

## REPRODUCIBILIDAD DE LA BATERÍA EPESE DE DESEMPEÑO FÍSICO EN ATENCIÓN PRIMARIA

Julio Cabrero García\*  
Abilio Reig-Ferrer\*\*\*  
Carmen Muñoz-Mendoza\*  
M<sup>a</sup> José Cabañero-Martínez\*  
Juan Diego Ramos-Pichardo\*  
Miguel Richart-Martínez\*  
Lorena González-Lloplis\*\*\*\*  
Ángela Sanjuán-Quiles\*  
Nieves Gómez-Moreno\*\*  
Soledad García-Beviá\*\*

Departamento de Enfermería. Universidad de Alicante\*

AVS. Centro de Salud Hospital Provincial-Plà. Alicante\*\*

Departamento de Psicología de la Salud. Universidad de Alicante\*\*\*

Centro de Saud de Anna. Alicante\*\*\*\*

### RESUMEN

*Objetivo. Estimar la reproducibilidad de tres medidas objetivas de desempeño físico de personas mayores en atención primaria.*

*Diseño. Estudio descriptivo y prospectivo con observación directa de la función física por parte de profesionales de la salud de acuerdo a un protocolo estandarizado.*

*Emplazamiento. Tres centros de atención primaria de las provincias de Alicante y Valencia.*

---

Nota: Este estudio está financiado por el Fondo de Investigación Sanitaria, nº expte. PI051538.  
Correspondencia: Julio Cabrero García. Departamento de Enfermería. Universidad de Alicante.  
Campus de San Vicente del Raspeig. E-03080 Alicante. E-mail: julio.cabrero@ua.es.

*Participantes.* Muestra de 66 personas de 70 y más años, evaluadas en dos ocasiones por el mismo profesional al objeto de replicar idénticas condiciones del estudio, en un intervalo temporal de dos semanas (mediana de 14 días).

*Mediciones principales.* Se evaluó el funcionamiento físico a través de tres pruebas objetivas de desempeño: el test de equilibrio, el de velocidad de la marcha, y la capacidad para levantarse y sentarse de una silla. Estas medidas provienen de los estudios EPESE (Established Populations for Epidemiologic Studies of the Elderly). Se ha calculado la fiabilidad test-retest mediante el coeficiente de correlación Intraclase.

*Resultados.* Los coeficientes de correlación intraclase (CCI) fueron de 0,55 para el test de equilibrio, de 0,69 para el test de levantarse de la silla, y de 0,79 para el de velocidad de la marcha. El valor para la puntuación total de la batería EPESE fue de 0,80.

*Conclusiones.* La reproducibilidad de estas medidas de desempeño es tan aceptable como las aportadas por la bibliografía de referencia. Estas pruebas de desempeño permiten evaluar con rigor cambios importantes en funcionamiento y salud que se producen con el tiempo.

**Palabras clave:** FIABILIDAD TEST-RETEST, EVALUACIÓN DEL DESEMPEÑO FÍSICO, BATERÍA EPESE, ATENCIÓN PRIMARIA.

## **SUMMARY**

*Objective.* To estimate the reproducibility of three direct measures of physical performance from the Short Physical Performance Battery, or EPESE Battery, in primary care.

*Design.* Prospective and descriptive study based on direct observation of physical function by health professionals using a standardised method.

*Setting.* Three primary care health centres in the provinces of Alicante and Valencia, Spain.

*Participants.* A sample of 66 subjects, 70 or more years, was selected. Subjects were tested in Primary Care contexts

twice (time interval of 2 weeks, median 14 days) by the same interviewer to replicate study conditions.

**Main measurements.** Performance-based functional status was evaluated by three performance tests: standing balance, gait speed, and chair stand. These tests were adapted from the EPESE studies (Established Populations for Epidemiologic Studies of the Elderly). Intra-class correlation coefficients (ICCs) were used to assess direct measures of performance.

**Results.** Intra-class correlations coefficients were 0.55 for standing balance, 0.69 for chair stands, and 0.79 for gait speed. An ICC of 0.80 for global EPESE battery was obtained. Pearson correlations values are of similar magnitude to the ICCs values.

**Conclusion.** Several commonly used measures of performance may be reliable as originally reported. Performance measures may be valuable for assessing major changes in health and functioning over time.

**Keywords:** TEST-RETEST RELIABILITY, PHYSICAL PERFORMANCE ASSESSMENTS, EPESE BATTERY, PRIMARY CARE.

## INTRODUCCIÓN

La discapacidad es un componente crítico que influye negativamente en la calidad de vida de las personas mayores (Cavazzini et al., 2004). En esta línea, es el mayor predictor, tras la edad, de resultados adversos, incluyendo la mortalidad, el deterioro en el funcionamiento, la incidencia de enfermedades agudas, caídas y lesiones, el uso de servicios sanitarios y la institucionalización (Guralnik, Fried y Salive, 1996). Mantener el funcionamiento y prevenir o reducir la discapacidad son de interés crítico para los profesionales de la salud que tratan con personas mayores. En Atención Primaria, los esfuerzos deben estar orientados, sobretodo, a detectar personas mayores en riesgo de desarrollar discapacidad o con pérdida de funcionalidad incipiente (Martín-Lesende, 2005).

El estatus funcional y la discapacidad se evalúan, por lo general, a través de autoinformes de actividades de la vida diaria y de autoinformes y tests objetivos de desempeño de limitaciones funcionales (Nagi, 1976; Verburgge y Jette, 1994; Guralnik y Ferrucci, 2003). Las medidas objetivas presentan ciertas ventajas sobre las medidas de autoinforme: mayor validez, mejor reproducibilidad, menor influencia del estatus cognitivo y menor influencia del lenguaje, el idioma y la cultura (Guralnik, Branch, Cummings y Curb., 1989). De esta forma, los tests de desempeño físico constituyen un elemento clave para la medición objetiva de las limitaciones funcionales. Estas medidas tienen el potencial de servir como indicadores de estado de salud, variables de resultado y predictores de sucesos futuros (Guralnik y Ferrucci, 2003). Dentro del grupo de medidas de desempeño físico, los tests basados en la medición de las extremidades inferiores (frente a las superiores) proporcionan la información más útil sobre el deterioro funcional preclínico en personas mayores. La Batería Abreviada de Desempeño Físico (*Short Physical Performance Battery*) o Batería EPESE representa quizás el mejor ejemplo de este tipo de medidas. Se compone de tres tests: equilibrio, marcha y levantarse y sentarse de una silla (Guralnik et al., 1994). Numerosos estudios han demostrado que esta batería es un buen predictor de importantes resultados adversos tales como institucionalización, utilización de los servicios de salud, caídas, mortalidad y comienzo de discapacidad (Guralnik et al., 1994; Guralnik, Ferrucci, Simonsick, Salive y Wallace, 1995; Ostir Markides, Black y Goodwin, 1998; Penninx et al., 2000). Otros estudios recientes han señalado también su valor predictivo en la identificación tanto del comienzo progresivo como catastrófico de la discapacidad en actividades de la vida diaria (Onder et al., 2005).

No existen estudios de fiabilidad en la literatura internacional sobre esta batería en el contexto de la atención primaria; en la literatura nacional en ningún contexto. Nuestro objetivo ha sido evaluar la estabilidad test-retest de la batería de desempeño físico en el ámbito de la Atención Primaria, en una muestra de personas mayores españolas de 70 y más años.

## MÉTODOS

### Personas

La muestra del estudio estuvo constituida inicialmente por 80 personas de 70 y más años pertenecientes a 4 centros de Atención primaria, dos urbanos, uno semi-urbano y uno rural. Las personas fueron reclutadas consecutivamente entre aquellas que asistieron a consulta de atención primaria en visita programada o a demanda. Se excluyeron a los pacientes con un puntuación de 4 o más errores en el test de Pfeiffer (Pfeiffer, 1975), a aquellos que fueron valorados por la enfermera o el médico como demasiado frágiles para realizar la batería de desempeño (con patologías cardíacas o respiratorias que pudiesen agravarse con el ejercicio), a los que tenían severos problemas de comunicación y a los dependientes en las actividades básicas de la vida diaria según el índice de Katz (Katz, Downs, Cash y Grotz, 1970). De las 80 personas que participaron en la evaluación inicial, 66 aceptaron volver para una nueva evaluación dentro de un plazo de dos semanas, cumplimentando una hoja de consentimiento informado.

### Variables

Se midió objetivamente el desempeño físico con la Batería Abreviada de Desempeño Físico (*Short Physical Performance Battery*, SPPB), desarrollada originalmente en el contexto de los estudios EPESE (Cornoni-Huntley, Brock, Ostfeld, Taylor y Wallace, 1986) para medir la función de las extremidades inferiores. La batería incluye tres pruebas: equilibrio, velocidad de la marcha y capacidad para levantarse y sentarse de una silla. En el test de equilibrio, el participante intenta mantener tres posiciones: pies juntos, semi-tándem y tándem durante 10 segundos. Las tres sub-pruebas siguen una secuencia jerárquica. En el test de velocidad de la marcha, el participante anda a su ritmo habitual una distancia de 4 metros. El test se realiza dos veces y se registra el tiempo más breve de los dos. Finalmente, en el test de capacidad de levantarse de la silla, el participante se levanta y se sienta en una silla 5 veces, de la forma más rápida posible, y se registra

el tiempo total empleado. La puntuación en cada uno de las tres pruebas se ordena en cinco niveles: entre 0 (peor rendimiento) y 4 (mejor rendimiento); se obtiene, además, una puntuación global para toda la batería mediante la suma de las puntuaciones de las pruebas, que oscila entre 0 y 12 puntos (Guralnik, et al., 1994).

Además, se recogió información sociodemográfica y se aplicó la subescala de función física (FF-10) del SF-36. El SF-36 es el instrumento genérico más utilizado en la bibliografía internacional para medir la calidad de vida relacionada con la salud (Garrat, Schmidt, Mackintosh y Fitzpatrick, 2002) y se dispone de valores de referencia de la versión española en población adulta de más de 60 años (López-García et al., 2003).

## **Procedimiento**

La batería se administró en los tres centros de salud participantes, en dos ocasiones, separadas entre 10 y 26 días (mediana de 14). Este intervalo temporal es el mismo que el empleado en otros estudios de reproducibilidad de esta misma batería (Wolinsky, Miller, Andresen, Malmstrom y Miller, 2005; Jette, Jette, Ng, Plotkin y Bach, 1999) y se adhiere a las recomendaciones generales para los estudios de reproducibilidad (Streiner y Norman, 2003, Scientific Advisory Committee of the Medical Outcomes Trust, 2002). Cuatro observadores, previamente entrenados, administraron la batería siguiendo el protocolo establecido por Guralnik et al. (2000), cada uno de ellos a los mismos participantes en el test y en el retest. El entrenamiento incluyó la administración supervisada de la batería a 20 personas mayores. Se evaluó posteriormente la fiabilidad entre observadores con 30 pacientes, con la misma definición de elegibles que la del estudio de reproducibilidad y de dos de los centros de salud participantes. Los cuatro observadores administraron las pruebas organizados por parejas, con un intervalo de 15 minutos entre las dos administraciones. Los observadores se alternaron en el orden de administración con cada nuevo paciente. Las observaciones fueron independientes. Los coeficientes de correlación intraclase globales fueron: 0,77; 0,84; 0,83 y 0,88, para los tests de equilibrio, de la marcha, de sentarse y levantarse de la silla y de la batería EPESE,

respectivamente. El equipamiento para administrar la batería consistió en un cronómetro, una silla de medida estándar, y un espacio disponible de cuatro metros de longitud.

Las variables sociodemográficas y la subescala de función física del SF-36 se evaluaron mediante entrevista personal.

### **Análisis estadístico**

Las características de la muestra se han analizado con estadísticos descriptivos y distribuciones de frecuencia. La fiabilidad test-retest se ha calculado mediante el coeficiente de correlación intraclass (CCI), estadístico recomendado para este tipo de estudios (Amstrong, White y Saracci, 1992, Scientific Advisory Committee of the Medical Outcomes Trust, 2002). Para su determinación se empleó un ANOVA bidimensional de efectos aleatorios. El CCI así obtenido es esencialmente equivalente al coeficiente Kappa ponderado basado en pesos cuadráticos para los desacuerdos, siempre que el número de observaciones sea suficientemente grande, como es el presente caso (Fleiss y Cohen, 1973). Alternativamente se calculó el CCI a partir de un ANOVA bidimensional de efectos mixtos, considerando los dos momentos de medida como efectos fijos (otro modelo de ICC frecuentemente empleado en estudios de estabilidad). Los resultados fueron casi idénticos (diferencias menores a cinco centésimas en todos los casos) por lo que no se presentan. Dado que el CCI calculado en este estudio es una medida de acuerdo absoluto, equivalente a un coeficiente Kappa, se han seguido las convenciones de Landis y Koch (1977) para interpretar su magnitud. Los análisis se realizaron con el módulo de escalas (fiabilidad) del programa estadístico SPSS versión 11.5.

## **RESULTADOS**

La edad media de esta muestra de pacientes de atención primaria fue de 76 años (con un arco de variación de entre 70 a 91 años) y con una presencia algo mayor de mujeres frente a varones (51,5 %). En la tabla 1 mostramos los datos descriptivos de la muestra así como los porcentajes relativos al nivel de estudios, convivencia, estado civil,

situación económica percibida y rendimiento medio en función física, según las puntuaciones del FF-10. En relación a los baremos de la población española, la muestra estudiada tiene un mayor nivel de función física.

**Tabla 1.** Descripción de la muestra del estudio (N=66)

Características	%	Media $\pm$ DE	Recorrido	Mediana
Edad		76,61 $\pm$ 4,81	70-91	76
<b>Género</b>				
Varones	48,5			
Mujeres	51,5			
<b>Nivel de estudios</b>				
No leer	12,3			
Leer/Estudios primarios	75,4			
Bachiller elemental	9,2			
Bachiller superior	3,1			
<b>Convivencia</b>				
Con pareja dependiente	1,5			
Con pareja	43,9			
Con pareja y familia	13,6			
Con familia	12,1			
Solo con hijo/a cerca	22,7			
Solo	6,1			
<b>Estado civil</b>				
Soltero/a	6,1			
Casado/a	59,1			
Separado/a	1,5			
Viudo/a	33,3			
<b>Situación económica</b>				
Vivir desahogadamente	53			
Atender necesidad básicas	37,9			
No necesidad básicas	9,1			
<b>Función física</b>				
FF-10 (SF-36)		70,9 $\pm$ 23	10-100	75



En la tabla 2 se presentan los coeficientes de correlación intraclase para cada uno de los tests de la batería así como para el global. Asimismo, se señalan las medias antes y después de la medición repetida a lo largo de dos semanas (mediana de 14 días, con una variación entre 10 y 26 días) y los coeficientes de correlación de Pearson.

**Tabla 2. Reproducibilidad de las pruebas de desempeño físico de la batería EPESE**

Test	n	CCI	IC 95%	Correlación	Media ± DE	Media ± DE
				Pearson	Tiempo 1	Tiempo 2
Velocidad andar	66	0,79	0,67-0,86	0,79	3,06± 0,84	3,06± 0,76
Levantarse silla	66	0,69	0,54-0,80	0,69	2,53± 1,10	2,47± 1,08
Equilibrio	62	0,55	0,35- 0,70	0,54	3,32± 1,11	3,26± 1,10
Batería EPESE	62	0,80	0,69-0,68	0,80	9,02± 2,46	8,92± 2,39

CCI= Coeficiente de Correlación Intraclase (ANOVA bidimensional de efectos aleatorios, definición de acuerdo absoluto); IC= Intervalo de confianza.

Los coeficientes de correlación intraclase de los tests constitutivos de la batería EPESE oscilan entre 0,55 y 0,79. Los coeficientes son más altos en los dos componentes de la batería que miden limitación funcional: test de velocidad al andar (CCI= 0,79) y test de levantarse de la silla (CCI= 0,69), siendo su magnitud en ambos casos notable. La magnitud del valor del CCI para el componente de deficiencia de la batería, el test de equilibrio, es sólo moderada (CCI= 0,55). El valor para la puntuación total de la batería EPESE alcanza el estándar de un kappa en la frontera de casi perfecto (CCI= 0,80).

Los coeficientes de correlación de Pearson muestran valores casi idénticos a los CCI tanto para cada uno de los componentes de la batería como para la puntuación global en la misma, lo que sucede porque las medias y las varianzas de las dos series son iguales (Fleiss y Cohen, 1973). Esta situación excluye la presencia de sesgo entre los dos momentos de medida, específicamente la vinculada con los efectos de memoria y aprendizaje. Además, un análisis diferencial

por género (no mostrado) reveló que no existían diferencias significativas en los niveles de actuación para cada uno de los tests, así como para el global.

En un intento por aclarar la fuente del moderado nivel de fiabilidad del test de equilibrio, calculamos el coeficiente Kappa no ponderado para cada categoría comparada con el resto de categorías combinadas en una sola (Amstrong et al., 1992). Los valores Kappa fueron 0.32, 0.24, 0.02 y 0.47, para las categorías 0-1, 2, 3 y 4, respectivamente (las categorías 0 y 1 se reagruparon en una sola). Calculamos además los porcentajes de acuerdo positivo y negativo (Cichetti y Feinstein, 1990). En estos análisis el acuerdo positivo se da cuando los observadores coinciden en valorar una ejecución en una categoría dada y el acuerdo negativo cuando los observadores «coinciden» en valorar la ejecución en cualquiera otra categoría indistintamente. Los porcentajes de acuerdo positivo fueron: 0.54, 0.31, 0.13 y 0.81 para las categorías 0-1, 2, 3 y 4 respectivamente. A su vez, los porcentajes de acuerdo negativo fueron: 0.92, 0.93, 0.88 y 0.65. Estos resultados sugieren que la fuente de fiabilidad del test de equilibrio estriba en la discriminación entre los que obtienen la máxima puntuación y el resto, y consiguientemente la discriminación entre los restantes niveles inferiores no es fiable.

## DISCUSIÓN

Aunque se ha demostrado que las medidas de desempeño físico son un buen predictor de importantes resultados adversos para la salud y la calidad de vida, no disponemos de estudios que establezcan la reproducibilidad de estas medidas en pacientes de atención primaria. Con el estudio que hemos llevado a cabo, se analiza por primera vez en una muestra española esta cuestión y se aportan datos que sostienen un más que razonable uso, desde el punto de vista psicométrico, de esta batería de desempeño físico. Los coeficientes de correlación intraclase obtenidos presentan unos valores considerados como moderados para el test de equilibrio y notables para el resto, según los estándares de Landis y Koch (1977). Un estándar más reciente y exigente (Scientific Advisory Committee of the Medical Outcomes Trust, 2002) sitúa en 0.70 la fiabilidad mínima necesaria para la toma de decisiones grupales; por tanto,

Tabla 3. Estudios de fiabilidad test-retest de la Batería EPESE

ESTUDIO	CONTEXTO	N	TIEMPO	ESTADÍSTICO
Wolinsky et al., 2005	Comunidad	80	5-45 días	CCI
Studensky et al., 2003	A. Primaria	No indica	2 semanas	CCI
Ostia et al., 2002	Comunidad	91	3 mediciones Intervalo: 1 semana	CCI
Guralnik et al., 1999	Comunidad	102	3 mediciones Intervalo: 1 semana	CCI
Jette et al., 1999	Varios	105	2 semanas	CCI
Tager et al., 1998	Comunidad	199	(0-132 días) 48 horas	CCI
Hoeymans et al., 1997	Comunidad	105	2 semanas	CC Pearson Kappa
Seeman et al., 1994	Comunidad	No indica	2 semanas	CC Pearson C.C Pearson

  

JUNTOS	BALANCE		TOTAL	MARCHA	SILLA	BATERÍA GLOBAL
	SEMI	TANDE M				
No indica	0.52	0.27)	No indica	0.56	0.72	No indica
No indica	No indica	No indica	No indica	No indica	No indica	0.72
No indica	No indica	0.71; 0.82; 0.70	0.86; 0.80; 0.89	0.86; 0.80; 0.89	0.76; 0.86; 0.90	0.92; 0.88; 0.91
No indica	No indica	No indica	0.88; 0.89; 0.88	0.88; 0.89; 0.88	0.76; 0.84; 0.89	No indica
No indica	No indica	No indica	0.79	0.79	0.67	No indica
No indica	No indica	No indica	0.78	0.78	0.82	No indica
No indica	No indica	No indica	0.93	0.93	0.81	No indica
No indica	No indica	K= 0.29	CCP=0.90	CCP=0.82	CCP=0.82	K=0.52
No indica	No indica	0.61	0.80	0.73	0.73	No indica

(rápido)

CCI es coeficiente de correlación Intraclase; CCP es coeficiente de correlación de Pearson; K es coeficiente Kappa

según este criterio, el test de equilibrio presenta un valor inadecuado para tal fin.

La literatura sobre la reproducibilidad de las medidas de desempeño de limitaciones funcionales revela que los tests de equilibrio son sólo de una fiabilidad moderada (Wolinsky et al., 2005; Seeman et al., 1994) o menos que moderada (Hoeymans, Wouters, Feskens, van der Bos, Kromhout, 1997), de menor fiabilidad que los otros dos tests (Ostir, Volpato, Fried, Chaves y Guralnik, 2002) y los menos estudiados o informados. (Véase la tabla 3 en la que se inventaría los estudios de fiabilidad con los tests de la batería EPESE). En nuestro estudio, su fiabilidad fue moderada y de menor valor que la de los otros dos tests. Sólo Ostir et al. (2002) obtuvieron un valor claramente superior ( $>0.70$ ), pero el intervalo entre las medidas fue de una semana, frente a las dos semanas de nuestro estudio. Además, otra diferencia importante fue que la muestra de éstos se compuso de mujeres con baja capacidad funcional, frente a la nuestra que se compuso de varones y mujeres con una capacidad funcional por encima de la media, según los baremos del FF-10/SF-36 (López-García et al., 2003). Los tests de equilibrio se diseñaron originalmente para discriminar entre pobre y aceptable grado de equilibrio (Curb et al., 2006), por lo que pueden tener un efecto techo en las poblaciones con mayor capacidad funcional. Nuestro análisis por categorías del desempeño de este test revela no sólo su efecto techo sino también que su fiabilidad se limita a la discriminación entre dos niveles: pobre equilibrio frente aceptable o superior. Con respecto a los otros dos tests, nuestros resultados son muy consistentes con los de otros estudios que han empleado un intervalo de tiempo similar al nuestro (Wolinsky et al., 2005; Jette et al., 1999 y Seeman et al., 1994), aunque algo inferiores a los de aquellos que han empleado un intervalo menor (Ostir et al., 2002; Guralnik et al., 1999; Tager, Swanson y Satariano, 1998). La conclusión es firme: estos dos tests presentan un grado de reproducibilidad sustancial. Por último, el valor de reproducibilidad para toda la batería está en el límite inferior de la concordancia casi perfecta. Nuestro valor es sin embargo inferior al de Ostir et al. (2002), pero recuérdese que su intervalo temporal fue sólo una semana. Sólo dos estudios más han informado de valores de reproducibilidad para toda la batería EPESE, y en ambos el valor fue inferior (Wolinsky et al., 2005; Studenski et al., 2003).

A pesar de que la batería EPESE reúne la mayor parte de la evidencia longitudinal sobre la relación entre movilidad y desarrollo de discapacidad y otros resultados adversos, muy pocos estudios han evaluado la fiabilidad de sus tres componentes y de la batería completa, y sólo el de Ostir et al (2002) y el nuestro han seguido el protocolo que estandariza tanto su administración como su puntaje. El resto de estudios seleccionan preferentemente dos de los tres tests y utilizan un escala de puntuación continua, incluso Guralnik et al (1999), a pesar de que la estandarización de la batería implica que los puntajes directos se transformen en escalas categóricas ordinales. Además, estos puntajes categóricos tienen el valor añadido de su significado clínico; por ejemplo, las personas no discapacitadas que puntúan entre 4 y 9 en la batería tienen un elevado riesgo de desarrollar discapacidad (Guralnik et al., 2000).

En este estudio se establece por vez primera la reproducibilidad de la batería ESESE en pacientes de atención primaria, pero sólo en personas mayores de 70 años. Los criterios de selección excluyeron a las personas con deterioro cognitivo y dependientes en actividades básicas de la vida diaria, siendo la muestra resultante de un nivel superior a la media en función física, lo que es relevante de cara a la generalización de los resultados. Dos de los tres tests que componen la batería, y la batería completa alcanzan los estándares de calidad recomendados; uno, el test de equilibrio, sólo un nivel moderado.

## REFERENCIAS BIBLIOGRÁFICAS

- Armstrong, B.K., White, E., y Saracci, R. (1992). *Principles of exposure measurement in epidemiology*. Oxford: Oxford University Press.
- Cavazzini, C., Conti, M., Bandinelli, S., Gangemi, S., Gallinella, M., Lauretani, F., Lucci, G., Windham, B.G., Guralnik, J.M., y Ferrucci, L. (2004). Screening for poor performance of lower extremity in primary care: the Camucia Project. *Aging Clinical and Experimental Research*, 16, 331-336.
- Cicchetti, D.V., y Feinstein, A.R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43, 551-558.
- Cornoni-Huntley, J., Brock, D., Ostfeld, A., Taylor, J., y Wallace, R.

- (1986). *Established Populations for Epidemiologic Studies of the Elderly: Resource Data Book*. Bethesda, Md: National Institute on Aging.
- Curb, J.D., Ceria-Ulep, C. D., Rodriguez, B.L., Grove, J., Guralnik, J., Willcox, B.J., Donlon, T.A., Masaki, K.H., y Chen, R. (2006). Performance-based measures of physical function for high-function populations. *Journal of the American Geriatric Society*, 54, 737-742.
- Fleiss, J.L., y Cohen, J. (1973). The equivalence of weighted kappa and the intraclass coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.
- Garratt, A., Schmidt, L., Mackintosh, A., y Fitzpatrick, R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal*, 324, 1417.
- Guralnik, J.M., Branch, L.G., Cummings, S.R., y Curb, J.D. (1989). Physical performance measures in aging research. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences*, 44, M141-M146.
- Guralnik, J.M., Simonsick, E.M., Ferrucci, L., Glynn, R.J., Berkman, L. F., Blazer, D. G., Scherr, P.A., y 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. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences*, 49, M85-M94.
- Guralnik, J.M., Ferrucci, L., Simonsick, E.M., Salive, M.E., y Wallace, R.B. (1995). Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *New England Journal of Medicine*, 332, 556-561.
- Guralnik, J.M., Fried, L.P., y Salive, M.E. (1996). Disability as a public health outcome in the aging population. *Annual Review of Public Health*, 17, 25-46.
- Guralnik, J.M., Ferrucci, L., Penninx, B. W., Kasper, J.D., Leveille, S. G., Bandeen-Roche, K., y 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., y 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., y Ferrucci, L. (2003). Assessing the building blocks of function: utilizing measures of functional limitation. *American Journal of Preventive Medicine*, 25, 112-121.
- Hoeymans, N., Wouters, E.R., Feskens, E.J., van den Bos, G.A., y 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.
- Jette, A.M., Jette, D.U., Ng, J., Plotkin, D.J., y 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.
- Katz, S., Downs, T.D., Cash, H.R., y Grotz, R. C. (1970). Progress in development of the index of ADL. *The Gerontologist*, 10, 20-30.
- Landis, J.R., y Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lopez-Garcia, E., Banegas, J.R., Graciani, A., Gutiérrez, J.L., Alonso, J., y Rodriguez-Artalejo, F. (2003). Valores de referencia de la versión española del Cuestionario de Salud SF-36 en población adulta de más de 60 años. *Medicina Clínica*, 120, 568-573.
- Martin-Lesende, I. (2005). Detección de ancianos de riesgo en atención primaria: recomendación. *Atención Primaria*, 36, 273-277.
- Nagi, S.Z. (1976). An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly. Health and Society*, 54, 439-467.
- Onder, G., Penninx, B.W., Ferrucci, L., Fried, L.P., Guralnik, J.M., y 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 Gerontology: Serie A: Biological Sciences and Medical Sciences*, 60, 74-79.

- Ostir, G.V., Markides, K.S., Black, S.A., y Goodwin, J.S. (1998). Lower body functioning as a predictor of subsequent disability among older Mexican Americans. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences*, 53, M491-M495.
- Ostir, G.V., Volpato, S., Fried, L.P., Chaves, P., y 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.
- Penninx, B.W., Ferrucci, L., Leveille, S.G., Rantanen, T., Pahor, M., y Guralnik, J. M. (2000). Lower extremity performance in nondisabled older persons as a predictor of subsequent hospitalization. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences*, 55, M691-M697.
- Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatric Society*, 23, 433-441.
- 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., y Rowe, J.W. (1994). Predicting changes in physical performance in a high-functioning elderly cohort: MacArthur studies of successful aging. *Journals of Gerontology: Serie A: Biological Sciences and Medical Sciences*, 49, M97-108.
- Streiner, D.L., y Norman, G.R. (2003). *Health measurement scales. A practical guide to their development and use.* (3<sup>o</sup> ed.) Oxford: Oxford University Press.
- Studenski, S., Perera, S., Wallace, D., Chandler, J. M., Duncan, P. W., Rooney, E., Fox, M., y 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., y 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.



Verbrugge, L.M., y Jette, A.M. (1994). The disablement process. *Social Science & Medicine*, 38, 1-14.

Wolinsky, F.D., Miller, D.K., Andresen, E.M., Malmstrom, T.K., y Miller, J.P. (2005). Reproducibility of physical performance and physiologic assessments. *Journal of Aging and Health*, 17, 111-124.