

# UCLA

## UCLA Previously Published Works

### Title

Performance of the Physical Functioning Activities of Daily Living Scale in the 2020 Medicare Health Outcomes Survey

### Permalink

<https://escholarship.org/uc/item/8t11d34j>

### Journal

Archives of Physical Medicine and Rehabilitation, 105(4)

### ISSN

0003-9993

### Authors

Hays, Ron D

Elliott, Marc N

### Publication Date

2024-04-01

### DOI

10.1016/j.apmr.2023.11.005

Peer reviewed



## ORIGINAL RESEARCH

# Performance of the Physical Functioning Activities of Daily Living Scale in the 2020 Medicare Health Outcomes Survey

Ron D. Hays, PhD,<sup>a</sup> Marc N. Elliott, PhD<sup>b</sup>

From the <sup>a</sup>UCLA Division of General Internal Medicine, Department of Medicine, Los Angeles, CA; and <sup>b</sup>RAND Corporation, Santa Monica, CA.

## Abstract

**Objective:** Assessing functional limitations for adults at high risk of frailty yields valuable information for identifying those in need of therapy. We evaluate a self-report measure used to assess physical function among Medicare recipients in the United States.

**Design:** Secondary analysis of the 2020 Medicare Health Outcomes Survey.

**Setting:** A random sample of adult enrollees of 510 managed care plans.

**Participants:** 287,476 adults (37% completion rate): 58% women; 16% were <65 years old (entitled via disability), 50% 65-74, and 34% 75 or older; 77% White, 14% Black, and 8% another race; 19% had <high school education.

**Interventions:** Not applicable.

**Main Outcome Measure:** We evaluate item distributions, dimensionality, monotonicity of response options, reliability, and validity of the 8-item Physical Functioning Activities of Daily Living (PFADL) scale.

**Results:** Most reported they could do 6 basic activities of daily living without difficulty. More limitations were reported for the other 2 PFADL items: 32% were *not limited at all* in climbing several flights of stairs and 40% in moderate activities. Product-moment correlations among the 8 items ranged from  $r=0.19$  between the easiest-to-do (eating) and most difficult-to-do (climbing several flights of stairs) items to  $r=0.73$  between bathing and dressing. The coefficient alpha and omega for the 8-item scale were both 0.86. Item slopes ranged from 2.6 (climbing several flights of stairs and eating) to 4.8 (dressing). Item characteristic curves revealed that response options were most likely to be selected in the appropriate order along the physical functioning continuum. The PFADL had at least 0.80 reliability between about -3 SDs below the mean to the mean. It was negatively correlated with comorbid condition count, disability days, problems with balance or walking, falling, and obesity.

**Conclusions:** The PFADL is useful for assessing average or below physical function in Medicare recipients.

Archives of Physical Medicine and Rehabilitation 2023;000:1–8

© 2023 by the American Congress of Rehabilitation Medicine.

Physical function predicts hospitalizations, institutionalization, and mortality.<sup>1-3</sup> It is associated with medical conditions such as chronic obstructive pulmonary disease, back pain, osteoarthritis,<sup>4</sup> loneliness,<sup>5</sup> and social factors like neighborhood and physical environment, economic stability, and education.<sup>6</sup> Assessing functional limitations in older individuals is important,<sup>7,8</sup> especially for patients at high risk of frailty (eg, kidney disease).<sup>9</sup> Assessment of physical function yields valuable information for physical

therapy, self-management, and referral to community-based exercise programs.<sup>10</sup>

Random samples of patients from all managed care organizations with Medicare contracts in the United States complete the Medicare Health Outcomes Survey (HOS) annually. Results are used for quality improvement, pay for performance, public reporting, and program oversight. The Physical Functioning Activities of Daily Living (PFADL) scale is included in the HOS.

Six of the PFADL items assess basic activities of daily living (BADL): walking, dressing, bathing, getting in/out of chairs, eating, and toileting: (1) No, I do not have difficulty, (2) Yes, I have difficulty, and (3) I am unable to do this activity. The scoring of the physical function items by the Centers for Medicare & Medicaid Services (CMS) assumes that the response selected is

This study was supported by Centers for Minority Aging Research Center for Health Improvement of Minority Elderly (RCMAR/CHIME) under NIH/NIA Grant P30-AG021684. The sponsor had no role in the design, methods, participant recruitment, data collection, analysis, or preparation of the paper.

Disclosures: The authors report no conflict of interest.

consistent with underlying level of physical function. Each response category of the items should have the highest likelihood of being selected somewhere along the underlying distribution consistent with *No, I do not have difficulty* representing the best functioning, *Yes, I have difficulty* the next best functioning, and *I am unable to do this activity* the worst functioning.<sup>11</sup> Some investigators score these items by collapsing the second and third response categories.<sup>12,13</sup> Doing this may discard potentially useful information about physical function.

The other 2 PFADL items are from the RAND-36 health survey.<sup>14</sup> They assess (1) moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf and (2) climbing several flights of stairs. For these items, patients are asked to select “No, not limited at all; Yes, limited a little; or Yes, limited a lot.” The extent to which these RAND-36 items are associated with the 6 BADL items and whether they improve the measurement of higher levels of physical functioning is unknown.

The assumptions underlying the scoring of the PFADL are that the 8 items are unidimensional and that the response options monotonically represent the level of physical function. Evidence in support of the reliability and validity of the PFADL for use among Medicare patients is also needed. Testing these assumptions can inform the selection and development of items to optimize future measurements of physical functioning. This study assesses whether the PFADL is unidimensional, its response options are monotonically related to underlying physical function, the reliability of the scale across the underlying physical function continuum, and its construct validity.

## Methods

### Ethics approval

The data collected in this study were collected in accordance with the Declaration of Helsinki. We obtained the public use files dataset that was constructed in accordance with CMS and Department of Health and Human Services policies and other applicable statutes and laws. All identifying information was excluded from the files, and demographic categories were aggregated so that identification of any given individual is not possible.

### Sample

CMS conducted a random sample of adult enrollees of 510 managed care plans, including those 65 and older and younger enrollees entitled via disability. Data were collected between August 17 through November 9, 2020. Of the 783,389 individuals sampled, 287,476 (37%) completed at least some of the items in the baseline survey (83% by mail and 17% by phone); 13,219 (5%) had missing data for more than half the items. The survey was completed

in either English, Spanish, or Chinese language. Further information about the Medicare HOS 2020 (cohort 23) baseline file is available online.<sup>15</sup>

### Measures

The PFADL items are shown in the [appendix](#). Each item is scored so that a higher number represents better physical functioning: limited a lot or unable to perform the ADL=0 points, limited a little or difficulty performing the ADL=1 point, and having no limitation or difficulty performing the ADL=2 points. The PFADL scale score sum ranges from 0 to 16 and a higher score represents better functioning.

We created a count of the number of 19 medical conditions that were included in the HOS survey: (1) blind or serious difficulty seeing; (2) serious difficulty hearing; (3) serious difficulty concentrating, remembering, or making decisions; (4) difficulty doing errands alone; (5) hypertension or high blood pressure; (6) angina pectoris or coronary artery disease; (7) congestive heart failure; (8) myocardial infarction or heart attack; (9) other heart conditions such as problems with heart values or the rhythm of the heartbeat; (10) stroke; (11) emphysema, or asthma, or chronic obstructive pulmonary disease; (12) Crohn’s disease, ulcerative colitis, or inflammatory bowel disease; (13) arthritis of the hip or knee; (14) arthritis of the hand or wrist; (15) osteoporosis; (16) sciatica; (17) diabetes; (18) depression; and (19) cancer (other than skin cancer). Also included were the Centers for Disease Control’s unhealthy physical days in the last 30 days,<sup>16</sup> a question about having a problem with balance or walking in the past 12 months, and a question about falling in the past 12 months. In addition, obesity was measured from self-reported weight in kilograms divided by the square of height in meters (obese=body mass index $\geq$ 30).

Education was categorized at eighth grade or less; some high school but did not graduate; high school graduate or General Education Diploma (GED); some college or 2-year degree; 4-year college graduate; more than a 4-year college degree. Age was categorized as less than 65, 65-74, and 75 or older.

### Analyses

We provide demographic information for the sample ([table 1](#)). Then, we compute frequencies and product-moment correlations among the 8 PFADL items. We examine principal component eigenvalues based on polychoric correlations to assess unidimensionality (ie, number of eigenvalues $>$ 1), estimate reliability (coefficient alpha<sup>18</sup> and omega<sup>19</sup>), scale means, and ceiling effects (percent with the best possible score). Next, we provide item characteristic curves, item thresholds and slopes, and scale information from an item response theory graded response model.<sup>20</sup> Finally, we assess our a priori hypotheses that worse physical functioning will be associated with the number of comorbid conditions, disability days due to physical health, having a problem with balance or walking in the past 12 months, falling in the past 12 months, obesity, and lower educational attainment. In addition, we hypothesized that Medicare recipients younger than 65 (entitled via disability) would have worse physical functioning than those 65 and older and that those 65-74 would have better physical functioning than those 75 and older. Analyses were conducted using SAS 9.4.<sup>a</sup>

#### List of abbreviations:

<b>BADL</b>	basic activities of daily living
<b>CAHPS</b>	Consumer Assessment of Healthcare Providers and Systems
<b>CMS</b>	Centers for Medicare & Medicaid Services
<b>HOS</b>	Medicare Health Outcomes Survey
<b>PFADL</b>	Physical Functioning Activities of Daily Living

**Table 1** Characteristics of the sample (n=287,476)

Variable	N (Percent)	Medicare Population <sup>17</sup>
<b>Women</b>	156,296 (58%)	55%
<b>Age</b>		
<65 years	46,005 (16%)	14%
65-74	142,812 (50%)	50%
75 or older	98,659 (34%)	37%
<b>Race</b>		
White	197,788 (77%)	77%
Black	36,753 (14%)	11%
Another race	21,670 (8%)	12%
<b>Education</b>		
Did not graduate high school	51,462 (19%)	14%
High school graduate/general education diploma	86,034 (32%)	33%
More than high school	128,452 (48%)	53%
<b>Marital status</b>		
Married	118,771 (44%)	51%
<b>Survey language</b>		
English	273,109 (95%)	
Spanish	14,003 (5%)	
Chinese	364 (<1%)	

## Results

The demographic characteristics of the sample were very similar to the Medicare population in 2019 (see [table 1](#)). The sample was 58% women; 77% White, 14% Black, and 8% another race; 16% were less than 65 years old, 50% 65-74, and 34% 75 or older. Nineteen percent had less than high school education, 32% high school or GED, and 48% more than a high school education. Forty-four percent were married. Ninety-five percent completed the survey in English, 5% in Spanish, and <1% in Chinese. The median number of 19 medical conditions reported was 3.8. The most common condition was hypertension (66%), followed by arthritis of the hip or knee (46%) and arthritis of the hand or wrist (39%).

### PFADL item frequencies

Frequencies for the PFADL items are shown in [table 2](#), ordered from the most difficult item to the least difficult (to report no limitations or difficulty performing the activity) item. Thirty-two percent of respondents reported not being limited when climbing several flights of stairs while 94% reported no difficulty with

eating. The 6 activities of daily living items are “easy” items for Medicare Advantage Organization members—that is, most report that they can do them without difficulty. The percentage of the sample not limited at all in moderate activities was 40%.

[Table 3](#) compares the percentage of the sample able to do 6 of the PFADL items with 2 other datasets. The current sample (2020 HOS baseline cohort 23) reported similar but slightly more limitations than the 2010 Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS) sample that also included those less than 65 years of age. The 2019 Medicare CAHPS sample, consisting of adults 65 and older in fee-for-service, reported slightly better physical functioning than the current sample.

### Associations among items and PFADL score distribution

Product-moment correlations among the items ranged from  $r=0.19$  between the easiest (eating) and most difficult item (climbing several flights of stairs) to  $r=0.73$  between bathing and dressing. Item correlations with the sum of the other 7 items in the PFADL scale

**Table 2** Frequencies for the Physical Functioning Activities of Daily Living Items

Activities of Daily Living	Not Limited at All	Limited a Little	Limited a Lot
Climbing several flights of stairs	88,615 (32%)	9702 (35%)	88,231 (32%)
Moderate activities	111,147 (40%)	96,519 (35%)	69,040 (25%)
	<b>No difficulty</b>	<b>Have difficulty</b>	<b>Unable to do</b>
Walking	168,686 (62%)	92,188 (34%)	9144 (3%)
Getting in or out of chairs	201,882 (75%)	64,143 (24%)	4959 (2%)
Bathing	221,099 (81%)	39,874 (15%)	10,348 (4%)
Dressing	231,088 (85%)	33,692 (12%)	6488 (2%)
Using the toilet	240,899 (89%)	24,393 (9%)	4080 (2%)
Eating	254,111 (94%)	14,819 (5%)	2600 (1%)

NOTE. Sample sizes for items vary because of missing data.

**Table 3** Percentage able to do 6 Physical Functioning Activities Without Difficulty in 3 Datasets

	2010 Medicare CAHPS* (n=366,701)	2019 Medicare CAHPS† (n=79,725)	2020 Health Outcomes Survey (HOS) Baseline Cohort 23
Walking	253,024 (69%)	58,120 (73%)	168,686 (62%)
Getting in or out of chairs	286,027 (78%)	64,737 (81%)	201,882 (75%)
Bathing	311,696 (85%)	69,600 (87%)	221,099 (81%)
Dressing	322,697 (88%)	71,194 (89%)	231,088 (85%)
Using the toilet	333,698 (91%)	73,347 (92%)	240,899 (89%)
Eating	344,699 (94%)	75,420 (95%)	254,111 (94%)

\* Medicare managed care and fee-for-service respondents including those less than 65 years old.<sup>21</sup>

† Medicare fee-for-service 65 years of age and older (unpublished).

ranged from 0.44 (eating) to 0.72 (walking), indicating that there is substantial common variance among them. The principal component eigenvalues of the polychoric correlations support a single underlying factor (first 2 eigenvalues=5.99 and 0.85). The average residual correlation was 0.05.

The coefficient alpha and omega for the 8-item scale were both 0.86, indicating that the 8-item scale is reliable enough for group-level comparisons. The mean was 12.73 (SD=3.43, range 0-16) and 26% reported the highest possible (ceiling) score.

### Item response theory graded response model parameters

Figure 1 shows the item characteristic curves. These curves support the monotonicity assumption—that is, each response option is most likely to be selected somewhere along the physical functioning continuum in the expected order. This supports scoring the items as 3 levels rather than collapsing categories.

Table 4 provides item means and graded response model parameter estimates. Item thresholds are consistent with the item frequencies and item means. The item thresholds indicate that the ordering from most difficult to easiest items is climbing stairs, moderate activities, walking, getting in and out of chairs, bathing, dressing, using the toilet, and eating. The thresholds indicate the level of physical function on a *z* score metric needed to have a 50% chance of scoring below vs above the threshold. For example, a person with a physical function score of -0.56 (a little more than a half standard deviation below the mean) would have a 50% probability of selecting *yes, limited a lot* vs either *yes, limited a little*, or *No, not limited at all* on the question about climbing stairs. Item slopes (discrimination) ranged from 2.60 (climbing stairs and eating) to 4.76 (dressing), indicating that the dressing is the item most strongly associated with underlying physical functioning.

Figure 2 provides the test (scale) information curve. The scale information curve has 2 peaks and shows that the scale has at least 0.80 reliability between a little higher than -3 SDs below the mean up to a little bit above the mean.

### Construct validity

Each of the hypothesized associations was confirmed. The PFADL was negatively correlated with number of comorbid conditions ( $r=-0.56$ ,  $P<.0001$ ), disability days due to physical health ( $r=-$

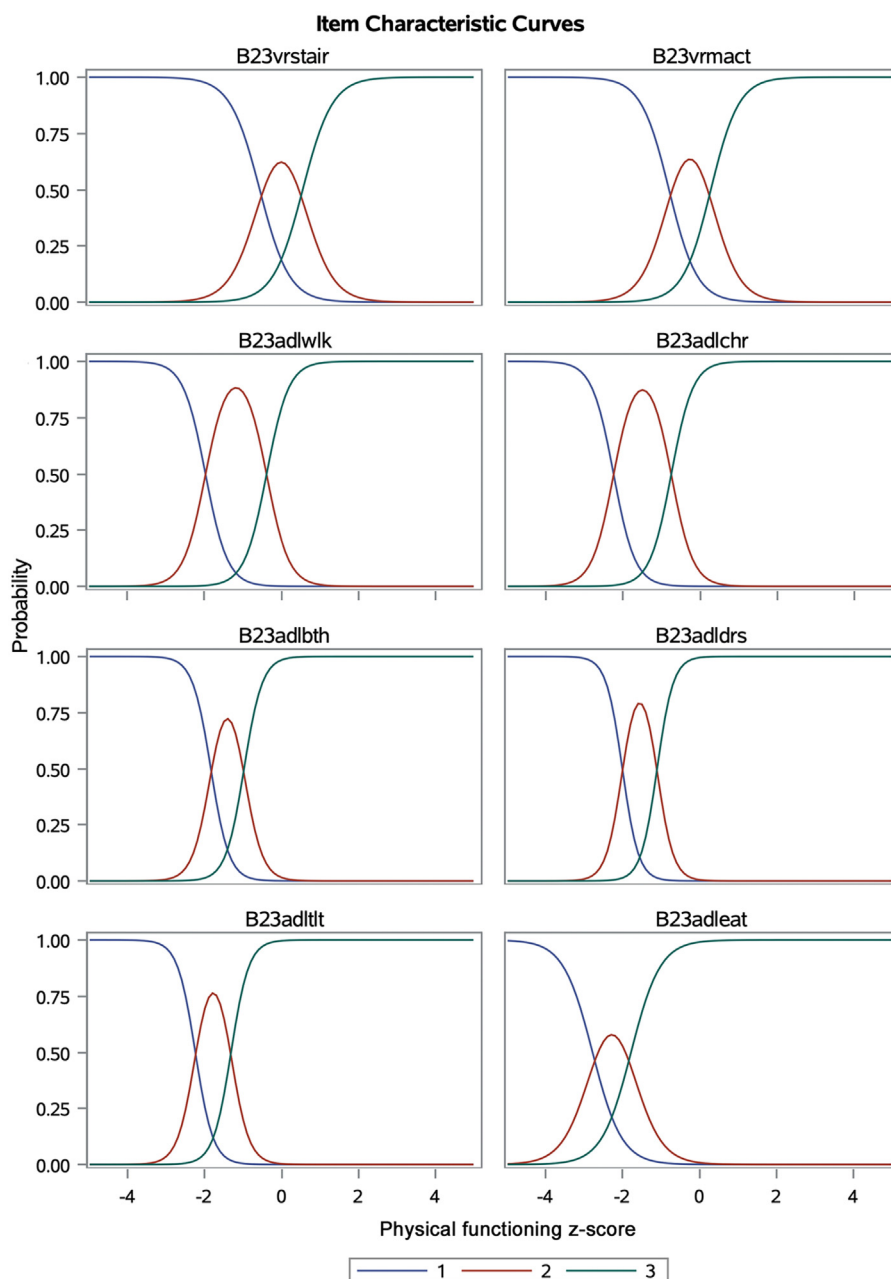
0.55,  $P<.0001$ ), having a problem with balance or walking in the past 12 months ( $r=-0.56$ ,  $P<.0001$ ), falling in the past 12 months ( $r=-0.32$ ,  $P<.0001$ ), obesity ( $r=-0.18$ ,  $P<.0001$ ), and PFADL was positively related to educational attainment ( $r=0.18$ ,  $P<.0001$ ). Moreover, there were significant differences by age ( $F(2, 283,220)=9539.34$ ,  $P<.0001$ ). Medicare respondents less than 65 years old (disabled) had significantly worse physical functioning than those 65 and older (effect sizes=0.70 and 0.38 for differences with 65-74 and with 75 and older, respectively), and those 75 and older had significantly worse physical functioning than those 65-74 (effect size=0.31); all pairwise comparisons significant at  $P<.0001$  using Tukey-Kramer adjustment for multiple comparisons.

### Discussion

This study provides a comprehensive psychometric evaluation of the PFADL measure in a large sample of Medicare recipients in the United States. The results of this study indicate that the PFADL items satisfy the monotonicity assumption underlying the scale scoring of the response options (0=*I am unable to do this activity/Yes, limited a lot*; 1=*Yes, I have difficulty/Yes, limited a little*; 2=*No, I do not have difficulty/No, not limited at all*), which supports preserving 3 categories in scoring. The item characteristic curves indicate that collapsing response categories would discard useful information. The analyses provide support for the unidimensionality of the PFADL items, consistent with prior assessments of physical function measures.<sup>21,22</sup>

Ceiling effects for measures are minimized by matching item difficulty to the level of physical function of respondents. Twenty-six percent of the participants in the current sample scored at the ceiling (best possible physical function) on the PFADL scale. In contrast, 65% of a sample of CAHPS Medicare respondents scored at the ceiling of a physical function scale that was limited to the 6 BADL items in the PFADL.<sup>21</sup> The lower ceiling effect found in this study was because the PFADL includes 2 RAND-36 items that assess higher levels of physical function than the 6 BADL items. Even smaller ceiling effects have been observed in other studies: 2% of inpatient geriatric rehabilitation patients had the best possible score on the PROMIS 24-item physical function scale.<sup>23</sup>

The reliability of the PFADL scale was 0.86, indicating sufficient reliability for group-level comparisons.<sup>24</sup> But scale



**Fig 1** Item characteristic curves. B23vrstair, climbing stairs; B23vrmact, moderate activities; B23adlwk, walking; B23adlchr, getting in/out of chairs; B23adlbth, bathing; B23adldrs, dressing; B23adlitt, using the toilet; B23adleat, eating.

information curves showed that the scale provides the most reliable information for the HOS respondents who have average or lower levels of physical functioning. The greater reliability of measurement among those with limitations compared with those with higher levels of functioning is consistent with other self-report measures such as the 4-item PROMIS-29 physical function scale,<sup>25</sup> the 6-item BADL scale administered to Medicare recipients,<sup>21</sup> and a 15-item Medical Outcomes Study scale.<sup>26</sup> Scale information might be enhanced for those with better physical functioning by including items that capture higher levels of function, such as items that assess doing chores like vacuuming or yard work, running a short distance, and engaging in 2 hours of physical labor that are included in the PROMIS 20-item short form.<sup>27</sup>

## Study Limitations

This study has limitations. Direct comparisons of the results of this study to other measures of physical function are limited by differences in the wording of survey questions.<sup>28</sup> In addition, mode effects are possible because the survey was administered using mail with phone follow-up. There was a small positive correlation of PFADL score with mail administration ( $r=0.09$ ,  $P<.0001$ ), but this may reflect the tendency for those with poorer physical functioning to complete surveys by phone. Because of the 37% response rate, it is unknown how well the results represent Medicare recipients more generally. Nonetheless, the findings of the study provide useful information about the PFADL scale administered annually to large samples of adults in Medicare.



**Table 4** Physical Functioning Activities of Daily Living Graded Response Model Item Parameters

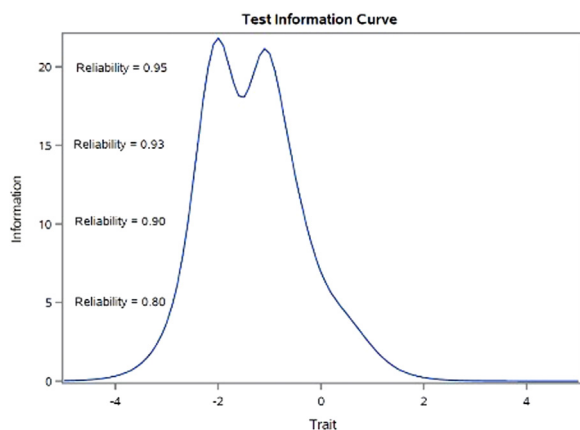
Item	N (%)	Threshold	Slope
B23vrstair (mean=2.00) Climbing stairs			2.60
Yes, limited a lot	88,231 (32)	-0.56*	
Yes, limited a little	97,021 (35)	0.56 <sup>†</sup>	
No, not limited at all	88,615 (32)		
B23vrmact (mean=2.15) Moderate activities			2.75
Yes, limited a lot	69,040 (25)	-0.81 <sup>‡</sup>	
Yes, limited a little	96,519 (35)	0.28 <sup>§</sup>	
No, not limited at all	111,147 (40)		
B23adlwlk (mean=2.59) Walking			3.49
I am unable to do this activity	9144 (3)	-1.97	
Yes, I have difficulty	92,188 (34)	-0.38	
No, I do not have difficulty	168,686 (63)		
B23adlchr (mean=2.73) Getting in/out of chairs			3.58
I am unable to do this activity	4959 (2)	-2.25	
Yes, I have difficulty	64,143 (24)	-0.75	
No, I do not have difficulty	201,882 (74)		
B23adlbth (mean=2.78) Bathing			4.22
I am unable to do this activity	10,348 (4)	-1.84	
Yes, I have difficulty	39,874 (15)	-0.98	
No, I do not have difficulty	221,099 (81)		
B23adldrs (mean=2.83) Dressing			4.76
I am unable to do this activity	6488 (2)	-2.01	
Yes, I have difficulty	33,692 (12)	-1.10	
No, I do not have difficulty	231,088 (85)		
B23adldtlt (mean=2.88) Using the toilet			4.32
I am unable to do this activity	4080 (2)	-2.25	
Yes, I have difficulty	24,393 (9)	-1.32	
No, I do not have difficulty	240,899 (89)		
B23adleat (mean=2.93) Eating			2.63
I am unable to do this activity	2600 (1)	-2.79	
Yes, I have difficulty	14,819 (5)	-1.78	
No, I do not have difficulty	254,111 (94)		

\* -1.82 in Patient-Reported Outcomes Measurement and Information System (PROMIS) general population sample.

<sup>†</sup> -0.09 in PROMIS general population.

<sup>‡</sup> -2.20 in PROMIS general population.

<sup>§</sup> -0.49 in PROMIS general population.



**Fig 2** Physical functioning activities of Daily Living Scale Information Curve.

## Conclusion

This study provides support for the reliability and validity of the PFADL for use in assessing Medicare enrollees but suggests potential value in adding items to assess more difficult activities of daily living. The findings can inform the selection of items to optimize future assessments of physical functioning.

## Suppliers

SAS 9.4; SAS Institute Inc

## Keywords

Aged; Functional status; Medicare; Rehabilitation; Self-assessment

## Corresponding author

Ron D. Hays, UCLA Department of Medicine, 1100 Glendon Avenue, Suite 850, Los Angeles, CA 90024. *E-mail address:* drhays@ucla.edu.

## Appendix Physical Functioning Activities of Daily Living items

2. The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

- Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf (*B23vrmact*)
  - Yes, limited a lot
  - Yes, limited a little
  - No, not limited at all
- Climbing several flights of stairs (*B23vrstair*)
  - Yes, limited a lot
  - Yes, limited a little
  - No, not limited at all

10. Because of a health or physical problem, do you have any difficulty doing the following activities without special equipment or help from another person?

- Bathing (*B23adlbth*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity
- Dressing (*B23adldrs*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity
- Eating (*B23adleat*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity
- Getting in or out of chairs (*B23adlchr*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity
- Walking (*B23adlwlk*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity
- Using the toilet (*B23adltrl*)
  - No, I do not have difficulty
  - Yes, I have difficulty
  - I am unable to do this activity.

## References

- Deardorff WJ, Jing B, Jeon SY, et al. Do functional status and Medicare claims data improve the predictive accuracy of an electronic health record mortality index? Findings from a national Veterans Affairs cohort. *BMC Geriatr* 2022;22:434.
- 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;43:527–39.
- Liotta G, Lorusso G, Madaro O, et al. Predictive validity of the short functional geriatric evaluation for mortality, hospitalization and institutionalization in older adults: a retrospective cohort survey. *Int J Nurs Sci* 2023;10:38–45.
- Wei MY, Luster JE, Ratz D, et al. Development, validation, and performance of a new physical functioning-weighted multimorbidity index for use in administrative data. *J Gen Intern Med* 2021;36:2427–33.
- Pedroso-Chaparro MDS, Cabrera I, Vara-Garcia C, et al. Physical limitations and loneliness: the role of guilt related to self-perception as a burden. *J Am Geriatr Soc* 2023;71:903–8.
- Tang J, Chen Y, Liu H, et al. Examining racial and ethnic differences in disability among older adults: a polysocial score approach. *Maturitas* 2023;172:1–8.
- O'Neill D, Forman DE. The importance of physical function as a clinical outcome: assessment and enhancement. *Clin Cardiol* 2020;43:108–17.
- Sin MK, Lee JA, Murphy PJM, Faselis C, Ahmed A. Upper extremity weakness: a novel risk factor for non-cardiovascular mortality among community-dwelling older adults. *Arch Gerontol Geriatr* 2023;112:105021.
- Roshanravan B, Patel KV. Assessment of physical functioning in the clinical care of the patient with advanced kidney disease. *Semin Dial* 2019;32:351–60.
- Shuman V, Brach JS, Bean JF, et al. Prevalence and predictors of ambulatory care physicians' documentation of mobility limitations in older adults. *Arch Phys Med Rehabil* 2023;104:719–27.
- Hambleton RK, Swaminathan H. Item response theory principles and applications. Boston, MA: Kluwer-Nijhoff; 1985.



12. Anderson C, Olshan A, Bae-Jump V, et al. Falls, walking or balance problems, and limitations in activities of daily living (ADLs) among older endometrial cancer survivors. *Support Care Cancer* 2022;30:6339–51.
13. Blackwood J, Karczewski H, Huang MH, et al. Katz activities of daily living disability in older cancer survivors by age, stage, and cancer type. *J Cancer Surviv* 2020;14:769–78.
14. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. *Ann Med* 2001;33:350–7.
15. Health Services Advisory Group. Medicare HOS Research Data Files. Available at: <https://www.hosonline.org/en/data-dissemination/research-data-files/>. February 23, 2023. Accessed April 24, 2023.
16. Centers for Disease Control and Prevention. Measuring Healthy Days. Atlanta, Georgia: CDC; 2000.
17. Tarazi W, Welch W, Nguyen N, et al. Medicare beneficiary enrollment trends and demographic characteristics. Issue Brief. Office of the Assistant Secretary for Planning and Evaluation. 2022. Available at: <https://aspe.hhs.gov/reports/medicare-enrollment>. Accessed April 24, 2023.
18. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
19. McDonald RP. Test theory: a unified treatment. Mahwah, NJ: Lawrence Erlbaum; 1999.
20. Samejima F. The graded response model. In: van der Linden WJ, Hambleton R, eds. *Handbook of modern item response theory*, New York, NY/USA: Springer; 1996:85–100.
21. Hays RD, Mallett JS, Gaillot S, et al. Performance of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) physical functioning items. *Med Care* 2016;54:205–9.
22. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measures in the 21st century. *Med Care* 2000;38. II-28-42.
23. Smit EB, Bouwstra H, Roorda LD, et al. A patient-reported outcomes measurement information system short form for measuring physical function during geriatric rehabilitation: test-retest reliability, construct validity, responsiveness, and interpretability. *J Am Med Dir Assoc* 2021;22:1627–32.
24. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York, NY: McGraw-Hill; 1994.
25. Hays RD, Spritzer KL, Reise SP. Using item response theory to identify responders to treatment: Examples with the Patient-Reported Outcomes Measurement Information System (PROMIS) physical functioning and emotional distress scales. *Psychometrika* 2021;86:781–92.
26. Hays RD, Liu H, Spritzer K, et al. Item response theory analyses of physical functioning items in the Medical Outcomes Study. *Med Car* 2007;45:S32–8.
27. 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;74:104–7.
28. Ankuda CK, Covinsky K, Freedman VA, et al. The devil's in the details: variation in estimates of late-life activity limitations across national cohort studies. *J Am Soc* 2023;71:858–68.