UCLA
UCLA Previously Published Works

Title

Linking Physical and Mental Health Summary Scores from the Veterans RAND 12-Item Health Survey (VR-12) to the PROMIS® Global Health Scale

Permalink

https://escholarship.org/uc/item/7kb5t57t

Journal

Journal of General Internal Medicine, 30(10)

ISSN

0884-8734

Authors

Schalet, Benjamin D
Rothrock, Nan E
Hays, Ron D
et al.

Publication Date

2015-10-01

DOI

10.1007/s11606-015-3453-9

Peer reviewed

Linking Physical and Mental Health Summary Scores from the Veterans RAND 12-Item Health Survey (VR-12) to the PROMIS® Global Health Scale

Benjamin D. Schalet, Ph.D.[1], Nan E. Rothrock, Ph.D.[1], Ron D. Hays, Ph.D.[2], Lewis E. Kazis, Sc.D.[3], Karon F. Cook, Ph.D.[1], Joshua P. Rutsohn, M.P.H.[1], and David Cella, Ph.D.[1]

[1]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; [2]Department of Medicine, University of California, Los Angeles, Los Angeles, USA; [3]Center for the Assessment of Pharmaceutical Practices (CAPP), Department of Health Policy and Management, Boston University School of Public Health, Boston, USA.

BACKGROUND: Global health measures represent an attractive option for researchers and clinicians seeking a brief snapshot of a patient's overall perspective on his or her health. Because scores on different global health measures are not comparable, comparative effectiveness research (CER) is challenging.

OBJECTIVE: To establish a common reporting metric so that the physical and mental health scores on the Veterans RAND 12-Item Health Survey (VR-12©) can be converted into scores on the corresponding Patient Reported Outcomes Measurement Information System (PROMIS®) Global Health scores.

DESIGN: Following a single-sample linking design, participants from an Internet panel completed items from the PROMIS Global Health and VR-12 Health Survey. A common metric was created using analyses based on item response theory (IRT), producing score cross-walk tables for the mental and physical health components of each measure. The linking relationships were evaluated by calculating the standard deviation of differences between the observed and linked PROMIS scores and estimating confidence intervals by sample size.

PARTICIPANTS: Participants (*N*=2025) were 49 % male and 73 % white; mean age was 46 years.

MAIN MEASURES: Mental and physical health subscales of the PROMIS Global Health and the VR-12. The mean VR-12 physical component and mental component scores were 45.2 and 46.6, respectively; the mean PROMIS physical and mental health scores were 48.3 and 48.5, respectively.

KEY RESULTS: We found evidence that the combined set of VR-12 and PROMIS items were relatively unidimensional and that we could proceed with linking. Linking worked better between the physical health than mental health scores using VR-12 item responses (vs. linking based on algorithmic scores). For each of the cross-walks, users can minimize the impact of linking error with modest increases in sample sizes.

CONCLUSIONS: VR-12 scores can be expressed on the PROMIS Global Health metric to facilitate the evaluation of treatment, including CER. Extending these results to other common measures of global health is encouraged.

INTRODUCTION

Global health measures are patient-reported evaluations that are used to assess a patient's overall perspective on his or her health. They contrast with domain-specific patient-reported outcome measures, which target narrower health concepts, such as fatigue, anxiety, or mobility. Given their brevity and broad scope, global health measures are often utilized by health services researchers to assess a patient's overall perspective on his or her health. Global health perceptions are predictive of future health care utilization and mortality.[1,2] Inclusion of global health is particularly important in comparative effectiveness research (CER) studies of patients with multiple chronic conditions.[3] Examples of these measures include the Patient Reported Outcomes Measurement Information System Global Health Scale (PROMIS Global Health)[4] and the Veterans RAND 12-Item Health Survey (VR-12).[5–10]

Administering global health measures allows clinical investigators to measure health change beyond the direct target of the intervention. When treating back pain, for example, it may be useful to know whether global perceptions of mental health also improved and by how much (relative to physical health). In addition, because global health measures are typically brief, their routine administration allows clinicians to compare self-reported health across diverse patient groups.

A problem, arises, however, when we wish to compare patient scores on different global health instruments. Comparison is challenging because of differences in the question (or item) content, the response options, and scoring rules. We need a common metric to compare scores. Converting scores

from each sample to percentile ranks or z-scores produces instruments with comparable units, but the resulting metric would be sensitive to the peculiarities of each sample, such as the size, restricted range, and the distribution of scores (e.g., skewed vs. normal).[11] The instruments can also differ in terms of coverage with respect to the levels of health and dysfunction being assessed. For these reasons, it is desirable to conduct a "linking" study in which a large sample of participants responds to questions from *both* instruments. With such a design, it may be possible to align the scores on different measures to a common metric with greater accuracy.[12]

Recognizing this issue, the National Cancer Institute funded the PROsetta Stone® project (www.prosettastone.org) to align scores from many outcome instruments.[13] We report here on the results conducted as part of this project to link the VR-12 scores to the PROMIS Global Health metric. PROMIS Global Health is a 10-item questionnaire that assesses self-reported overall health.[4] It was recently developed as part of the PROMIS initiative (www.nihpromis.org), which utilized state-of-the-science qualitative and quantitative methods.[14,15] An example of a PROMIS Global Health item is, "In general, how would you rate your mental health, including your mood and your ability to think?"

The VR-12 is a short-form version of the Veterans RAND 36-Item Health Survey (VR-36), which was developed and modified from the RAND-36 v1.0 (MOS SF-36).[5–8,10,16] The VR-12 is freely available from the principal developers and is included in an ongoing evaluation of the CMS Medicare Advantage program.[17] The VR-12 is also included in the Ambulatory Care Survey of Healthcare Experiences of Patients (SHEP), sponsored by the Veterans Health Administration.[10] The VR-12 and the longer VR-36 have been administered approximately 7 million times over the past 15 years, and are represented in the literature by over 150 published articles.[9]

We linked items from the VR-12 on the PROMIS Global Health scale. Once linked, cross-walk charts for these measures were created such that scores on one instrument were matched with corresponding scores on the other.

METHOD
Measures

PROMIS Global Health. The PROMIS Global Health instrument produces two scores—the Global Physical Health (GPH) and Global Mental Health (GMH)—which are each based on four items. The GPH scale comprises items on physical health, physical functioning, pain intensity, and fatigue. The GMH scale includes items on overall quality of life, mental health, satisfaction with social activities and relationships, and emotional problems. GPH and GMH items were calibrated using the graded response model (GRM).[18] The GRM estimates parameters representing item location (level

of health) and discrimination (ability to distinguish people at different levels of health). The scores were centered on the 2000 US Census with respect to age, sex, education, and race-ethnicity.[4,19] The scores are on a T-score metric that has a mean of 50 and standard deviation of 10. Two PROMIS Global items (general health and social roles) are not used to score the GPH or GMH.

Veterans RAND 12-Item Health Survey (VR-12). The VR-12 is a patient-reported instrument from which physical and mental health component summary scores (PCS and MCS) are derived. The VR-12 items assess physical functioning, role limitations due to physical or mental health problems, pain, energy, mental health, social functioning, and general health. The VR-12 modified some of the role limitation questions to improve the instrument's performance and to make it easier to administer; there are formulas available to convert VR-12 scores to MOS SF scores.[7,8] PCS and MCS scores are derived using an algorithm that is referenced to a metric centered at 50.0 using the 2000–2002 US Medical Expenditure Panel Survey population.[10] The algorithm imputes missing responses using a modified regression estimate. The model is based on extraction of component scores and an orthogonal rotation in order to minimize the correlation between subscale scores. Responses to each item are used to create the PCS and MCS summary scores. The item weights account for differences in the strength of relationships between individual items and the PCS and MCS. As with PROMIS Global Health, VR-12 scores are standardized using a T-score metric with a mean of 50 and a standard deviation of 10.

Sample

Participants were recruited and data collected by an Internet survey company (www.op4g.com) that maintains a panel of respondents from the general population. To ensure adequate demographic representation, we imposed minimum enrollment requirements for age, gender, race, ethnicity, and education. Only those who indicated that they were 18 years of age or older were allowed to complete the survey. In addition to providing sociodemographic and clinical information and responding to questions on other health domains, participants completed the items of the PROMIS Global Health and the VR-12. Table 1 shows the demographic characteristics of the sample of 2025 participants who completed both measures: 49 % were male, and the mean age was 46 years (SD=18, range 18 to 100). The sample mean score on the VR-12 was 45.2 (SD=10.0) for the physical component and 46.6 (SD=11.1) for the mental component. On PROMIS Global Health, the mean score was 48.3 (SD=9.0) for physical health and 48.5 (SD=10.0) for mental health. To check the robustness of the analyses, the sample was split into training and validation halves and analyzed separately, but the results were very

Table 1 Demographic Characteristics of Participants Linking VR-12 to PROMIS Global Health (*N*=2025)

Participant characteristic	Percentage
Gender	
Male	49
Ethnicity	
Hispanic	16
Race	
White	73
Black/African American	12
Asian	5
Multiracial	3
Other	7
Education	
Less than high school	14
High school diploma or GED	31
Some college/technical degree/vocational program	28
Further educational attainment	27
Mean age (range), years	46 (18–100)

similar to those of the entire sample, so only the entire sample results are presented.

Analyses

In the PROsetta Stone project, each pair of instruments was typically subjected to multiple linking methods.[13,20] This approach includes methods based on IRT and one commonly used non-IRT method (equipercentile linking).[21,22] IRT is a family of mathematical models that estimate properties of each item response relative to a single dimension that is measured by all the items.[23,24] Previous studies in several health domains have demonstrated that results are highly congruent across methods.[13,20,25] Here we report only the results of the fixed IRT calibration, consistent with other PROsetta reports. We fit the data to GRM,[18] the standard IRT model for the calibration of PROMIS instruments.[26] Details on methods are provided in Appendix A; we report on the accuracy of linking in Appendix C.

Linking to VR-12 Algorithmic and Sum Scores. Most available linking methods, including all IRT-based links, use individual item scores (or parameters) as the basis for the link. By contrast, equipercentile linking is based on scale scores.[21,22] In the present context, there are both strengths and weaknesses to item score and scale score linking. The VR-12 item-weighted scale scores are amenable to equipercentile linking but not to item score-based methods. In addition, the VR-12 can also be linked using individual item scores. In order to use the cross-walk tables derived from such links, however, researchers would need access to item-level response data. We elected to conduct linking both on PCS/MCS summary scores and item-level response data. Fixed-parameter IRT linking was conducted using the GRM,[18] and incorporated the established PROMIS calibrations.[4]

VR-12 MCS scores were linked to PROMIS GMH scores; VR-12 PCS were linked to PROMIS GPH scores. For each component (mental or physical), we provide two sets of score cross-walk tables. One cross-walk table associates VR-12 MCS and PCS *algorithmic* summary scores with PROMIS GMH and GPH T-scores; the other table associates the *summed* item scores for each VR-12 subscale with PROMIS GPH and GMH T-scores.

Linking Assumptions. There are a number of assumptions made prior to linking, although the single-group design obviates these to some extent. The first assumption is that the linked instruments are measuring essentially the same concept.[12,27] We tested this by inspecting item content, calculating the item–total correlations and estimating the proportion of general factor variance of the combined set of items. A second linking assumption is that scores of the measures to be linked are highly correlated. We estimated correlations between PROMIS Global Health and VR-12 scores (algorithm- and sum-score based). In addition to linking assumptions, we tested the unidimensionality assumption of IRT using both confirmatory and exploratory factor analyses. Since IRT calibrations require that the combined item response data are essentially unidimensional, we conducted these analyses only on the combined items (e.g., all physical health items from PROMIS and VR-12). For details, see Appendices B and D.

Score Cross-walk Tables. While cross-walk tables are readily derived from equipercentile results, an extra step is necessary following IRT-based linking. We used the item parameter estimates derived from the fixed-parameter calibration to construct a cross-walk table by applying expected a posteriori (EAP) summed scoring.[28,29] Cross-walk tables map simple raw summed scores from each legacy instrument to T-score values on the PROMIS Global Health metric.

RESULTS

Item Content Overlap

We found substantial overlap between the VR-12 and the PROMIS Global Health item content, but also some differences (see Table 2). Each mental health scale includes an item that targets social activities (an additional PROMIS question on social roles is not used to score any subscale). Both assess feelings or emotion: PROMIS has one question, while the VR-12 has two. While PROMIS GMH has two questions that reference "mental health" and "quality of life," this wording is not included in the VR-12. Unlike PROMIS, the VR-12 MCS has two questions on how emotional problems interfere with personal productivity.

Each physical health scale has a question on pain intensity. Both measures assess physical function: PROMIS has one question, while the VR-12 has two. The VR-12 has an item on "energy," while PROMIS has an analogous "fatigue" question. Although both measures have a general health question, the VR-12 version contributes mostly to the PCS, while

Table 2 Mental and Physical Health Items (Questions) that were Combined in Order to Create a Cross-Walk Table of Scores

Instrument	Item	Text/Abbreviation of Combined Mental Health Items
PROMIS	Global02	In general, would you say your quality of life is:
PROMIS	Global04	In general, how would you rate your mental health, including your mood and ability to think?
PROMIS	Global05	In general, how would you rate your satisfaction with social activities and relationships?
PROMIS	Global10	How often have you been bothered by emotional problems such as feeling anxious, depressed or irritable? [Reverse-scored]
VR-12	4a	Accomplished less - due to emotional problems [Reverse-scored]
VR-12	4b	Limited - due to emotional problems [Reverse-scored]
VR-12	6a	Feeling peaceful
VR-12	6b	Have energy
VR-12	6c	Feeling downhearted [Reverse-scored]
VR-12	7	Interference with social activities [Reverse-scored]

Instrument	Item	Text/Abbreviation of Combined Physical Items
PROMIS	Global03	In general, how would you rate your physical health?
PROMIS	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?
PROMIS	Global07	How would you rate your pain on average? [Reverse-scored]
PROMIS	Global08	How would you rate your fatigue on average? [Reverse-scored]
VR-12	1	General health
VR-12	2a	Physical functioning - moderate activities
VR-12	2b	Physical functioning - climbing several stairs
VR-12	3b	Accomplished less - due to physical problems [Reverse-scored]
VR-12	3b	Limited - due to physical problems [Reverse-scored]
VR-12	5	Pain interference [Reverse-scored]
VR-12	6b	Have energy

The VR-12 energy item was used in both physical and mental health links; it contributes relatively equally to both physical and mental health component scores. PROMIS Global also includes two items not used to score any subscale, namely Global01 (general health) and Global09 (social roles). VR-12 Veterans RAND 12-Item Health Survey

the PROMIS general health question is not used to score any subscale. Finally, unlike PROMIS, the VR-12 PCS also has two questions that ask about role limitations due to physical health.

Correlations and Item Statistics

Pearson correlations were generally below the ideal range for linking (r<0.80).[13,20,30] The correlations between the PROMIS Global Health scales and the VR-12 raw sum scores were higher than correlations between the PROMIS Global Health scales and the VR-12 algorithmic scores (0.80 vs. 0.69 for physical health and 0.69 vs. 0.63 for mental health). Factor analytic results were mixed, with some evidence of multidimensionality, and yet also the presence of a large general factor. Despite less than optimal conditions, we proceeded, because our single sample design allowed us to directly test the accuracy of the link by comparing linked scores to actual scores. For details, please see Appendix B (factor analysis) and Appendix E (Table 1, correlations).

Score Cross-walk Tables

To construct cross-walk tables for IRT-based links, we used the item parameter estimates derived from the fixed-parameter calibrations.[28,29] Tables 3, 4, 5 and 6 show four cross-walk tables that associate VR-12 component scores with PROMIS Global Health scores. For both mental and physical health, we provide a summed VR-12 and an algorithmic VR-12 cross-walk.

Inspection of the algorithmic cross-walk tables shows that the population-based mean T-scores are very close for both components. A VR-12 MCS score of 50 is paired with a 50.3

on PROMIS GMH, while a VR-12 PCS score of 50 is paired with a score of 50.6 on PROMIS GPH. This result is expected and confirming, since both measures were centered on a US general population and both use the T-score metric. Given the

Table 3 Sum-Based VR-12 Mental Health Component Scores Associated with PROMIS Global Mental Health T-Scores

VR-12 MCS Score	PROMIS GMH T-Score	T-Score SE
6	17.9	4.3
7	20.4	4.5
8	22.7	4.5
9	24.9	4.5
10	26.8	4.5
11	28.6	4.5
12	30.3	4.4
13	31.9	4.4
14	33.4	4.4
15	34.9	4.4
16	36.4	4.4
17	37.8	4.4
18	39.2	4.4
19	40.6	4.4
20	42.0	4.4
21	43.4	4.4
22	44.9	4.4
23	46.4	4.5
24	47.9	4.5
25	49.5	4.6
26	51.1	4.6
27	52.9	4.7
28	54.7	4.8
29	56.7	4.9
30	59.0	5.0
31	61.8	5.1
32	65.4	5.4
33	70.2	6.0

GMH global mental health, MCS mental component score, VR-12 Veterans RAND 12-Item Health Survey

Table 4 Algorithm-Based VR-12 Mental Health Component Scores Associated with PROMIS Global Mental Health T-Scores

VR-12 MCS Score	PROMIS GMH T-score	VR-12 MCS Score	PROMIS GMH T-score
9	20.5	43	45.1
10	21.3	44	45.8
11	22.1	45	46.5
12	22.9	46	47.3
13	23.6	47	48.0
14	24.3	48	48.8
15	25.1	49	49.5
16	25.8	50	50.3
17	26.6	51	51.0
18	27.4	52	51.8
19	28.2	53	52.6
20	28.9	54	53.5
21	29.7	55	54.4
22	30.4	56	55.4
23	31.1	57	56.5
24	31.8	58	57.7
25	32.5	59	59.0
26	33.1	60	60.3
27	33.8	61	61.9
28	34.5	62	63.7
29	35.2	63	65.9
30	35.9	64	68.4
31	36.6	65	70.6
32	37.4	66	71.0
33	38.1	67	71.0
34	38.8	68	71.1
35	39.5		
36	40.2		
37	40.9		
38	41.6		
39	42.3		
40	43.1		
41	43.8		
42	44.4		

GMH global mental health, MCS mental component score, VR-12 Veterans RAND 12-Item Health Survey

Table 5 Sum Score-Based VR-12 Physical Health Component Scores Associated with PROMIS Global Physical Health T-Scores

VR-12 PCS Score	PROMIS GPH T-Score	T-Score SE
7	19.4	4.2
8	22.2	4.1
9	24.4	3.9
10	26.4	3.8
11	28.1	3.7
12	29.7	3.7
13	31.2	3.6
14	32.7	3.6
15	34.1	3.6
16	35.5	3.6
17	36.9	3.6
18	38.2	3.6
19	39.6	3.6
20	41.0	3.6
21	42.4	3.7
22	43.9	3.7
23	45.4	3.8
24	46.9	3.8
25	48.6	4.0
26	50.4	4.1
27	52.3	4.3
28	54.4	4.5
29	56.8	4.7
30	59.6	4.9
31	63.4	5.5
32	67.8	6.2

PCS physical component score, GPH global physical health, VR-12 Veterans RAND 12-Item Health Survey

Table 6 Algorithm-Based VR-12 Physical Health Component Scores Associated with PROMIS Global Physical Health T-Scores

VR-12 PCS Score	PROMIS GPH T-score	VR-12 PCS Score	PROMIS GPH T-score
10	16.6	41	43.1
11	17.4	42	43.9
12	18.4	43	44.7
13	19.6	44	45.5
14	20.9	45	46.3
15[*]	21.6	46	47.2
16	22.2	47	48.0
17	23.6	48	48.8
18	24.6	49	49.7
19	25.5	50	50.6
20	26.4	51	51.5
21	27.3	52	52.5
22	28.1	53	53.5
23	28.9	54	54.7
24	29.8	55	56.0
25	30.6	56	57.5
26	31.5	57	59.1
27	32.3	58	60.8
28	33.1	59	62.7
29	33.9	60	64.8
30	34.6	61	67.2
31	35.4	62	69.8
32	36.1	63	71.2
33	36.9	64	71.4
34	37.7	65	71.5
35	38.4	66	71.7
36	39.2		
37	40.0		
38	40.7		
39	41.5		
40	42.3		

[*]No participant scored 15; we linearly interpolated the PROMIS T-score

PCS physical component score, GPH global physical health, VR-12 Veterans RAND 12-Item Health Survey

different times when the metrics were set (approximately a decade apart), small differences were expected. The linking relationships are approximately linear until a T-score of about 55 (see Appendix E, Fig. 7), at which point the slope decreases. This may be due to the different methodologies used in instrument development, differences in item content, or lack of precision in measuring respondents at upper ranges.

DISCUSSION

This study provides tables that allow clinical researchers to convert VR-12 scores to PROMIS Global scores and vice versa. We did this by administering both measures to a large sample (*N*=2025) and then co-calibrating the item responses. These links permit the conversion of the VR-12 regardless of whether the scoring is based on the algorithmic or the sum score methods. We favor the latter approach, as the sum score linking tables do not require complex computer algorithms or specialized software, and they are easily adapted. For users who have item-level data, we recommend summing VR-12 items (according to Table 2) and obtaining linked PROMIS T-scores from Tables 3 and 5. The correlations between the VR-12 sum scores and PROMIS Global Health scores were higher, and their deviations smaller, than those obtained after linking

based on algorithm-derived component scores. This was particularly the case for the mental health scales.

There are several important clinical applications of this work. For example, our tables can be used to migrate performance improvement data to a single reporting metric. Clinical decision-makers at medical centers where the VR-12 is routinely administered may wish to replace it with the newer PROMIS Global Health instrument. Using our conversion tables, they can take advantage of historical patient data collected with the VR-12. Likewise, individual clinicians or departments may wish to integrate PROMIS Global Health into their practice, given its brevity and simplicity.[4,31] Our results allow users to leverage the accumulated evidence in support of a "legacy" instrument (such as the VR-12) to provide benchmark data and reference points,[10] though further validation of such points would be required.

Our results also facilitate CER. In CER, the accurate aggregation of results from multiple outcome studies is important. The individual studies that such meta-analyses comprise often use different instruments. Our results will improve the accuracy of such pooling, because individual results can be converted from the VR-12 to the PROMIS metric (and vice versa). Ultimately, this will allow clinical researchers to better understand the effects of particular treatments on global health.

While these results make linking possible and justifiable for group-level data, they are not as robust as results from similar work with questionnaires measuring single concepts like depression[13] and fatigue.[25] The raw score Pearson correlations (0.63 to 0.80) were lower than suggested thresholds for linking.[13,30] This is likely due to the breadth of the concept of global health, and the observed differences in item content, wording, and format. In addition, the scoring of these questionnaires is more complex than the more typical approach based upon raw sum scoring. PROMIS Global Health is scored using IRT-based parameters as applied to each of the four items that contribute to the total score. Unlike the PROMIS Global Health scales, the VR-12 components are weighted according to a factor analytic technique that forces the components to be orthogonal (unrelated), which will lead to inconsistent results compared to simple sum scores.[32] The impact of this difference in scoring between PROMIS Global Health and the VR-12 is evident in the generally weaker justification for linking scores based on the algorithmic approach.

There are some general limitations inherent in the linking process that should be noted. First, scores linked to the PROMIS metric based on the VR-12 may have more error than scores obtained directly from the PROMIS Global Health measure and vice versa (i.e., linking error is added to measurement error). Note, however, that the impact of this linking error can be reduced when users convert scores for a sample and compute its mean. For example, if one converts the VR-12 scores of 50 patients into PROMIS T-scores, the mean of those T-scores will be much more likely to reflect the "true" mean T-score relative to the conversion of 10 patient scores or a single one (see Appendix C.) Also, our linking tables should be used with the knowledge that concordance between the scores of any two instruments may be sensitive to population differences.[12] This potential limitation, however, is mitigated by our relatively large sample size.

The linking tables provided in this article are an important tool for clinical researchers who rely on the new PROMIS Global Health measures but are interested in comparing findings of contemporary work with many studies previously conducted with the VR-12. Data collected previously using the VR-12 can also be translated for comparison with current PROMIS Global Health scores. Future linking will advance the field even further by bringing together other measures of global health into one common metric.

Conflict of Interest: *Karon F. Cook is an unpaid officer of the PROMIS Health Organization; David Cella is an unpaid member of the board of directors and officer of the PROMIS Health Organization. All other authors declare that they have no conflict of interest.*

Corresponding Author: *Benjamin D. Schalet, Ph.D.; Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 625 N. Michigan Avenue, Suite 2700, Chicago, IL 60611, USA (e-mail: b-schalet@northwestern.edu).*

REFERENCES

1. **Bjorner JB, Fayers PM, Fayers PM, Hays RD.** Self-rated health. Assessing quality of life. Oxford: Oxford University Press; 2005:309–23.
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 Manag. 2012;43(3):527–39.
3. **Ahmed S, Berzon RA, Revicki DA, et al.** The Use of Patient-Reported Outcomes (PRO) within comparative effectiveness research: implications for clinical practice and health care policy. Med Care. 2012;50(12):1060–70.
4. **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(7):873–80.
5. **Kazis LE, Nethercot VA, Ren XS, Lee A, Selim A, Miller DR.** Medication effectiveness studies in the United States veterans administration health care system: a model for large integrated delivery systems. Drug Dev Res. 2006;67(3):217–26.
6. **Kazis LE, Miller DR, Skinner KM, et al.** Patient-reported measures of health: the veterans health study. J Ambul Care Manage. 2004;27(1):70–83.
7. **Kazis LE, Lee A, Spiro A, et al.** Measurement comparisons of the medical outcomes study and veterans SF-36 health survey. Health Care Financ Rev. 2004;25(4):43–58.
8. **Kazis LE, Miller DR, Clark JA, et al.** Improving the response choices on the veterans Sf-36 health survey role functioning scales: results from the veterans health study. J Ambul Care Manage. 2004;27(3):263–80.
9. Boston University School of Public Health. Veterans Rand 12-Item Health Survey (Vr-12) Website - Vr-36, Vr-12 and Vr-6d. http://www.bu.edu/sph/research/research-landing-page/vr-36-vr-12-and-vr-6d/. Accessed 13 June 2015.

10. **Selim A, Rogers W, Fleishman J, et al.** Updated U.S. Population standard for the veterans rand 12-item health survey (Vr-12). Qual Life Res. 2009;18(1):43–52.

11. **Baguley T.** Standardized or simple effect size: what should Be reported? Br J Psychol. 2009;100(3):603–17.

12. **Dorans NJ.** Linking scores from multiple health outcome instruments. Qual Life Res. 2007;16(Supplement 1):85–94.

13. **Choi SW, Schalet B, Cook KF, Cella D.** Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess. 2014;26(2):513–27.

14. **DeWalt DA, Rothrock N, Yount S, Stone AA.** PROMIS cooperative group. Evaluation of item candidates: the PROMIS qualitative item review. Med Care. 2007;45(5 Suppl 1):S12–21.

15. **Cella D, Riley W, Stone A, et al.** The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179–94.

16. **Selim AJ, Berlowitz D, Kazis LE, et al.** Comparison of health outcomes for male seniors in the veterans health administration and medicare advantage plans. Health Serv Res. 2010;45(2):376–96.

17. **Iqbal S**, **Rogers W**, **Selim A**, **et al**. The Veterans Rand 12 Item Health Survey (Vr-12): What It Is and How It Is Used. 2009; http://www.hosonline.org/surveys/hos/download/Veterans_RAND_12_Item_Health_Survey_VR-12_2007.pdf. Accessed 13 June 2015.

18. **Samejima F**. Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika Monograph Supplement, No. 17*. Richmond, VA: Psychometric Society; 1969: http://www.psychometrika.org/journal/online/MN17.pdf. Accessed 13 June 2015.

19. **Liu H, Cella D, Gershon R, et al.** Representativeness of the PROMIS Internet panel. J Clin Epidemiol. 2010;63(11):1169–78.

20. **Schalet BD, Cook KF, Choi SW, Cella D.** Establishing a common metric for self-reported anxiety: linking the MASQ, PANAS, and GAD-7 to PROMIS anxiety. J Anxiety Disord. 2014;28(1):88–96.

21. **Kolen MJ, Brennan RL.** Test equating, scaling, and linking: methods and practices. New York: Springer; 2004.

22. **Lord FM.** The standard error of equipercentile equating. J Educ Stat. 1982;7(3):165–74.

23. **Reeve B, Fayers PM.** Applying item response theory modelling for evaluating questionnaire item and scale properties. In: Fayers PM, Hays R, eds. Assessing quality of life in clinical trials. Oxford: New York Oxford University Press; 2005:55–73.

24. **Edelen MO, Reeve BB.** Applying Item Response Theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16(Suppl 1):5–18.

25. **Lai J-S, Cella D, Yanez B, Stone A.** Linking fatigue measures on a common reporting metric. J Pain Symptom Manag. 2014;48(4):639–48.

26. **Reeve BB, Hays RD, Bjorner JB, et al.** Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007;45(5 Suppl 1):S22–31.

27. **Noonan VK, Cook KF, Bamer AM, Choi SW, Kim J, Amtmann D.** Measuring fatigue in persons with multiple sclerosis: creating a crosswalk between the modified fatigue impact scale and the PROMIS fatigue short form. Qual Life Res. 2012;21(7):1123–33.

28. **Lord FM, Wingersky MS.** Comparison of IRT true-score and equipercentile observed-score "equatings". Appl Psychol Meas. 1984;8(4):453–61.

29. **Thissen D, Pommerich M, Billeaud K, Williams VSL.** Item response theory for scores on tests including polytomous items with ordered responses. Appl Psychol Meas. 1995;19(1):39–49.

30. **Dorans NJ.** Equating, concordance, and expectation. Appl Psychol Meas. 2004;28(4):227–46.

31. **Wagner LI, Schink J, Bass M, et al.** Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer. 2014;121(6):927–34.

32. **Farivar SS, Cunningham WE, Hays RD.** Correlated physical and mental health summary scores for the SF-36 and Sf-12 health survey, V.I. Health Qual Life Outcomes. 2007;5:54.

33. **Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A.** Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the medical outcomes study. Med Care. 1995;33(4 Suppl):AS264–79.