UCLA
UCLA Previously Published Works

Title

U.S. General Population Estimate for "Excellent" to "Poor" Self-Rated Health Item

Permalink

https://escholarship.org/uc/item/43z5t9hr

Journal

Journal of General Internal Medicine, 30(10)

ISSN

0884-8734

Authors

Hays, Ron D
Spritzer, Karen L
Thompson, William W
et al.

Publication Date

2015-10-01

DOI

10.1007/s11606-015-3290-x

Peer reviewed

U.S. General Population Estimate for "Excellent" to "Poor" Self-Rated Health Item

Ron D. Hays, Karen L. Spritzer, William W. Thompson & David Cella

Springer

Springer

Special Symposium: Assessing Health in the Patient-Reported Outcomes Measurement Information System (PROMIS)

U.S. General Population Estimate for "Excellent" to "Poor" Self-Rated Health Item

Ron D. Hays, PhD[1], Karen L. Spritzer, MS[1], William W. Thompson, PhD[2], and David Cella, PhD[3]

[1]Division of General Internal Medicine, Department of Medicine, UCLA, Los Angeles, CA, USA; [2]National Center of Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, USA; [3]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

BACKGROUND: The most commonly used self-reported health question asks people to rate their general health from *excellent* to *poor*. This is one of the Patient-Reported Outcomes Measurement Information System (PROMIS) global health items. Four other items are used for scoring on the PROMIS global physical health scale. Because the single item is used on the majority of large national health surveys in the U.S., it is useful to construct scores that can be compared to U.S. general population norms.

OBJECTIVE: To estimate the PROMIS global physical health scale score from the responses to the single *excellent* to *poor* self-rated health question for use in public health surveillance, research, and clinical assessment.

DESIGN: A cross-sectional survey of 21,133 individuals, weighted to be representative of the U.S. general population.

PARTICIPANTS: The PROMIS items were administered via a Web-based survey to 19,601 persons in a national panel and 1,532 subjects from PROMIS research sites. The average age of individuals in the sample was 53 years, 52 % were female, 80 % were non-Hispanic white, and 19 % had a high school degree or lower level of education.

MAIN OUTCOME MEASURES: PROMIS global physical health scale.

KEY RESULTS: The product–moment correlation of the single item with the PROMIS global physical health scale score was 0.81. The estimated scale score based on responses to the single item ranged from 29 (*poor* self-rated health, 2.1 SDs worse than the general population mean) to 62 (excellent self-rated health, 1.2 SDs better than the general population mean) on a T-score metric (mean of 50).

CONCLUSIONS: This item can be used to estimate scores for the PROMIS global physical health scale for use in monitoring population health and achieving public health objectives. The item may also be used for individual assessment, but its reliability (0.52) is lower than that of the PROMIS global health scale (0.81).

The most widely used self-reported health question elicits a rating on an *excellent* to *poor* response scale.[1] This item provides a general perception of health that reflects both objective health conditions and the individual's values for different aspects of health-related quality of life. Multiple studies have found this and other global health items to be predictive of health care utilization and mortality.[2–5] The item is also used as a case mix variable for patient experiences with care measures.[6]

The Patient-Reported Outcomes Measurement Information System (PROMIS®) project was funded by the National Institutes of Health (NIH) in 2004 with the objective of developing, evaluating, and disseminating publicly available survey items assessing self-reported generic health (www.nihpromis. gov).[7] The PROMIS vision was to create efficient measures that would be feasible to implement in busy office practices and that could provide a system of population health surveillance normed to the U.S. general population.

Four items are used to assess global physical health. Three of these are administered using five-category response scales, and one item (rating of pain on average) uses a response scale of 0–10:[8]

1) In general, how would you rate your physical health? *Excellent, Very Good, Good, Fair, Poor* (global03)
2) To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair? *Completely, Mostly, Moderately, A little, Not at all* (global06)
3) In the past 7 days, how would you rate your pain on average? Scale of 0 to 10, where 0= no pain and 10= worst pain imaginable (global07)
4) In the past 7 days, how would you rate your fatigue on average? *None, Mild, Moderate, Severe, Very severe* (global08)

We recoded global07 from its numeric 0–10 scale to five categories, based on a previous work.[9] The 4-item global physical health scale had an internal consistency reliability (coefficient alpha) of 0.81. It is scored on a T-score metric, with the mean (50) and SD (10) relative to the U.S. general population.[10] The scale has been included in several nationally

representative surveys in the U.S., including the 2010 National Health Interview Survey (NHIS) and two cohorts of the HealthStyles survey: a non-probability mail sample in 2010 and a probability-based Internet panel in 2012.[11] The average global physical health score in the HealthStyles samples was approximately 50, matching the PROMIS U.S. general population. The mean score in the NHIS was slightly higher, as it was administered using face-to-face interviews. The PROMIS global physical health scale was selected for the Healthy People 2020 initiative.

PROMIS also includes the self-reported general health item: "In general, would you say your health is: *excellent*, *very good*, *good*, *fair*, or *poor*?" (global01) This item was not used in scoring the PROMIS global health scale because it was so highly correlated (polychoric $r=0.95$) and was "locally dependent" (residual correlation of 0.29 when a one-factor categorical confirmatory factor analytic model was estimated) with a similarly worded item in the global physical health scale ("In general, how would you rate your physical health?"). Because the *excellent* to *poor* self-reported health item is used on the majority of nationally representative health surveys in the U.S., there is interest in estimating the PROMIS global physical health scale from this item alone such that the scale score can be used for population health surveillance. This study provides estimates of PROMIS global physical health scale scores from the in general how would you rate your health (global01) item.

METHODS

PROMIS items were administered via a Web-based survey to 19,601 persons in a national Internet panel maintained by Polimetrix[7] (now YouGov/Polimetrix; see http://research.yougov.com/) and to 1,532 subjects from PROMIS research sites ($n=21,133$ overall). Demographic quotas were specified for the PROMIS study participants selected from the Internal panel, but the sample is not a random sample of the U.S. general population. PROMIS measures were normed using post-stratification adjustment of the sample to match the 2000 U.S. census data on gender, age groups, race/ethnicity, education, marital status and income.[10,12]

The PROMIS global physical health scale was scored using item response theory (IRT) methods. IRT provides potential advantages over classical test theory, such as standard errors of measurement that vary by estimated physical health score and estimation of distances between response categories rather than assuming equal spacing between them. Details regarding IRT are discussed elsewhere.[13]

Analysis Plan

The demographic characteristics of the PROMIS sample are provided in Table 1. We estimate product–moment and Spearman rank-order correlations of the single item (global01 in Table 2) with the PROMIS global physical health scale. To

Table 1 Characteristics of Patient-Reported Outcomes Measurement and Information System (PROMIS) Respondents

Characteristics	PROMIS respondents ($n=21,133$)
% Female	52
% Hispanic	9
% African-American	9
% White (%)	80
% Less than high school education	3
% High school diploma/GED	16
% Greater than high school education	82
% Married/living with someone	65
% None of 25 chronic conditions	18
Mean age	53

evaluate the comparability of global01 with the similar item in the PROMIS global physical health scale (global03), we substitute global01 for global03 and recalibrate the scale using the IRT graded response model, which yields a discrimination parameter and threshold parameters for each item. The discrimination parameter (a) as shown in Table 2 is similar to an item-total correlation. Higher values of this parameter are associated with items that are better able to discriminate between contiguous levels of global physical health. The threshold parameters (b1–b4) represent the level of global physical health necessary to respond above threshold with a 0.50 probability.

Note that we did not calibrate global01 and global03 together with the other four global physical health items because of inflation of discrimination parameters due to local dependency (see Table 9.1 of Hays[14]). We compare the item parameters (discrimination and thresholds) from the 4-item variant of the global physical health scale that uses the global01 item with that of the existing scale. In addition, we use the graded response model results to estimate category response curves for global01. These category response curves represent the

Table 2 Item Parameters for Two 4-Item Variants of Global Physical Health (Graded Response Model)

	a	b1	b2	b3	b4
Global03	2.31	−2.11	−0.89	0.29	1.54
Global01*	2.31	−2.21	−1.01	0.17	1.46
Global06	2.99	−2.80	−1.78	−1.04	−0.40
Global06*	2.99	−2.80	−1.78	−1.05	−0.40
Global07	1.74	−3.87	−1.81	−0.67	1.00
Global07*	1.76	−3.86	−1.80	−0.67	0.99
Global08	1.90	−3.24	−1.88	−0.36	1.17
Global08*	1.88	−3.26	−1.89	−0.36	1.17

Global01: In general, how would you rate your health? Excellent, Very Good, Good, Fair, Poor; Global03: In general, how would you rate your physical health? Excellent, Very Good, Good, Fair, Poor; Global06: To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair? Completely, Mostly, Moderately, A little, Not at all; Global07: In the past 7 days, how would you rate your pain on average? Scale of 0–10, where 0= no pain and 10=worst pain imaginable; Global08: In the past 7 days, how would you rate your fatigue on average? None, Mild, Moderate, Severe, Very severe
a=discrimination parameter; b1–b4 are threshold parameters
* Indicates calibrations based on substituting global01 for global03

probability of responding in a particular category conditional on the estimated global physical health score.

Reliability can be estimated for different locations along the continuum of scores and for the average (marginal reliability) across the continuum from the graded response model. Information is analogous to reliability, and indicates the precision (reciprocal of the error variance) of an item or scale along the underlying continuum. We estimate information curves that provide reliability estimates for global01 and the existing four-item global physical health scale. The IRTPRO version 2.1 software program (Scientific Software International Inc., Skokie, IL, USA) was used to estimate the parameters of the graded response model.

We estimate mean scores on the existing PROMIS global health scale by response to the global health item (global01) using ANOVA and evaluate the significance of difference between least-squares means by global01 response category with the Tukey–Kramer adjustment for multiple comparisons using SAS software, Version 9.3 (SAS Institute Inc., Cary, NC, USA).

RESULTS

The demographics of the 21,133 respondents are shown in Table 1. The average age of the respondents in the sample was 53 years, and 52 % were female. The majority were white (80 %); 9 % were Hispanic, and 9 % African-American. Eighty-two percent had more than a high school degree. In addition, 65 % were married or living with someone, and 18 %

reported having none of 25 chronic conditions (conditions assessed included hypertension, angina, coronary artery disease, heart failure, heart attack, stroke, liver disease, kidney disease, arthritis or rheumatism, osteoarthritis, migraines, asthma, chronic obstructive pulmonary disease, diabetes, cancer, depression, anxiety, alcohol or drug problems, sleep disorder, HIV/AIDS, spinal cord injury, multiple sclerosis, Parkinson's disease, epilepsy, and amyotrophic lateral sclerosis).

The single item (global01) had a product–moment correlation of 0.81 (Spearman rho of 0.80) with the PROMIS four-item global health scale. The graded response model item parameters for the PROMIS four-item global health scale and the 4-item variant of the scale in which global01 is substituted for global03 are shown in Table 2.

The parameters for the two 4-item variants of global health were similar. For example, the discrimination parameter (a) for global01 and global03 was the same, and the corresponding threshold parameters ($b1$–$b4$) were very close to one another. Scale score estimates for the two variants of the 4-item scale were essentially identical (intraclass correlation=0.98). Hence, it is immaterial whether global01 or global03 is used in the scale.

The category response curves for the PROMIS global health item (global01) are shown in Fig. 1. The curves provide strong support for differentiation along the physical health continuum for each response option. People with the most positive estimated physical health scores (about 2 SDs above the mean and higher) have the highest probability of selecting *excellent*, while those at the other extreme (2 SDs below the mean and lower) have the highest probability of selecting
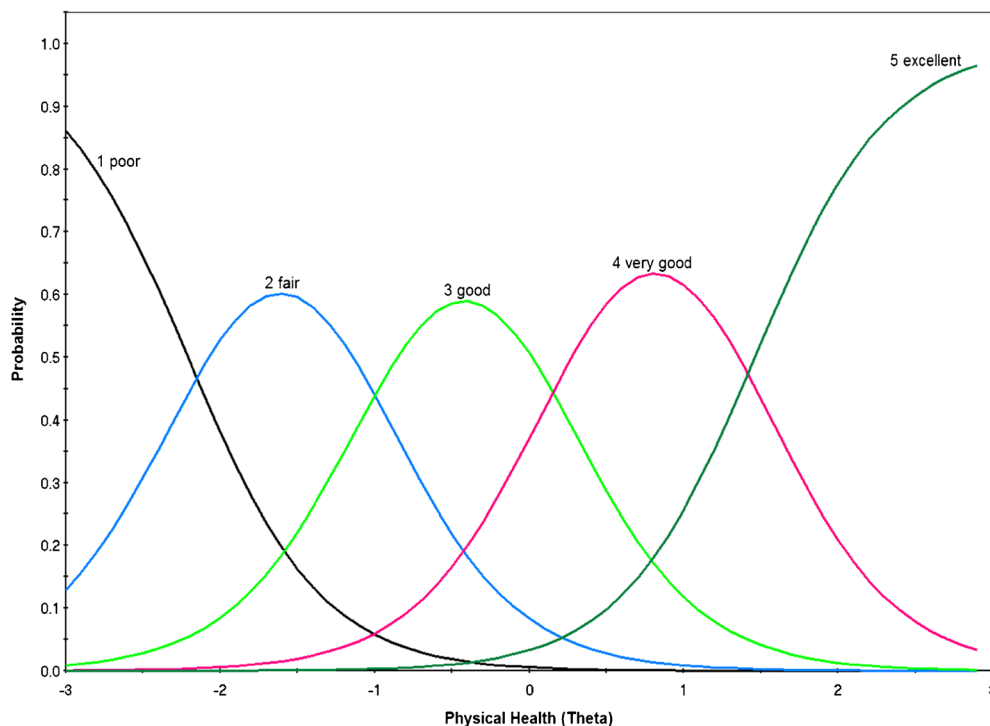

Fig. 1 Category Response Curves for PROMIS Global Health Item (global01) (Image file: Figure 1.tif).

poor. The *fair*, *good*, and very good response options are monotonically ordered between these two extremes.

Figure 2 provides the scale information and corresponding standard error of measurement (SEM=1/information$^{1/2}$) for the PROMIS 4-item global health scale on a z-score metric. Figure 3 provides the same detail for the global01 item. A SEM of 0.32 is equivalent to scale information of 10 and reliability of 0.90; a SEM of 0.45 is equivalent to scale information of 5 and reliability of 0.80. Information by level of physical health varied from about 1.5 to 6.8 for the scale (Fig. 2). The information for the single item varied from about 1.2 to 2.4 (Fig. 3). Hence, the reliability (1-SEM^2) for the PROMIS global health scale is 0.80 or higher from about −3 to 0.5 SDs relative to the mean, while the reliability for global01 is less than 0.80 throughout the range of physical health. The marginal reliability was 0.81 and 0.52 for the 4-item global physical health scale and the single item, respectively.

The ANOVA for the PROMIS global health scale score (dependent variable) by response category on global01 (independent variable) was statistically significant (F-statistic=9,836.82, $p<.0001$, dfs=4 and 21,099) for the overall PROMIS sample and the estimated means are provided in Table 3. These T-score means range from 29 (*poor* self-rated health) to 62 (*excellent* self-rated health), a difference of greater than 3 standard deviations.

DISCUSSION

This study provides estimated PROMIS global physical health scores from responses to the most widely used self-rated health item. Before discussing the implications of the study, it is important to recognize the tradeoffs in using a single item. While the item is strongly correlated with the 4-item global physical health scale score, its reliability is lower, and the standard error of measurement therefore higher, than the multi-item scale. In addition, future work is needed to compare the relative validity of the single item with the 4-item scale. Even with the seemingly strong correlation between the item and the scale, it is possible for the item to have different correlations with a criterion measure.[15] For example, Hays et al.[8] found that the PROMIS global physical health scale had a higher polyserial correlation than the single item with the EuroQol EQ-5D (0.82 vs. 0.65). In addition, the respondents to the survey were primarily members of an Internet panel. Despite this potential limitation, a previous study had shown that post-stratification adjustment produced characteristics similar to that of the U.S. general population.[10]

The present study reveals that individuals with *excellent* self-rated health are about 1.2 SDs better than the U.S. general population average, while those with *very good* health are about 0.4 SDs better. A self-rating of *good* is about 0.3 SDs worse than the U.S. general population mean, while *fair* and *poor* are 1.2 and 2.1 SDs worse, respectively. These results
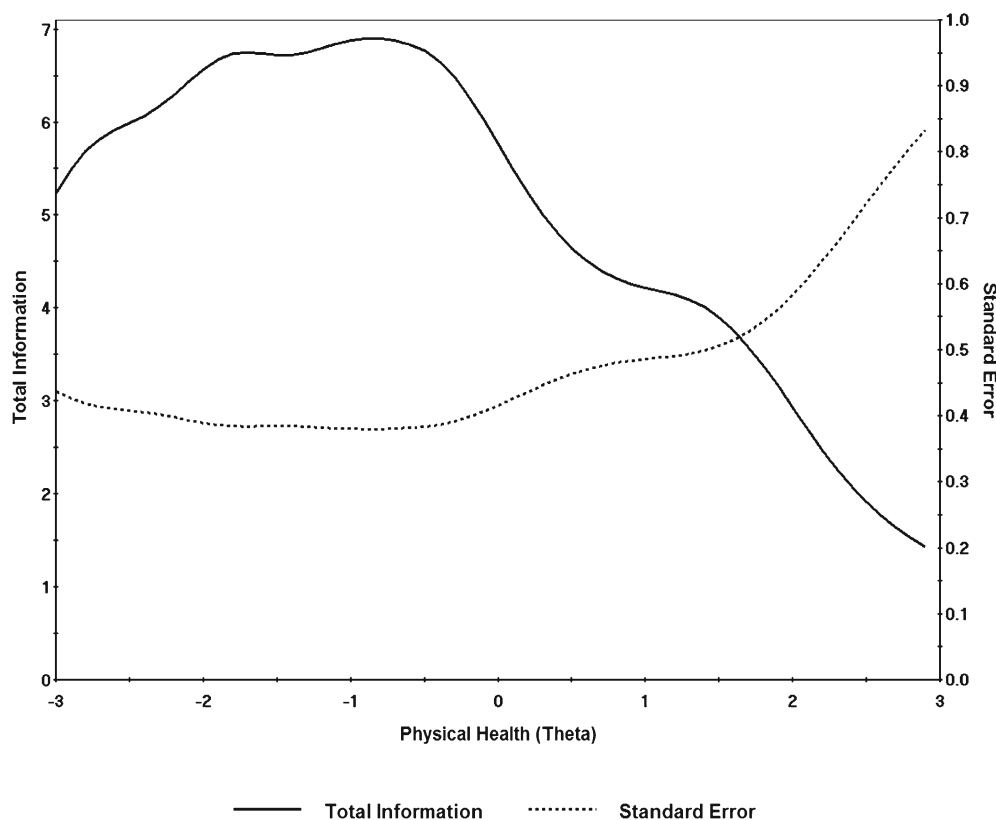

Fig. 2 Scale Information Curve for PROMIS 4-Item Global Physical Health Scale (Image file: Figure 2.tif).
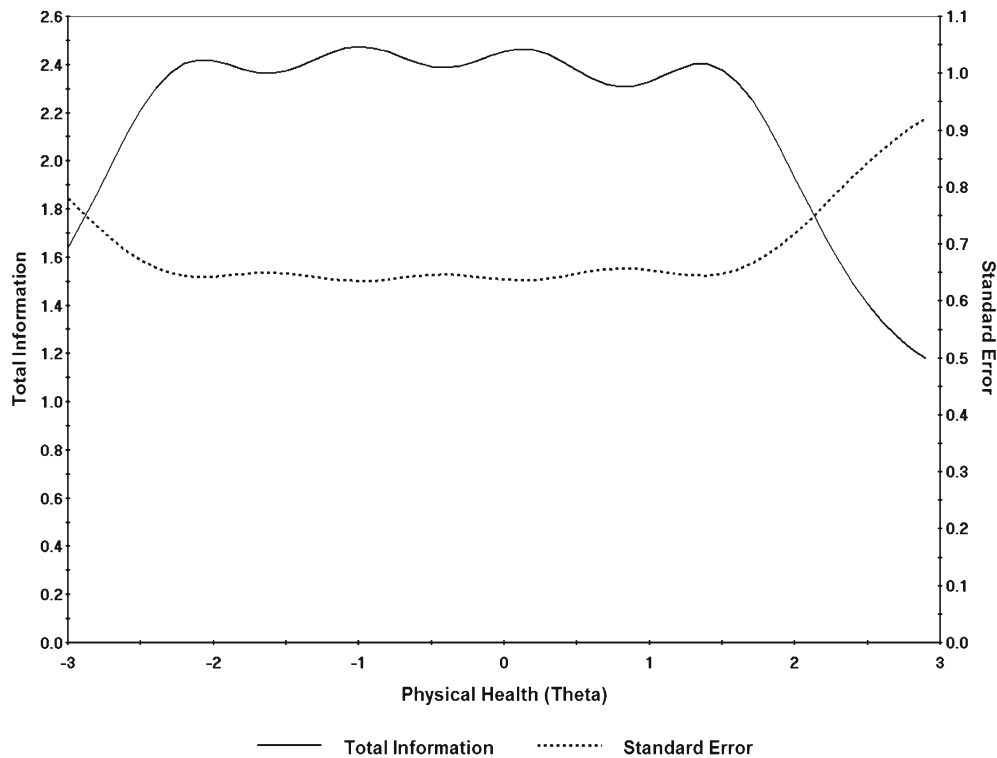
Fig. 3 Information Curve for PROMIS *Excellent* to *Poor* Global Health Item (Image file: Figure 3.tif).

have implications for the scoring of self-rated general health in other studies.

Some investigators who have administered the item have collapsed responses of *fair* or poor together and have collapsed responses of *good*, *very good* and *excellent*.[16,17] This study clearly shows that dichotomizing the five response levels results in the loss of important information. The more accurate scoring resulting from this study allows for better surveillance of the overall health of the U.S. population and for evaluating the attainment of Healthy People 2020 objectives.

The *excellent* to *poor* health item is one of the items in the SF-36 health survey. The original possible scoring range of 0–100 for this SF-36 item assigned 100 to *excellent* and 0 to *poor*, with 84, 61, and 25 for *very good*, good, and *fair*, respectively. This scoring was based on the logic employed by Stewart, Hays and Ware,[18] who "recoded the response

Table 3 PROMIS Global Physical Health Scale Means By Response to *Excellent* to *Poor* (global01) Item

In general, would you say your health is:	PROMIS Global Physical Health Scale Mean [standard error]
Excellent	62.45[a] [0.12]
Very good	54.42[b] [0.07]
Good	46.71[c] [0.07]
Fair	38.04[d] [0.10]
Poor	29.48[e] [0.20]

F (dfs=4 and 21,099)=9,836.82, p<0.0001. Superscript indicates that all pairwise comparisons of means in a column are significantly different (p<0.05) using the Tukey-Kramer adjustment of least-squares means for multiple comparisons.

choices of the overall health item … to better reflect the unequal intervals of the item" (p. 727). To derive the recoded values, the authors calculated the average score for the other four current health items for each response level of the excellent to poor health item. Recoding of the item was then based on transposing (interpolating) these means into the 0–100 possible range. The same methodology yielded very similar estimates (86, 64, and 28 were obtained for *very good*, good, and *fair*, respectively) in a sample of 1,844 adults from a general population mail survey in western Switzerland.[19]

If the estimated scores (means) for the *excellent* to *poor* response categories in Table 3 are transformed linearly to a possible range of 0–100, then *very good* is 76, *good* is 52, and *fair* is 26. The present study, for which the criterion for determining the distance between response categories is derived using a more accurate item response theory model than the simple summated score previously used for the SF-36, shows *very good* and *good* to be further away from *excellent* and closer to *fair* than is implied by the SF-36 scoring.

The excellent to poor item has been administered for decades in the U.S. on the National Health Interview Survey and the Behavioral Risk Factor Surveillance System (BRFSS). The results of our study provide a basis for more accurate estimates of health in national surveys of the U.S. general population and for comparisons of subgroup differences (e.g., age, gender). The item has been shown to be predictive of important criterion variables such as health care utilization and mortality.[5,20] It may also be useful for identifying patients for targeted interventions[20] and as part of the evaluation and planning of care for patients in clinical practice.[21] The item has

been used as part of behavioral change strategies by comparing responses to it with health-related behaviors and clinical measures.[22]

In summary, the *excellent* to *poor* self-rated health item is useful for population-level monitoring. Because it assesses general health perceptions, it reflects what is important to the patient in evaluating their health. The item can provide a useful complement to the specific measures used routinely to evaluate patients in clinical practice.[23] The study reported here provides essential information about how this item is scored and used for these purposes.

Corresponding Author: Ron D. Hays, PhD; Division of General Internal Medicine, Department of Medicine, UCLA, 911 Broxton Avenue, Los Angeles, CA 90024, USA (e-mail: drhays@g.ucla.edu).

REFERENCES

1. **Stewart AL, Hays RD, Ware JE.** Health perceptions, energy/fatigue, and health distress measures. In: Stewart AL, Ware JE, eds. Measuring Functioning and Well-being: The Medical Outcomes Study Approach. Durham, NC: Duke University Press; 1992:143–172.

2. **Bopp M, Braun J, Gutzwiller F, Faeh D.** for the Swiss National Cohort Study Group. Health risk or resource? Gradual and independent association between self-rated health and mortality persists over 30 years. PLoS One. 2012;7:e30795.

3. **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 Manag. 2012;43:527–539.

4. **Ware JE, Manning WG, Duan N, et al.** Health status and the use of outpatient mental health services. Am Psychol. 1984;39:1090–1100.

5. **Idler EL, Benyamini Y.** Self-rated health and mortality: a review of twenty-seven community studies. J Health Soc Behav. 1997;38:21–37.

6. **Elliott MN, Zaslavsky AM, Goldstein E, et al.** Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. Health Serv Res. 2009;44:501–518.

7. **Cella D, Riley W, Stone A, Rothrock N, Reeve B, Young S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, DeVellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Pilkonis P, Revicki D, Rose M, Weinfurt K, Lai J, & Hays RD.** Initial item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS) network: 2005–2008. J Clin Epidemiol. 63(11):1179–1194

8. **Hays RD, Bjorner J, Revicki DA, Spritzer K, Cella D.** Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. Qual Life Res. 2009;18:873–880.

9. **Norquist JM, Watson DJ, Yu Q, Paolini JF, McQuarrie K, Santanello NC.** Validation of a questionnaire to assess niacin-induced cutaneous flushing. Curr Med Res Opin. 2007;23:1549–1560.

10. **Liu HH, Cella D, Gershon R, et al.** Representativeness of the PROMIS Internet panel. J Clin Epidemiol. 2010;63:1169–1178.

11. **Riley W, Hays RD, Kaplan RM, Cella D.** Sources of comparability between probability sample estimates and nonprobability web samples estimates. Proceedings of the 2013 Federal committee on statistical methodology (FCSM) research conference. Available at: https://fcsm.sites.usa.gov/files/2014/05/B4_Riley_2013FCSM.pdf. Accessed March 2, 2015.

12. **Hays RD, Liu H, Kapteyn A.** Use of Internet panels to conduct surveys. Behavior Research Methods. Submitted.

13. **Cappelleri JC, Lundy JJ, Hays RD.** Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. Clin Ther. 2014;36(5):648–662.

14. **Hays RD.** Response 1 to Reeve's chapter: Applying Item response theory for questionnaire evaluation. In: Madans J, Miller K, Maitland A, Willis G, eds. Question Evaluation Methods: Contributing to the Science of Data Quality. Hoboken, New Jersey: Wiley & Sons, Inc; 2011:125–135.

15. **Hays RD, Reise S, Calderón JL.** How much is lost in using single items? J Gen Intern Med. 2012;27:1402–1403.

16. **Kawachi I, Kennedy BP, Glass R.** Social capital and self-rated health: a contextual analysis. Am J Public Health. 1999;89:1187–1193.

17. **Mithen J, Aitken Z, Ziersch A, Kavanagh AM.** Inequalities in social capital and health between people with and without disabilities. Soc Sci Med. 2015;126:26–35.

18. **Stewart AL, Hays RD, Ware JE.** The MOS short-form general health survey: reliability and validity in a patient population. Med Care. 1988;26:724–735.

19. **Pernegar TV, Gayet-Ageron A, Courvoisier DS, Agoritsas T, Cullati S.** Self-rated health: analysis of distances and transitions between response options. Qual Life Res. 2013;22:2761–2768.

20. **Bierman AS, Bubolz TA, Fisher ES, Wasson JH.** How well does a single question about health predict the financial health of Medicare managed care plans? Eff Clin Pract. 1999;2(2):56–62.

21. **Mavaddat N, Valderas JM, van der Linde R, Khaw KT, Kinmonth AL.** Association of self-rated health with multimorbidity, chronic disease and psychosocial factors in a large middle-aged and older cohort from general practice: a cross-sectional study. BMC Fam Pract. 2014;15:185.

22. **Bombak AE.** Self-rated health and public health: a critical perspective. Front Public Health. 2013;1:15.

23. **Jylhä M.** What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med. 2009;69:307–316.