Title

Performance of the Medicare Consumer Assessment of Health Care Providers and Systems (CAHPS) Physical Functioning Items

Journal

Medical Care, 54(2)

ISSN

0025-7079

Authors

Hays, Ron D

Mallett, Joshua S

Gaillot, Sarah

et al.

Publication Date

2016-02-01

DOI

10.1097/mlr.0000000000000475

Peer reviewed

Performance of the Medicare Consumer Assessment of Health Care Providers and Systems (CAHPS) Physical Functioning Items

Ron D. Hays, PhD, Joshua S. Mallett, MS,† Sarah Gaillot, PhD,‡ and Marc N. Elliott, PhD†*

Background: Physical functioning is an important health domain for adults.

Objective: Evaluate physical functioning items in Medicare beneficiaries.

Research Design: Survey data from the 2010 Consumer Assessment of Healthcare Providers and Systems Medicare survey.

Subjects: The 366,701 respondents were 58% female; 38% were 75 or older; 57% had high school education or less.

Measures: Walking, getting in or out of chairs, bathing, dressing, toileting, and eating assessed with 3 response choices: *unable to do, have difficulty, do not have difficulty*.

Results: Pearson correlations among the 6 items ranged from 0.515 to 0.835 (coefficient $\alpha = 0.92$). A single factor categorical factor analytic model fit the data well (comparative fit index = 0.998; root mean square error of approximation = 0.083). The item with the highest percentage of respondents reporting *no difficulty* was eating, followed by toileting, dressing, bathing, getting in and out of a chair, and walking. Threshold parameters from an item response theory–graded response model ranged from −1.983 (between *unable to do* and *have difficulty* eating) to −0.551 (between *have difficulty* and *no difficulty* walking). Item discrimination parameters ranged from 4.632 (walking) to 8.228 (dressing). IRT-scored physical functioning scores correlated with self-rated general health ($r = 0.389$, n = 344,843, $P < 0.0001$) mental health ($r = 0.296$, n = 351,254, $P < 0.0001$) and number of chronic conditions ($r = -0.229$, n = 284,507, $P < 0.0001$).

Conclusions: The physical functioning items target relatively easy activities, providing information for a minority of people in the sample with the lowest levels of physical functioning. Items representing higher levels of physical functioning are needed for the majority of the Medicare beneficiaries.

Key Words: physical functioning, Medicare, CAHPS, self-reported health

(*Med Care* 2015;00: 000–000)

Physical functioning is the ability to conduct a variety of activities ranging from self-care to more challenging and vigorous activities that require increasing degrees of mobility, strength, or endurance. It is one of the strongest predictors of hospitalizations, institutionalization, and mortality.[1–3] Four major components of physical functioning have been proposed: (1) instrumental activities of daily living; (2) lower extremity (mobility); (3) upper extremity (dexterity); and (4) central (neck and back) activities.[4] Empirical analyses provide consistent support for a single underlying dimension of physical functioning.[5–7]

A large number of self-reported measures are available, differing in the number and type of activities represented. For example, the Katz Index of Independence in Activities of Daily Living[8] assesses 6 activities (bathing, dressing, toileting, transferring, continence, feeding) that assess mobility and dexterity; whereas the Patient-reported Outcomes Measurement Information System (PROMIS) physical functioning item bank of 124 items captures instrumental activities of daily living, mobility, dexterity, and central activities.[9] The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare managed care surveys included 6 physical functioning items (bathing, dressing, eating, getting in or out of chairs, walking, using the toilet) from 2007 to 2010, and on the fee-for-service survey from 2007 to 2013. The common item stem was "Because of a health or physical problem, are you unable to do or have any difficulty doing the following activities?" Three response options were provided: (1) I am unable to do this activity; (2) Yes, I have difficulty; (3) No, I do not have difficulty. These items are similar, but not identical, to the Katz items.

The physical functioning items were included to assess the level of physical functioning in Medicare beneficiaries overall and between health plans. The information yield of a physical functioning measure depends on the extent to which the activities it includes match the level of physical functioning of the target population. If the activities assessed are too difficult or too easy relative to the ability of a respondent,

From the *UCLA Department of Medicine, Los Angeles; †RAND Corporation, Santa Monica, CA; and ‡Centers for Medicare & Medicaid Services, Baltimore, MD.

Supported by CMS contract HHSM-500-2005-000281 to RAND. R.D.H. was also supported in part by grants from AHRQ (2U18 HS016980), NIA (P30AG021684), NIMHD (2P20MD000182), and NCI (1U2-CCA186878-01).

The authors declare no conflict of interest.

Reprints: Ron D. Hays, PhD, UCLA Department of Medicine, 911 Broxton Avenue, Los Angeles, CA 90024. E-mail: drhays@ucla.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

ISSN: 0025-7079/15/000-000

then the measure does not provide information about the level of physical functioning.

This paper evaluates the physical functioning items included in an annual survey of patient experiences with Medicare in the United States. We estimate on the reliability and validity of the items and an indication of how well they provide information on the physical functioning of Medicare beneficiaries.

MATERIALS AND METHODS

We analyzed data from the 2010 Medicare managed care and fee-for-service data collected from February 18 to June 15, 2010. The analytic sample included 366,701 Medicare managed care and fee-for-service beneficiaries in 2010: 58% female; 14% was 18–64, 48% was 65–74, 29% was 75–84, and 9% was of 85 and older group; 57% high school education or less. The average number of the 6 chronic conditions (angina, cancer, congestive heart failure, diabetes, heart attack, stroke) reported was 0.89.

Analysis Plan

We first treated the 6 items as having an equal interval scale, scoring the items as *unable to do* = 0, *able to do with difficulty* = 50, and *have no difficulty* = 100. We computed item frequencies and estimated internal consistency reliability (Cronbach α)[10] for the 6-item scale. We averaged the 6 items to create a 0–100 physical functioning scale and estimated the mean, SD, range, % at min, % at max, skewness, and kurtosis.

We then estimated a categorical confirmatory factor analytic model to assess whether the 6 items were unidimensional. Next, we fit a graded response model[11] to estimate between category threshold parameters (1 less than the number of response categories) and a discrimination parameter for each item. The model was also used to produce category response curves, a person-item map, and the physical functioning scale information curve. Finally, we estimated correlations of the physical functioning scale with the number of chronic conditions, self-rated general health (*poor*, *fair*, *good*, *very good*, *excellent*) and self-rated mental health (*poor*, *fair*, *good*, *very good*, *excellent*). We hypothesized that the physical functioning scale would be negatively related to the number of conditions[12] and positively associated with self-ratings of health.[13]

The majority of the analyses were conducted using SAS 9.4 (TS1M2). The confirmatory factor analytic model was estimated using Mplus Version 7.[14] A person-item map was created by manually combining output produced by SAS 9.4 PROC SGPLOT (TS1M2).[15]

RESULTS

Classic Test Theory

As seen in Table 1, the hardest item to report *no difficulty* performing was walking (69% reporting no difficulty), followed by getting in and out of chairs (78%), bathing (85%), dressing (87%), using the toilet (91%), and eating (94%). The percentage of persons who were unable to do an activity was similar across items (3%–4%). Variation in responses to items is evidenced in the having *some difficulty* and *no difficulty* response choices.

TABLE 1. Percentage of Medicare Beneficiaries (n = 366,701) Selecting Each Response Option for the 6 Physical Functioning Items

Item	Unable to Do	Have Difficulty	No Difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

The product-moment correlations among the 6 items ranged from 0.515 to 0.835 with listwise deletion of cases and 0.514 to 0.838 with pairwise deletion of cases (Appendix, Supplemental Digital Content 1, http://links.lww.com/MLR/B82). Internal consistency reliability for the 6-item physical functioning scale was 0.92 (0.93) with pairwise (listwise) deletion of cases. Item-scale correlations (corrected for overlap) ranged from 0.71 (walking) to 0.85 (dressing) for pairwise deletion of cases and 0.71 (walking) to 0.86 (dressing) for listwise deletion of cases. The mean physical functioning scale score (scored as an average of responses to the 6 items and transformed linearly to a 0–100 possible range) was 89 (SD = 21) with a skewness of −2.69 and kurtosis of 7.38. Two percent of the 372,743 respondents scored at the floor and 65% were at the ceiling.

Item Response Theory (IRT)

We evaluated the IRT assumption of unidimensionality by fitting a single-factor categorical confirmatory factor analytic model and it fit the data well ($\chi^2 = 22{,}820.511$; n = 366,701; *df* = 9; comparative fit index = 0.998, root mean square error of approximation = 0.083). Local independence was supported by the small residual correlations (magnitude of residual correlations were 0.04 or smaller). Factor loadings were statistically significant and ranged from 0.930 to 0.977 (Table 2). Item discrimination parameters (Table 2) ranged from 4.632 (walking) to 8.228 (dressing); the range of threshold parameters was −1.983 (between *unable to do* and *have difficulty* eating) to −0.551 (between *have difficulty* and *no difficulty* walking).

Category response curves for the 6 physical functioning items are given in Figure 1. The curves show the probability of picking each response choice on the *y*-axis as a function of underlying physical functioning on the *x*-axis (logit). The 3 response categories for the 6 items are appropriately monotonically ordered and working as desired because each category is most likely to be selected for some level of underlying physical functioning. For all items, the *unable to do* response choice has the greatest probability of being selected for persons with an estimated physical functioning score ranging from about −2 and below on the logit scale (low level of physical functioning). The *no difficulty* response choice has the greatest probability of being selected for those with estimated physical functioning scores of about ≥ -1.

TABLE 2. Categorical Confirmatory Factor Analysis Loadings and Graded Response Model Item Parameters

	Walking	Chairs	Bathing	Dressing	Toileting	Eating
Loading (SE)	0.930 (0.001)	0.950 (0.000)	0.961 (0.000)	0.977 (0.000)	0.970 (0.000)	0.943 (0.001)
t statistic	1730.273	2249.092	2515.998	3236.282	2279.103	1302.112
Item thresholds						
Unable to do to *have difficulty*	−1.861	−1.914	−1.719	−1.785	−1.872	−1.983
Have difficulty to *no difficulty*	−0.551	−0.806	−1.025	−1.101	−1.268	−1.527
Item discrimination	4.632	5.652	6.341	8.228	7.232	4.870

Figure 2 provides a person-item map that displays the range of the items in the lower panel, and the person score distribution in the upper panel. The endpoints labeled 1 and 2 on the lines in the lower panel represent the 2 thresholds for each item. Note that the items are located to the left (easy) of the density of the distribution of person scores. That is, the items are targeted at the minority of people in the sample with low levels of physical functioning. Figure 3 provides the physical functioning scale (test) information curve. On the *z* score metric shown in Figure 3, reliability is equal to (information − 1)/information. Hence, information of 30, 20, and 10 represent reliabilities of 0.97, 0.95, and 0.90, respectively. Consistent with the person-item map, the scale information peaks below the average physical functioning score (logit of 0).

The IRT-scored physical functioning scale correlated significantly with a count of the number of chronic conditions ($r = -0.229$, n = 284,507, $P < 0.0001$), self-rated general health ($r = 0.389$, n = 344,843, $P < 0.0001$), and self-rated mental health ($r = 0.296$, n = 351,254, $P < 0.0001$). The physical functioning scale (scored as an average of responses to the 6 items) also correlated significantly with a count of the number of chronic conditions ($r = -0.164$, n = 284,507, $P < 0.0001$), self-rated general health ($r = 0.290$, n = 344,843, $P < 0.0001$), and self-rated mental health ($r = 0.233$, n = 351,254, $P < 0.0001$).

DISCUSSION

The 6 physical functioning items examined here had a high level of internal consistency reliability (exceeding the 0.90 minimum for use of measures at the individual level)[16] and provided adequate fit to a 1-factor model. The 3 response categories performed well in representing the underlying physical functioning concept. In addition, associations of the 6-item physical functioning scale score with the number of chronic conditions and self-rated health were consistent with hypotheses and were larger than the corresponding correlations for the simple-summated physical functioning score. The correlation of number of chronic conditions with the IRT-estimated physical functioning score was −0.229, generally consistent with a prior study that reported a 0.40 SD lower score on physical functioning for those reporting 1 condition versus no chronic conditions.[12] The correlations of the physical functioning scale with self-rated general health ($r = 0.389$) and mental health ($r = 0.296$) are

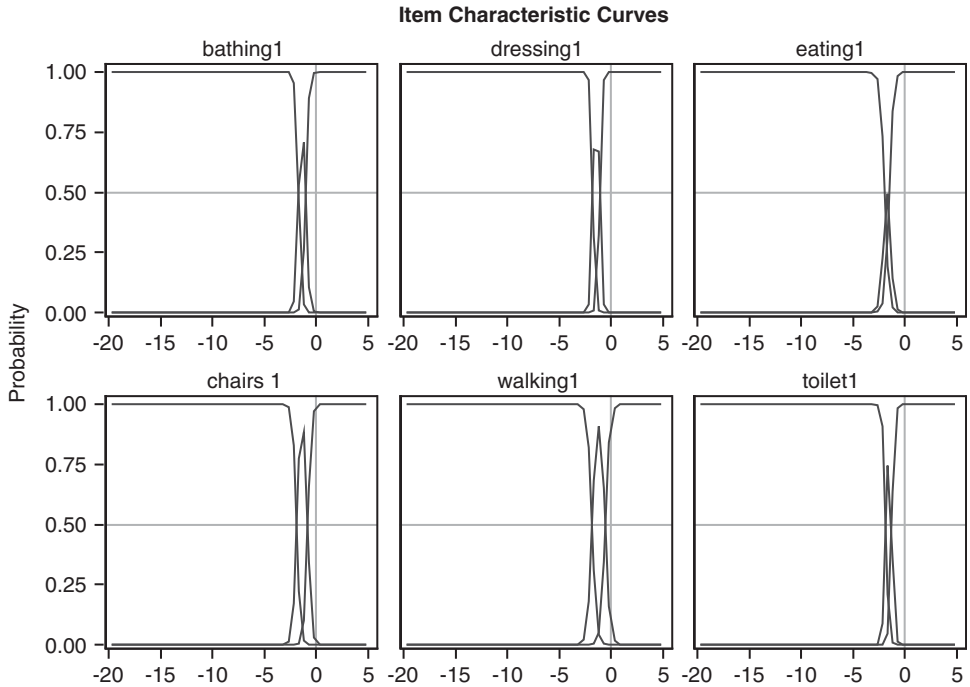

FIGURE 1. Category response curves.

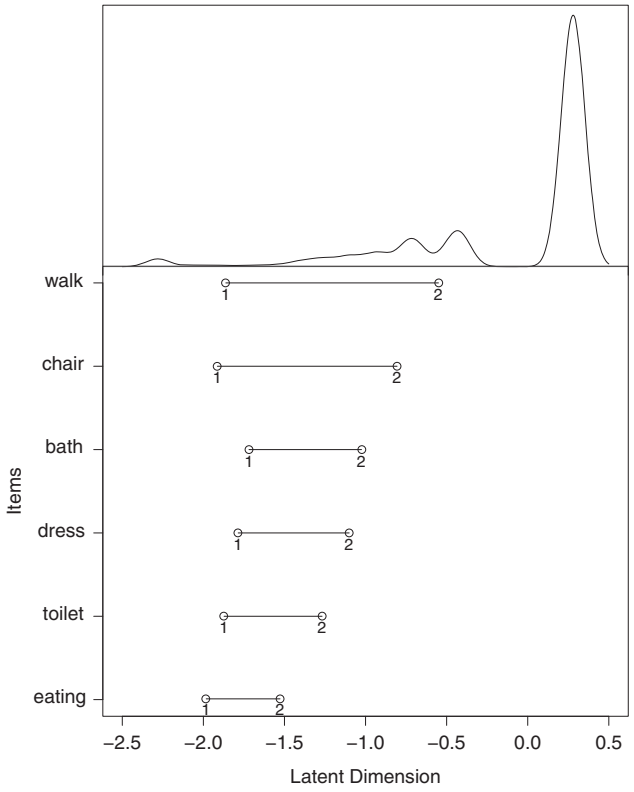

FIGURE 2. Person-item map.

similar to those reported for parallel measures in the PROMIS project.[13]

The physical functioning items target relatively easy activities with the majority of the sample reporting no difficulty walking, transferring from a chair, bathing, dressing, toileting, and eating. Sixty-five percent of the sample had the highest possible physical functioning score (no difficulty in any of the 6 activities). Hence, these items provide useful information for a minority of people in the sample with the lowest levels of physical functioning. The scale provides limited information for those scoring at or above the average level of physical functioning in the Medicare population. The 6-item physical functioning scale can be used to document the levels of physical functioning in less physically healthy Medicare beneficiaries. Items representing higher levels of physical functioning would improve measurement for the majority of the Medicare beneficiary sample.

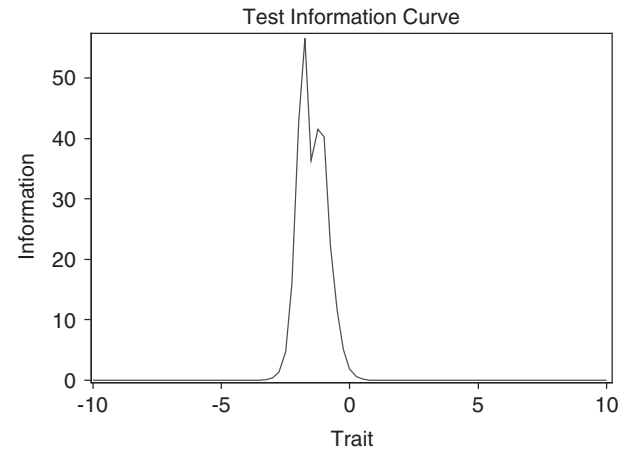

FIGURE 3. Physical functioning scale information curve.

The 6 physical functioning items in the CAHPS Medicare survey have parallel items in the PROMIS 20-item short form measure: walking (Does your health now limit you in walking more than a mile?), chairs (Are you able to transfer from a bed to a chair and back?), bathing (Are you able to wash and dry your body?), dressing (Are you able to dress yourself, including tying shoelaces and doing buttons?), toileting (Are you able to get on and off the toilet?), and eating (Are you able to hold a plate full of food?).[17] One possibility to improve measurement of physical functioning in CAHPS Medicare surveys is computer-adaptive administration of an item bank. For example, items in the PROMIS physical functioning item bank can be administered iteratively based on responses to previously administered items. This is the most efficient approach to obtaining maximal information about each respondent. Reliabilities of 0.90 or above can typically be obtained after administering about 5 items.[18] Another possibility is to add items that assess more difficult physical functioning activities. For example, the 4-item physical functioning scale in the PROMIS-29 profile measure includes an item about ability to run errands and shop.[19] The PROMIS 20-item short form includes more difficult physical activities such as doing chores like vacuuming or yard work, running a short distance, limitations in vigorous activities, lifting or carrying groceries, and engaging in 2 hours of physical labor.[17] Including more difficult physical functioning items in future CAHPS Medicare surveys would provide better information for the majority of the sample. To ensure broader content coverage for Medicare beneficiaries, some items could be selected to represent mobility and others to represent upper-extremity activities.[13]

In summary, this paper provides evidence supporting the psychometric properties of the 6 physical functioning items included in the CAHPS Medicare surveys but also indicates that the scale is only informative for those with limited physical functioning. If the only objective is to identify those with very low levels of functioning, then the set of 6 physical functioning items examined is satisfactory. If there is a desire for more precise information about functioning for the majority of the sample, then the item set would need to be bolstered with questions assessing more advanced physical functioning activities.

REFERENCES

1. Bień B, Bień-Barkowska K, Wojskowicz A, et al. Prognostic factors of long-term survival in geriatric inpatients. Should we change the recommendations for the oldest people? *J Nutr Health Aging*. 2015;19:481–488.
2. Han PKJ, Lee M, Reeve BB, et al. Development of a prognostic model for six-month mortality in older adults with declining health. *J Pain Symptom Manage*. 2012;43:527–539.

3. Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. *J Gerontol*. 1992;47:M106–M110.
4. Fries JF, Cella D, Rose M, et al. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. *J Rheumatol*. 2009;36:2061–2066.
5. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. 2000;38(suppl): II-28–II-42.
6. Hays RD, Liu H, Spritzer K, et al. Item response theory analyses of physical functioning items in the Medical Outcomes Study. *Med Care*. 2007;45:S32–S38.
7. Rose M, Bjorner JB, Becker J, et al. Evaluation of a preliminary physical function item bank supported the expected advantage of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2008;61:7–33.
8. Katz S, Down TD, Cash HR, et al. Progress in the development of the index of ADL. *Gerontologist*. 1970;10:20–30.
9. Paz SH, Spritzer KL, Morales LS, et al. Evaluation of the Patient-Reported Outcomes Information System (PROMIS®) Spanish physical functioning items. *Qual Life Res*. 2013;22:1819–1830.
10. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297–334.
11. Samejima F. The graded response moel. In: van der Liden WJ, Hambleton R, eds. *Handbook of Modern Item Response Theory*. New York, NY: Springer; 1996:85–100.
12. Rothrock NE, Hays RD, Spritzer K, et al. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010; 63:1195–1204.
13. Hays RD, Spritzer KL, Amtmann D, et al. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Arch Phys Med Rehabil*. 2013;94:2291–2296.
14. Muthén LK, Muthén BO. *Mplus User's Guide*, 7th ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.
15. SAS Institute. 2014*SAS® 94 TS Level 1M2*. Cary, NC: SAS Institute.
16. Nunnally J. *Psychometric Theory*, 2nd ed. New York: McGraw-Hill; 1978.
17. Hays RD, Spritzer KL, Fries JF, et al. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement and Information System (PROMIS) 20-Item Physical Functioning Short-Form in a prospective observational study of rheumatoid arthritis. *Ann Rheum Dis*. 2015;74:104–107.
18. Cella D, Riley W, Stone A, et al. Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005–2008. *J Clin Epidemiol*. 2010;63: 1179–1194.
19. Craig BM, Reeve BB, Brown PM, et al. U.S. valuation of health outcomes measured using the PROMIS-29. *Value Health*. 2014;17: 846–853.