69 feet)	ICC = .96 (*n* = 96; TI = same day) (Steffen *et al.*, 2002); ICC = .90 (*n* = 20; TI = 48 hours) (Rehm-Gelin *et al.*, 1997); ICC = .95 (*n* = 9; TI = 1 week) (Thomas and Hageman, 2002)[b]; ICC = .94 (*n* = 23; TI = same day) (Brusse *et al.,* 2005)[c]; ICC = .94 (*n* = 30; TI= one day[d] and one week[e]) (Sherrington and Lord, 2005)	ICC = .99 (*n* = 20) (Rehm-Gelin *et al.*, 1997)

Note. [a]Analysis of multiple sequential measures of walking speed; [b]Dementia patients; [c]Parkinson disease sample; [d]Hospital inpatients sample; [e]Community dwellers sample; ICC= intraclass correlation coefficients; MDC = minimal detectable change; r = correlation coefficient; r_p= Pearson correlation coefficient; TI = time interval.

Validity

Twenty studies provide data about predictive validity (Table 2). 50% of these studies focus on 4 and 6-meter tests carried out at usual gait speed; 6 and 4 studies, respectively. The findings presented are linked to: hospitalization, health deterioration, dependence in ADL, disability in mobility, start of progressive, and catastrophic disability in ADL, mobility and disability in the upper limbs, mortality, cognitive deterioration and hip fracture (Atkinson *et al*., 2005; Cesari *et al*., 2005; Dargent-Molina, Douchin, Cormier, Meunier, and Breart, 2002; Dargent-Molina *et al*., 1996; Fried, Bandeen-Roche, Chaves, and Johnson, 2000; Guralnik *et al*., 2000; Onder *et al*., 2005; Perera, Studenski, Chandler, and Guralnik, 2005; Rolland *et al*., 2006; Studenski *et al*., 2003). The monitoring periods varied between 1 and 6 years. All the tests reported construct validity in one way or another (Table 2). Nine studies evaluate the association with disability measurements but none of them formulates specific prior hypotheses. The correlations vary between .13 (2.44 meters test and ADL subscale on the WHO [World Health Organization] questionnaire) (Hoeymans *et al*., 1996), and .74 (6 meters test and ADL/IADL [instrumental activities of daily living] scale) (Creel, Light, and Thigpen, 2001). Three studies of the 5 meters test provide a concurrent analysis between usual gait speed and maximum gait speed. The values observed, in this case, were very heterogeneous (.30, .62, and .76, respectively) (Kinugasa, Nagasaki, Furuna, and Itoh, 1996; Nagasaki, Itoh,

TABLE 2. Validity of the walking speed tests.

Tests	*Validity*		
	Predictive validity	*Construct validity*	
		Disability measures	*Functional limitation measures*
	Usual gait speed		
2.44 meters (8 feet)	- Onset of ADL (ORs 95% CI: 5.4 [category 1: slowest]; 4.3 [category 2]; 3.6 [category 3] when compared to category 4 [fastest]) and mobility (ORs 3.4; 2.6 and 2.1) related disabilities over a 2 years period (Ostir, Markides, Black, and Goodwin, 1998) - Mortality (ORs 95% CI: 3.64 [quartile 1: slowest]; 2.57 [quartile 2] and 2.16 [quartile 3] compared to quartile 4 [fastest] over a 2 years period (Markides *et al*., 2001)	- r_s = .13 and .16 with ADL; r_s = .17 and .33 with IADL, 1990 and 1993, respectively (Hoeymans *et al.,* 1996)[c] - Associated with increased levels of disability (.13) (Peek, Ottenbacher, Markides, and Ostir, 2003)[d]	- r_s = .48 with chair stands test (Guralnik *et al*., 1994) - r_s = .17 and .32 with mobility scale, 1990 and 1993, respectively (Hoeymans *et al.,* 1996)[c] - r_p = -.81, -.77 y -.94 with chair stands test (baseline, week 6 and week 12), respectively (Schaubert and Bohannon, 2005) - Comparison with reported ability to walk across a small room (men: S = 71% and Sp = 91%; women: S = 82% and Sp= 92%) (Merrill, Seeman, Kasl, and Berkmann*et,* 1997)
3 meters (9.84 feet)	Recurrent falls (≥ 2 falls within 1 year): OR 2.2, 95% CI 1.1-4.1) (Stel, Smit, Pluijm, and Lips, 2003)	- Not significant difference between the severely disabled and not disabled or moderately disabled (Dunnet C-test, p > .05); r = .39 with KI and r = .63 with functional ability scales (Nybo *et al*., 2001) - r_s = .47 with scale of ADL capacity[c]; high rates of disability were reported by those men who were unable to complete the walking test (Kivinen, Sulkava, Halonen, and Nissinen, 1998)	r_p = .54 with chair stands test (Thapa *et al*., 1994)
4 meters (13.12 feet)	- Progressive and catastrophic disability in ADL (RRs 95% CI .65 and .72, mobility (.27 and .57 and upper extremity function (.70 and .64), during 3 years of follow-up (Onder *et al.,* 2005) - Hospitalization (OR .62), decline in health (OR .49 decline in Global Health) and OR .63 decline in Euroqol Score) and decline in function (OR .72 SF-36 ≥ 70) for the first 12 months of follow-up (Studenski *et al*., 2003)[a,b] - Risk of ADL disability (AUC SS .75, and .69, and GS .70, and .67, 1, and 4 years, respectively) and mobility disability (AUC SS .70 and .69 and GS .67 and .65, 1 and 4 years, respectively), during 1 to 6 years of follow-up (Guralnik *et al.,* 2000) - Disability in mobility after 18 months (OR 2.04, 95% CI 1.02-4.09 [Model A: with difficulty for walking ½ mile]) (Fried *et al*., 2000) - Mortality (HR 2.23, 95% CI 1.44-3.46) within 5 years (Perera *et al.,* 2005) - Physical decline (OR .70) and combined decline [physical and cognitive] (OR .46), during 3 years of follow-up (Atkinson *et al.,* 2005)	Not reported	- Comparison with reported need of help to walk: k = .55 and difficulty to walk: k = .41 (Ferrer, Lamarca, Orfila, and Alonso, 1999) - r_s = 0.87 with SPPB (Cesari *et al*., 2006)
5 meters (16.40 feet)	- Onset of functional dependence (65-74 years: HR 2.43 and ≥ 75 years: HR 6.18), during 6 years of follow-up (Shinkai *et al.,* 2000) - Recurrent nonsyncopal falls (RR 1.6, 95% CI 1.2-2.2) [unadjusted ORs], during 1 year of follow-up (Nevitt, Cummings, Kidd, and Black, 1989) - Not predict maintain physical functioning (IADL) or mortality (Furuna, Nagasaki, Nishizawa, and Okuzumi, 1998)	Not reported	- r_p = .30 with M-Walk (Nagasaki *et al*., 1995a)[e] - r_p = .76 with M-Walk (Nagasaki *et al*., 1995b)[e] - r_p = .62 with M-Walk (Kinugasa *et al*., 1996)[e]
6meters (19.69 feet)	- Persistent lower extremity limitation, persistent severe lower extremity limitation, death and hospitalization (≤ 1 m/s RRs 95% CI 2.20; 2.29; 1.64 and 1.48, respectively), during 1 year of follow-up (Cesari *et al.,* 2005) - Risk of hip fracture (RR 1.4 for 1 SD decrese [95% CI 1.1-1.6]), during a mean follow-up of 1.94 years (Dargent-Molina *et al*., 2002; Dargent-Molina *et al.,* 1996) - Mortality (HR 1.50, 95% CI .97-2.33), during a mean follow-up of 3.8 years (Rolland *et al.,* 2006)	- r_s = .74 with ADL/IADL scale and r_s= -.68 with BI (Creel *et al.,* 2001) - Comparison with BI (cut-off point .42 m/s) S = 78% and Sp = 91% (Curcio *et al.,* 2000) - Not correlated with UPDRS/ADL (Brusse *et al*., 2005) - A statistical significant difference was found between medium and low functioning (p ≤ 0.001) (Berkman *et al.,* 1993)	- r_p = .85 with TUG (Creel *et al*., 2001) - r_p = -.67 with TUG (Brusse *et al.,* 2005)[f] - Self-report: walking 1 mile (r_p = .25); walking up 20 steps (r_p = .27) and lifting/carrying 20 lbs. (r_p = 0.19); Performa sc nce-based: SPPB (r_p = .49), Health ABC performance ore (r_p = .71), chair stands test (r_p = .41) and 20-m walking speed test (r_p= .68) (Simonsick *et al.,* 2001)

TABLE 2. Validity of the walking speed tests.

Tests	*Validity*		
	Predictive validity	*Construct validity*	
		Disability measures	*Functional limitation measures*
3 meters (9.84 feet)	Not reported	Not reported	- r = -.48 with SPPB (Seeman *et al.*, 1994) - r_p = .29 whith chair stands test (women) (Guralnik *et al.*, 1994)
5 meters (16.40 feet)	- Onset of functional dependence (65-74 years: HR 5.15 and $\geq$ 75 years: HR 3.45), during 6 years of follow-up (Shinkai *et al.*, 2000) - Maintain physical functioning [IADL] (OR .38) and mortality (OR .22) over a 4-year period (Furuna *et al.*, 1998) - Falls [1 or more falls] (RR .23, 95% CI .11-.50), during 1 year of follow-up (Chu, Chi, and Chiu, 2005)	Not reported	- r_p= .30 with P-Walk (Nagasaki *et al.*, 1995a)[e] - r_p= .69 with P-Walk (Nagasaki *et al.*, 1995b)[e] - r_p= .62 with P-Walk (Kinugasa *et al.*, 1996)[e]
6 meters (19.69 feet)	ADL dependence (RR 2.4, 95% CI 1.4-4.2) (Gill, Richardson, and Tinetti, 1995; Gill, Williams, and Tinetti, 1995) and (OR 5.3, 95% CI 3.0-9.3) (Gill, Williams, Mendes de León, and Tinetti, 1997) during 1 and 3 years of follow-up, respectively	- r_s = .71 with ADL/IADL scale and r_s = -.56 with BI (Creel *et al.*, 2001) - Participants who became dependent at one year had greater declines compared with those who remained independent ($p < .001$) (Gill *et al.*, 1997)	- r_p = .82 with TUG (Creel *et al.*, 2001) - r_p = -.69 with TUG and r_p= .89 with CGS (Brusse *et al.*, 2005)[f]

Note. [a]Differences in areas under the receiver operating characteristic (ROC) curves for these models were used to assess added value; [b]Primary care sample; [c]Adapted from the World Health Organization questionnaire; [d]Structural equation models estimated with LISREL 8.30; [e]Walking speed was measured over a distance of 5 meters; [f]Walking speed was measured over a distance of 6 meters; ADL= activities of daily living; AUC = area under curve; BI= Barthel Index; CI = confidence interval; CGS = comfortable gait speed; GS = gait speed; HR= hazard ratio; IADL= instrumental activities of daily living; k = Kappa statistical; KI = Katz Index; M-Walk= maximal walking speed; OR = odds ratio; P-Walk= preferred walking speed; r_p = Pearson correlation coefficient; RR= relative risk; r_s = Spearman correlation coefficient; S = sensitivity; SD = standard desviation; SF-36 = Medical Outcomes Study 36-item short form heatlh survey; Sp = Specificity; SPPB = Short Physical Performance Battery; SS = summary performance score; TUG = Timed Up & Go test; UPDRS = Unified Parkinson´s Disease Rating Scale.

and Furuna, 1995a; Nagasaki, Itoh, and Furuna, 1995b). Along these same lines are the results observed when correlating walking tests with other objective measurements of physical performance in lower limbs (8-foot test with a test that evaluates capacity to stand up from a seated position: .48 and -.81) (Guralnik Simonsick, *et al*., 1994; Schaubert and Bohannon, 2005).

Responsiveness

Responsiveness is another of the attributes with scant information (5 intervention studies (Bean *et al*., 2004; English, Hiller, Stiller, and Warden-Flood, 2006; Sayers *et al*., 2003; Sharpe *et al*., 1997; Thomas and Hageman, 2003) and 3 observational studies (Guralnik *et al*., 1999; Lan, Deeg, Guralnik, and Melzer, 2003; Onder *et al*., 2002). The interventions are aimed at training programmes to improve muscular strength, balance, and mobility. Only one study includes the 3-meter test carried out at usual and fast gait speed on the intervention group using the effect size statistic (Sharpe *et al*., 1997). As for the observational studies, only one compares the responsiveness of the 3-meter test, carried out at fast gait speed, with the mobility limitation index (MOBLI index); the responsiveness index (RI) of the MOBLI was only higher than the speed test for the group that reported a deterioration in mobility (Lan *et al*., 2003).

Summary assessment

Table 3 presents a summary assessment of the walking speed test attributes. The 4 meters test achieved the best rating with a total of six positive scores. However, certain negative aspects and gaps are observed in some of the attributes: negative score in floor or ceiling effect (community sample with a low proportion of people with difficulty walking), absence of interrater reliability data, and uncertain results for responsiveness.

In spite of the above, it is important to highlight the positive scores in the attributes test-retest reliability and predictive validity, the latter being the only attribute to obtain the maximum score possible, since it not only predicts numerous adverse results but is also used in more than one area of study: community and primary healthcare.

TABLE 3. Summary of evaluation of the quality of the walking speed tests.[a]

Tests	*Conceptual model*	*Burden*	*Interpretability*	*Floor or ceiling effects*	*Reliability*		*Validity*		*Responsiveness*
					Test-retest	*Interrater*	*Predictive*	*Construct*	
Usual gait speed									
2.44 meters (8 feet)	+	?	+	-	+	-	+	?	-
3 meters (9.84 feet)	?	?	+	0	?	+	+	?	+
4 meters (13.12 feet)	+	+	+	-	+	0	++	+	?
5 meters 16.40 feet)	?	0	0	-	0	0	+	+	?
6 meters (19.69 feet)	?	?	0	0	?	+	+	+	-

TABLE 3. Summary of evaluation of the quality of the walking speed tests.[a] (Cont.)

Tests	*Conceptual model*	*Burden*	*Interpretability*	*Floor or ceiling effects*	*Reliability*		*Validity*		*Responsiveness*
					Test-retest	*Interrater*	*Predictive*	*Construct*	
Fast gait speed									
3 meters (9.84 feet)	?	?	0	0	?	?	0	-	+
5 meters (16.40 feet)	?	0	0	0	?	0	+	+	-
6 meters (19.69 feet)	?	?	0	0	?	+	+	+	+

Note. [a](+) = positive rating; (-) = negative rating; (?) = indeterminate rating; (0) = no information available.

Discussion

This review has identified 18 walking speed tests as measurements of functional limitations in elderly people. The evidence obtained about these tests has been evaluated using standardised attributes and quality criteria (Terwee *et al*., 2007; Terwee *et al*., 2006). The most frequently used tests covered distances of 2.44 meters, 4 meters, and 6 meters, carried out at usual gait speed. These results contradict the findings of a recent systematic review of the evaluation of walking speed in clinical research, which signals that the most frequently used distance is 10-meter, especially in patients with neurological pathologies (Graham, Ostir, Fisher, and Ottenbacher, 2008). However, in that case, the selection criteria only included methodological aspects related with the administration of the tests, without taking into account the type of population or psychometric information provided by the studies, as is the case with this review.

A lack of standardisation is observed in the administration of most tests. Furthermore, there is little information about the training process of the examiners. These findings are similar to those reported in the systematic review performed by Graham *et al.* (2008), which indicated the high degree of methodological variability in the administration of the different tests.

There are few studies about the interpretability of the tests. Most of them are based on normative data in samples of the general population and only one study reports on MIC. The estimation of MIC is particularly useful as it helps to plan, evaluate, and compare the effectiveness of interventions that use the results of objective physical performance tests as a measurement (Perera *et al*., 2006). Mean speed is reported in most tests. In general, as the distance increases so does the walking speed. Similar results have been reported by Cesari *et al.* (2005), who attributed the differences to a possible effect of acceleration from the initial stationary position.

In relation to reliability, test-retest reliability is the most frequently reported attribute. The values found were higher than the recommended quality standards, for both group and individual decisions. However, there are few interrater reliability tests and, therefore, there is no information about the level of training or test administration protocol required for reliable application.

In terms of validity, most of the data focus on predictive validity. The most frequently reported adverse results were: dependence in ADL, deterioration of mobility, and mortality. This is because objective measurements of physical performance, just like ADL measurements, provide highly valuable information as indicators of adverse results (Studenski *et al*., 2003).

Construct validity was established through association with other objective physical performance tests or self-reporting measurements for functional limitations of the lower limbs and disability measurements. The correlation coefficients, as expected, were higher between measurements used on the same construct than measurements used on different constructs. Similar results have been reported by Coman and Richardson (2006), examining studies that relate self-reports of functional limitations and disability with objective measurements of physical performance. However, very few studies have analysed two or more walking speed tests concurrently. The information yielded in this regard is very heterogeneous and offers surprisingly low values.

Responsiveness is another of the least reported psychometric properties. The data in this regard are insufficient to determine the capacity of the walking test to detect changes over time, either changes derived from intervention or from other situations similar to functional limitations such as illnesses or impairments. Based on current evidence, these tests seem to be fairly unresponsive.

As for methodological limitations, two considerations should be taken into account. The first is related to the quality criteria used. Due to the lack of specific quality standards for objective measurements of physical performance, attributes, and quality criteria developed for measurements of HRQOL had to be adapted for this study; hence some of these criteria are not fully suited to this review, for example, reliability through internal consistency. The second consideration is linked to search strategies. Owing to the fact that the database thesauri contain no specific walking speed test descriptors, the searches focused solely on keywords. However, to guarantee the exhaustiveness of this process, multiple terms were selected using different databases and completing the electronic searches with manual reviews and reference tracking.

In conclusion, test-retest reliability and predictive validity data support the use of walking speed tests in epidemiological studies as predictors of important adverse events related to health in elderly people. However, further evidence is required to support their viability and applicability in clinical practice, either for screening purposes or to monitor and evaluate change.

References

Atkinson, H.H., Cesari, M., Kritchevsky, S.B., Penninx, B.W.J., Fried, L.P., Guralnik, J. M., and Williamson, J.D. (2005). Predictors of combined cognitive and physical decline. *Journal of the American Geriatric Society, 53,* 1197-1202.

Bean, J.F., Herman, S., Kiely, D.K., Frey I.C., Leveille, S.G., Fielding, R.A., and Frontera, W.R. (2004). Velocity Exercise Specific to Task (InVEST) training: A pilot study exploring effects on leg power, balance, and mobility in community-dwelling older women. *Journal of the American Geriatric Society, 52,* 799-804.

Berkman, L.F., Seeman, T.E., Albert, M., Blazer, D., Kahn, R., Mohs, R., Finch, C., Schneider, E., Cotman, C., and McClearn, G. (1993). High, usual and impaired functioning in community-dwelling older men and women: findings from the MacArthur Foundation Research Network on Successful Aging. *Journal of Clinical Epidemiology, 46,* 1129-1140.

Bohannon, R.W. (2009). Measurement of gait speed of older adults is feasible and informative in a home-care setting. *Journal of Geriatric Physical Therapy, 32,* 22-23.

Brusse, K.J., Zimdars, S., Zalewski, K.R., and Steffen, T.M. (2005). Testing functional performance in people with Parkinson disease. *Physical Therapy, 85,* 134-141.

Buchner, D.M., Guralnik, J.M., and Cress, M.E. (1995). The clinical assessment of gait, balance, and mobility in older adults. In L.Z. Rubenstein, D. Wieland, and R. Bernabei (Eds.), *Geriatric Assessment Technology: The State of the Art.* (pp. 75-88). Milan: Editrice Kurtis.

Carretero-Dios, H. and Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology, 7,* 863-882.

Cavazzini, C., Conti, M., Bandinelli, S., Gangemi, S., Gallinella, M., Lauretani, F., Lucci, G., Windham, B.G., Guralnik, J.M., and Ferrucci, L. (2004). Screening for poor performance of lower extremity in primary care: the Camucia Project. *Aging Clinical and Experimental Research, 16,* 331-336.

Cesari, M., Kritchevsky, S.B., Penninx, B.W., Nicklas, B.J., Simonsick, E.M., Newman, A.B., Tylavsky, F.A., Brach, J.S., Satterfield, S., Bauer, D.C., Visser, M., Rubin, S.M., Harris, T.B., and Pahor, M. (2005). Prognostic value of usual gait speed in well-functioning older people—results from the Health, Aging and Body Composition Study. *Journal of the American Geriatrics Society, 53,* 1675-1680.

Cesari, M., Onder, G., Russo, A., Zamboni, V., Barillaro, C., Ferrucci, L. Pahor, M., Bernabei, R., and Landi, F. (2006). Comorbidity and physical function: Results from the aging and longevity study in the Sirente geographic area (ilSIRENTE study). *Gerontology, 52,* 24-32.

Cesari, M., Pahor, M., Marzetti, E., Zamboni, V., Colloca, G., Tosato, M., Patel, K.V., Tovar, J.J., and Markides, K. (2009). Self-assessed health status, walking speed and mortality in older Mexican-Americans. *Gerontology, 55,* 194-201.

Chu, L.W., Chi, I., and Chiu, A.Y. (2005). Incidence and predictors of falls in the chinese elderly. *Annals of the Academy of Medicine, Singapore, 34,* 60-72.

Coman, L. and Richardson, J. (2006). Relationship between self-Report and performance measures of function: A systematic review. *Canadian Journal on Aging, 25,* 253-270.

Creel, G.L., Light, K.E., and Thigpen, M.T. (2001). Concurrent and construct validity of scores on the Timed Movement Battery. *Physical Therapy, 81,* 789-798.

Curcio, C.L., Gómez, J.F., and Galeano, I.C. (2000). Validez y reproducibilidad de medidas de evaluación funcional basadas en la ejecución. *Revista Española de Geriatría y Gerontología, 35,* 82-88.

Dargent-Molina, P., Douchin, M.N., Cormier, C., Meunier, P.J., and Breart, G. (2002). Use of clinical risk factors in elderly women with low bone mineral density to identify women at higher risk of hip fracture: The EPIDOS prospective study. *Osteoporosis International, 13,* 593-599.

Dargent-Molina, P., Favier, F., Grandjean, H., Baudoin, C., Schott, A.M., Hausherr, E., Meunier, P.J., and Breart, G. (1996). Fall-related factors and risk of hip fracture: The EPIDOS prospective study. *Lancet, 348,* 145-149.

English, C.K., Hillier, S.L., Stiller, K., and Warden-Flood, A. (2006). The sensitivity of three commonly used outcome measures to detect change amongst patients receiving inpatient rehabilitation following stroke. *Clinical Rehabilitation, 20,* 52-55.

Fernández-Ríos, L. and Buela-Casal, G. (2009). Standards for the preparation and writing of Psychology review articles. *International Journal of Clinical and Health Psychology, 9,* 329-344.

Ferrer, M., Lamarca, R., Orfila, F., and Alonso, J. (1999). Comparison of performance-based and self-rated functional capacity in Spanish elderly. *American Journal of Epidemiology, 149,* 228-235.

Ferrucci, L., Guralnik, J.M., Salive, M.E., Fried, L.P., Bandeen-Roche, K., Brock, D.B., Simonsick, E.M., Corti, M.C., and Zeger, S.L. (1996). Effect of age and severity of disability on short-term variation in walking speed: The Women's Health and Aging Study. *Journal of Clinical Epidemiology, 49,* 1089-1096.

Fox, K.M., Felsenthal, G., Hebel, J.R., Zimmerman, S.I., and Magaziner, J. (1996). A portable neuromuscular function assessment for studying recovery from hip fracture. *Archives of Physical Medicine and Rehabilitation, 77,* 171-176.

Fried, L.P., Bandeen-Roche, K., Chaves, P.H., and Johnson, B.A. (2000). Preclinical mobility disability predicts incident mobility disability in older women. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 55,* M43-M52.

Furuna, T., Nagasaki, H., Nishizawa, S., and Okuzumi, H. (1998). Longitudinal change in the physical performance of older adults in tne community. *Journal of the Japanese Physical Therapy Association, 1,* 1-5.

Gill, T.M., Richardson, E.D., and Tinetti, M.E. (1995). Evaluating the risk of dependence in activities of daily living among community-living older adults with mild to moderate cognitive impairment. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 50,* M235-M241.

Gill, T.M., Williams, C.S., Mendes de León, C.F., and Tinetti, M.E. (1997). The role of change in physical performance in determining risk for dependence in activities of daily living among nondisabled community-living elderly persons. *Journal of Clinical Epidemiology, 50,* 765-772.

Gill, T.M., Williams, C.S., and Tinetti, M.E. (1995). Assessing risk for the onset of functional dependence among older adults: The role of physical performance. *Journal of the American Geriatric Society, 43,* 603-609.

Graham, J.E., Ostir, G.V., Fisher, S.R., and Ottenbacher, K.J. (2008). Assessing walking speed in clinical research: A systematic review. *Journal of Evaluation in Clinical Practice, 14,* 552-562.

Guralnik, J.M., Branch, L.G., Cummings, S.R., and Curb, J.D. (1989). Physical performance measures in aging research. *Journal of Gerontology, 44,* M141-M146.

Guralnik, J.M. and Ferrucci, L. (2003). Assessing the building blocks of function: Utilizing measures of functional limitation. *American Journal of Preventive Medicine, 25,* 112-121.

Guralnik, J.M., Ferrucci, L., Penninx, B.W., Kasper, J.D., Leveille, S.G., Bandeen-Roche, K., and Fried, L.P. (1999). New and worsening conditions and change in physical and cognitive performance during weekly evaluations over 6 months: The Women's Health and Aging Study. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 54,* M410-M422.

Guralnik, J.M., Ferrucci, L., Pieper, C.F., Leveille, S.G., Markides, K.S., Ostir, G.V., Studenski, S., Berkman, L.F., and Wallace, R.B. (2000). Lower extremity function and subsequent disability: Consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 55,* M221-M231.

Guralnik, J.M., Seeman, T.E., Tinetti, M.E., Nevitt, M.C., and Berkman, L.F. (1994). Validation and use of performance measures of functioning in a non-disabled older population: MacArthur studies of successful aging. *Aging (Milan, Italy), 6,* 410-419.

Guralnik, J.M., Simonsick, E.M., Ferrucci, L., Glynn, R.J., Berkman, L.F., Blazer, D.G., Scherr, P.A., and Wallace, R.B. (1994). A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *Journal of Gerontology, 49,* M85-M94.

Hoeymans, N., Feskens, E.J., van den Bos, G.A., and Kromhout, D. (1996). Measuring functional status: Cross-sectional and longitudinal associations between performance and self-report (Zutphen Elderly Study 1990-1993). *Journal of Clinical Epidemiology, 49,* 1103-1110.

Hoeymans, N., Wouters, E.R., Feskens, E.J., van den Bos, G.A., and Kromhout, D. (1997). Reproducibility of performance-based and self-reported measures of functional status. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 52,* M363-M368.

Imms, F.J. and Edholm, O.G. (1981). Studies of gait and mobility in the elderly. *Age and Ageing, 10,* 147-156.

Jette, A.M., Jette, D.U., Ng, J., Plotkin, D.J., and Bach, M.A. (1999). Are performance-based measures sufficiently reliable for use in multicenter trials? Musculoskeletal Impairment (MSI) Study Group. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 54,* M3-M6.

Kinugasa, T., Nagasaki, H., Furuna, T., and Itoh, H. (1996). Physical performance measures for characterizing high functioning older persons. *Journal of Aging and Physical Activity, 4,* 338-348.

Kivinen, P., Sulkava, R., Halonen, P., and Nissinen, A. (1998). Self-reported and performance-based functional status and associated factors among elderly men: The Finnish cohorts of the Seven Countries Study. *Journal of Clinical Epidemiology, 51,* 1243-1252.

Lan, T., Deeg, D. J.H., Guralnik, J.M., and Melzer, D. (2003). Responsiveness of the index of mobility limitation: Comparison with gait speed alone in the Longitudinal Aging Study Amsterdam. *Journals of Gerontology: Serie A: Biological Sciences and Medical Science, 58,* 721-727.

Markides, K.S., Black, S.A., Ostir, G.V., Angel, R.J., Guralnik, J.M., and Lichtenstein, M. (2001). Lower body function and mortality in Mexican American elderly people. *Journals of Gerontology: Serie A: Biological Sciences and Medical Science, 56,* M243-M247.

Merrill, S.S., Seeman, T.E., Kasl, S.V., and Berkman, L.F. (1997). Gender differences in the comparison of self-reported disability and performance measures. *Journals of Gerontology: Serie A: Biological Sciences and Medical Science, 52,* M19-M26.

Montero, I. and León, O.G. (2007). A guide for naming research studies in Psychology. *International Journal of Clinical and Health Psychology, 7,* 847-862.

Nagasaki, H., Itoh, H., and Furuna, T. (1995a). A physical fitness model of older adults. *Aging (Milan, Italy), 7,* 392-397.

Nagasaki, H., Itoh, H., and Furuna, T. (1995b). The structure underlying physical performance measures for older adults in the community. *Aging (Milan, Italy), 7,* 451-458.

Nagasaki, H., Itoh, H., Hashizume, K., Furuna, T., Maruyama, H., and Kinugasa, T. (1996). Walking patterns and finger rhythm of older adults. *Perceptual and Motor Skills, 82,* 435-447.

Nagi, S. (1991). Disability concepts revisited: Implications for prevention. In A. Pope and A. Tarlov (Eds.), *Disability in America: Toward a national agenda for prevention.* (pp. 309-327). Washington, D.C: National Academy Press.

Nevitt, M.C., Cummings, S.R., Kidd, S., and Black, D. (1989). Risk factors for recurrent nonsyncopal falls. A prospective study. *JAMA: The Journal of the American Medical Association, 261,* 2663-2668.

Nybo, H., Gaist, D., Jeune, B., McGue, M., Vaupel, J.W., and Christensen, K. (2001). Functional status and self-rated health in 2,262 nonagenarians: The Danish 1905 Cohort Survey. *Journal of the American Geriatric Society, 49,* 601-609.

Onder, G., Penninx, B.W., Ferrucci, L., Fried, L.P., Guralnik, J.M., and Pahor, M. (2005). Measures of physical performance and risk for progressive and catastrophic disability: Results from the Women's Health and Aging Study. *Journals of Gerontoloty: Serie A: Biological Sciences and Medical Sciences, 60,* 74-79.

Onder, G., Penninx, B.W., Lapuerta, P., Fried, L.P., Ostir, G.V., Guralnik, J.M., and Pahor, M. (2002). Change in physical performance over time in older women: the Women's Health and Aging Study. *Journals of Gerontoloty: Serie A: Biological Sciences and Medical Sciences, 57,* M289-M293.

Ostchega, Y., Harris, T.B., Hirsch, R., Parsons, V.L., Kington, R., and Katzoff, M. (2000). Reliability and prevalence of physical performance examination assessing mobility and balance in older persons in the US: Data from the Third National Health and Nutrition Examination Survey. *Journal of the American Geriatric Society, 48,* 1136-1141.

Ostir, G.V., Kuo, Y.F., Berges, I.M., Markides, K.S., and Ottenbacher, K.J. (2007). Measures of lower body function and risk of mortality over 7 years of follow-up. *American Journal of Epidemiology, 166,* 599-605.

Ostir, G.V., Markides, K.S., Black, S.A., and Goodwin, J.S. (1998). Lower body functioning as a predictor of subsequent disability among older Mexican Americans. *Journals of Gerontoloty: Serie A: Biological Sciences and Medical Sciences, 53,* M491-M495.

Ostir, G.V., Volpato, S., Fried, L.P., Chaves, P., and Guralnik, J.M. (2002). Reliability and sensitivity to change assessed for a summary measure of lower body function: Results from the Women's Health and Aging Study. *Journal of Clinical Epidemiology, 55,* 916-921.

Peek, M., Ottenbacher, K., Markides, K., and Ostir, G. (2003). Examining the disablement process among older Mexican American adults. *Social Science & Medicine, 57,* 413-425.

Perera, S., Mody, S.H., Woodman, R.C., and Studenski, S.A. (2006). Meaningful change and responsiveness in common physical performance measures in older adults. *Journal of the American Geriatric Society, 54,* 743-749.

Perera, S., Studenski, S., Chandler, J.M., and Guralnik, J.M. (2005). Magnitude and patterns of decline in health and function in 1 year affect subsequent 5-year survival. *Journals of Gerontoloty: Serie A: Biological Sciences and Medical Sciences, 60,* 894-900.

Rehm-Gelin, S.L., Light, K.E., and Freund, J.E. (1997). Reliability of timed-functional movements for clinical assessment of a frail elderly population. *Physical & Occupational Therapy in Geriatrics, 15,* 1-19.

Rolland, Y., Lauwers-Cances, V., Cesari, M., Vellas, B., Pahor, M., and Grandjean, H. (2006). Physical performance measures as predictors of mortality in a cohort of community-dwelling older French women. *European Journal of Epidemiology, 21,* 113-122.

Sayers, S., Bean, J., Cuoco, A., LeBrasseur, N., Jette, A., and Fielding, R. (2003). Changes in function and disability after resistance training: Does velocity matter? A pilot study. *American Journal of Physical Medicine & Rehabilitation, 82,* 605-613.

Schaubert, K.L. and Bohannon, R.W. (2005). Reliability and validity of three strength measures obtained from community-dwelling elderly persons. *Journal of Strength and Conditioning Research, 19,* 717-720.

Scientific Advisory Committee of the Medical Outcomes Trust (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research, 11,* 193-205.

Seeman, T.E., Charpentier, P.A., Berkman, L.F., Tinetti, M.E., Guralnik, J.M., Albert, M., Blazer, D., and Rowe, J.W. (1994). Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *Journal of Gerontology, 49,* M97-M108.

Sharpe, P.A., Jackson, K.L., White, C., Vaca, V.L., Hickey, T., Gu, J., and Otterness, C. (1997). Effects of a one-year physical activity intervention for older adults at congregate nutrition sites. *The Gerontologist, 37,* 208-215.

Sherrington, C. and Lord, S.R. (2005). Reliability of simple portable tests of physical performance in older people after hip fracture. *Clinical Rehabilitation, 19,* 496-504.

Shinkai, S., Watanabe, S., Kumagai, S., Fujiwara, Y., Amano, H., Yoshida, H., Ishizaki, T., Yukawa, H., Suzuki, T., and Shibata, H. (2000). Walking speed as a good predictor for the onset of functional dependence in a Japanese rural community population. *Age and Ageing, 29,* 441-446.

Simonsick, E.M., Maffeo, C.E., Rogers, S.K., Skinner, E.A., Davis, D., Guralnik, J. M., and Fried, L.P. (1997). Methodology and feasibility of a home-based examination in disabled older women: The Women's Health and Aging Study. *Journals of Gerontology: Series A: Biological Sciences and Medical Sciences, 52,* M264-M274.

Simonsick, E.M., Newman, A.B., Nevitt, M.C., Kritchevsky, S.B., Ferrucci, L., Guralnik, J.M., and Harris, T. (2001). Measuring higher level physical function in well-functioning older adults: Expanding familiar approaches in the Health ABC study. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 56,* M644-M649.

Simpson, J.M., Valentine, J., and Worsfold, C. (2002). The Standardized Three-metre Walking Test for elderly people (WALK3m): repeatability and real change. *Clinical Rehabilitation, 16,* 843-850.

Solway, S., Brooks, D., Lacasse, Y., and Thomas, S. (2001). A qualitative systematic overview of the measurement properties of functional walk tests used in the cardiorespiratory domain. *Chest, 119,* 256-270.

Steffen, T.M., Hacker, T.A., and Mollinger, L. (2002). Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Physical Therapy, 82,* 128-137.

Stel, V.S., Smit, J.H., Pluijm, S.M., and Lips, P. (2003). Balance and mobility performance as treatable risk factors for recurrent falling in older persons. *Journal of Clinical Epidemiology, 56,* 659-668.

Studenski, S., Perera, S., Wallace, D., Chandler, J.M., Duncan, P.W., Rooney, E., Fox, M., and Guralnik, J.M. (2003). Physical performance measures in the clinical setting. *Journal of the American Geriatric Society, 51,* 314-322.

Tager, I.B., Swanson, A., and Satariano, W. (1998). Reliability of physical performance and self-reported functional measures in an older population. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 53,* M295-M300.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M., and de Vet H.C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60,* 34-42.

Terwee, C.B., Mokkink, L.B., Steultjens, M.P., and Dekker, J. (2006). Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: A systematic review of measurement properties. *Rheumatology (Oxford, England), 45,* 890-902.

Thapa, P.B., Gideon, P., Fought, R.L., Kormicki, M., and Ray, W.A. (1994). Comparison of clinical and biomechanical measures of balance and mobility in elderly nursing home residents. *Journal of the American Geriatric Society, 42,* 493-500.

Thomas, V.S and Hageman, P.A. (2002). A preliminary study on the reliability of physical performance measures in older day-care center clients with dementia. *International Psychogeriatrics, 14,* 17-23.

Thomas, V.S. and Hageman, P.A. (2003). Can neuromuscular strength and function in people with dementia be rehabilitated using resistance-exercise training? Results from a preliminary intervention study. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences, 58,* 746-751.

Received February 9, 2009
Accepted July 16, 2009

APPENDIX 1. Summary of atributes and criterias for the evaluation of walking speed tests

Atribute	*Definition*	*Quality criteria*[a]
1) Conceptual model	Extent to which a reasoned description is provided of the concept and population evaluated	(+) Detailed description provided of the foundations for using the test as a measure of functional limitations and/or full description of the test and population (?) Clear description lacking about the use of the test as a measure of functional limitations and/or incomplete description of the test or population (-) No mention of the conceptual foundations of the test and/or only one aspect of the walking test described (*e.g.,* distance covered) or population description missing
2) Burden	Extent to which a description is provided of the time, effort, requirements and demands of test administration for the interviewee and interviewer	(+) Detailed information provided about the administration time for the participant and examiner and/or demands of the task and special requirements of the participants and examiner (?) No clear information provided about the administration time for the participant and examiner and/or demands of the task and special requirements of the participants and examiner (0) No information found on burden
3) Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	(+) Mean and SD presented in representative samples of the general population and/or MIC are defined (?) Doubtful design or method and/or non-representative sample and/or MIC not defined (0) No information found on interpretability
4) Floor or ceiling effects	Number of participants who achieved the lowest or highest possible score	(+) $\leq$ 15% of the participants achieve the minimum and maximum possible score (?) Doubtful design or method (-) > 15% of the participants achieve the minimum and maximum possible score (0) No information found on floor or ceiling effects
5) Reliability		
5.1) Test-retest reliability	Extent to which similar results can be obtained through repeat measures in stable people	(+) ICC or Kappa $\geq$ .70; time interval of one or two weeks (?) Doubtful design or method (time interval not indicated or < 1 week, sample size < 50) (-) ICC or Kappa < .70 despite adequate design and method (0) No information found on reliability
5.2) Interrater reliability	Extent to which similar results can be obtained in stable people examined by two different observers	(+) ICC or Kappa $\geq$ .70 (?) Doubtful design or method (-) ICC or Kappa < .70 despite adequate design and method (0) No information found on interrater reliability
6) Validity		
6.1) Predictive validity	Extent to which the test is able to predict important future clinical results	(++) The test predicts numerous adverse results in at least two different population groups (+) The text predicts adverse results in just one population group

APPENDIX 1. Summary of atributes and criterias for the evaluation of walking speed tests. (Cont.)

Atribute	*Definition*	*Quality criteria*[a]
	results	(+) The text predicts adverse results in just one population group (?) Doubtful design or method (-) The test does not predict adverse results despite adequate design and method (0) No information found
6.2) Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	(+) Measurements of the same correlation construct $\geq$ .60 and/or measurements of different constructs and at least 75% of the hypotheses are confirmed (?) Doubtful design or method (*e.g.,* lack of hypothesis formulation, sample < 50) (-) Measurements of the same correlation construct < .60 and/or measurements of different constructs and less than 75% of the hypotheses are confirmed. (0) No information found on construct validity
7) Responsiveness	The ability of the instrument to detect changes	(+) Evidence of change in scores through ET measurement; longitudinal design with comparison between a stable group and a group that has changed (?) Doubtful design or method. Contradictory results (-) No changed detected in scores despite adequate design and method 0 No information found on responsiveness