ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2022;000: 1−5

ORIGINAL RESEARCH

Crosswalking the National Institutes of Health Impact Stratification Score to the PEG

Ron D. Hays, PhD,[a,b] Nabeel Qureshi, MPH, MPhil,[b] Maria Edelen, PhD,[b,c] Anthony Rodriguez, PhD,[b] Mary Slaughter, PhD,[b] Patricia M. Herman, ND, PhD[b]

From the [a]UCLA Department of Medicine, Los Angeles, CA; [b]Health Care Division, RAND Corporation, Santa Monica, CA; and [c]Patient Reported Outcomes, Value and Experience (PROVE) Center, Department of Surgery, Brigham and Women's Hospital, Boston, MA.

Abstract

Objective: To crosswalk the National Institutes of Health (NIH) Pain Consortium's Research Task Force proposed Impact Stratification Score (ISS) to the PEG (Pain Intensity, Interference With Enjoyment of Life, Interference With General Activity) Scale.

Design: Cross-sectional data collected in 2021. Ordinary least squares regression analyses of ISS and PEG.

Setting: Amazon Mechanical Turk workers.

Participants: 1931 adults with back pain with an average age of 41 (range, 19-77); 48% were female, 16% Hispanic, 7% non-Hispanic Black, 5% non-Hispanic Asian, and 71% non-Hispanic White (N=1931).

Interventions: Not applicable.

Main Outcome Measures: The Patient-Reported Outcomes Measurement Information System (PROMIS)-29+2 v2.1 survey that includes the ISS, and the 3-item PEG.

Results: The ISS and PEG had a correlation coefficient of 0.74. The ISS accounted for 55% of the adjusted variance in the PEG and the standardized average deviation between observed and predicted scores (normalized mean absolute error) was 0.53. Likewise, the PEG explained 55% of the variance in the ISS with a normalized mean absolute error of 0.52.

Conclusions: This study provides a crosswalk between the ISS and PEG that can be used to predict one from the other. The regression equations can facilitate comparisons in studies that use different measures.

Archives of Physical Medicine and Rehabilitation 2022;000:1−5

An extensive body of research has evaluated interventions directed at adults with chronic pain using patient-reported outcomes.[1] It is challenging to synthesize findings across studies because of the plethora of measures. The National Institutes of Health (NIH) Pain Consortium's Research Task Force (RTF) on chronic low back pain noted that because of variations in study design and measures used, it is "difficult to compare epidemiologic data and studies of similar or competing interventions, replicate findings, pool data from multiple studies, resolve conflicting conclusions, develop multidisciplinary consensus, or even achieve consensus within a discipline regarding interpretation of findings."[2] (p 2029)

The NIH RTF focused on effect in terms of pain intensity and interference with activities and physical function.[2,3] They proposed an Impact Stratification Score (ISS) for chronic low back pain consisting of the Patient-Reported Outcomes Measurement Information System (PROMIS)-29 physical function, pain interference, and pain intensity measures. Hays et al[4] found internal consistency reliability estimates of 0.92-0.93 for the ISS. In addition, the ISS had a correlation coefficient of 0.75 to 0.84 with the Roland-Morris Disability Questionnaire, 0.51 to 0.75 with a single-item rating of average pain, and 0.64 to 0.71 with the PROMIS-29 v1.0 satisfaction with social role participation. The ISS was also found to be responsive to change. The area under the curve for the ISS predicting improvement on the retrospective rating of change from baseline to 6 weeks later was 0.83.

Funded by the National Center for Complementary and Integrative Health (NCCIH), Grant No. 1R01AT010402-01A1. NCCIH had no role in the design; data collection, analysis, or interpretation; or writing of this article.

Disclosures: none.

0003-9993/$36 - see front matter © 2022 by the American Congress of Rehabilitation Medicine. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
https://doi.org/10.1016/j.apmr.2022.08.006

The responsiveness of the ISS to change was shown in a prospective comparative effectiveness clinical trial of 750 active-duty US military personnel with low back pain.[5] As hypothesized, ISS scores improved for a substantial proportion of the sample. Thirty-seven percent of the sample improved significantly on the ISS over these 6 weeks and 59% reported on a retrospective change item that they were better (16% a little better, 14% moderately better, 23% much better, and 6% completely gone). Among those who improved significantly on the ISS, 89% reported that they were better on the retrospective rating item. Thirty-three percent of the sample improved significantly and reported improvement on the retrospective change item, 4% improved significantly but did not report that they were better on the retrospective change item, 26% did not improve significantly but reported improvement on the change item, and 37% did not improve significantly on the ISS or report improvement on the retrospective change item.

One measure increasingly used to assess the pain experience is the PEG (Pain Intensity, Interference With Enjoyment of Life, Interference With General Activity), a 3-item subset of the Brief Pain Inventory. The developers reported internal consistency reliability of 0.73 and 0.89 in 2 samples and construct validity comparable to the full Brief Pain Inventory.[6] In a subsequent clinical trial of 244 patients with persistent musculoskeletal pain of moderate severity, the PEG was better able to detect symptom change than the SF-36 Bodily Pain and PROMIS Pain Interference measures.[7]

The usefulness of the ISS and PEG will be enhanced with the availability of empirical crosswalks from one to the other so that researchers can interpret their results in the context of the other measure. Crosswalks also serve to help with the integration of results from studies using only the ISS or only the PEG and can be used for meta-analyses. This study provides regression equations to predict the ISS from the PEG and vice versa.

Methods

Data source

The data were collected in 2021 from Amazon Mechanical Turk (MTurk). MTurk is a source of temporary workers who are paid to complete tasks. The job or tasks are referred to as human intelligence tasks and include completing surveys, writing product descriptions, coding, or identifying content in images or videos. Eligible study participants had to complete a minimum of 500 previous human intelligence tasks on MTurk with a successful completion rate of at least 95%. All participants provided electronic consent at the start of the survey. Those who completed a general health survey and reported currently having back pain were asked to complete a back pain survey. Those who completed the general health and back pain survey were paid $4 for participation. The study was designed to administer the general health survey to approximately 6000 adults in order to obtain about 2000 completed back pain surveys. All procedures were reviewed and approved by the research team's institutional review board (RAND Human Subjects Research Committee FWA00003425; IRB00000051).

List of abbreviations:

ISS	**Impact Stratification Score**
MTurk	**Mechanical Turk**
NIH	**National Institutes of Health**
NMAE	**normalized mean absolute error**
OLS	**ordinary least squares**
PEG	**Pain Intensity, Interference With Enjoyment of life, Interference With General Activity Scale**
PROMIS	**Patient-Reported Outcomes Measurement Information System**
PROPr	**patient-reported outcome preference score**
RTF	**Research Task Force**

Measures

ISS

The PROMIS-29+2 v2.1 was administered. The ISS is made up of 9 PROMIS-29 items including 4 physical function items, 4 pain interference items, and 1 pain intensity item. Physical function (1=*without any difficulty*, 5=*unable to do*) and pain interference (1=*not at all*, 5=*very much*) each contribute from 4 to 20 points, and pain intensity (0-10 rating) contributes 0-10 points. The ISS has a possible range of 8 (least pain effect) to 50 (greatest pain effect).

PROMIS-29±2 v2.1

In addition to the 9 ISS items, the PROMIS-29+2 includes 5 multi-item scales with 4 items each (fatigue, sleep disturbance, depression, anxiety, ability to participate in social roles and activities) and a 2-item cognitive function scale.[8] In addition, physical health and mental health summary scores[9] and a single preference-based score, the patient-reported outcome preference score (PROPr), can be estimated.[10]

PEG

The 3 PEG items are (1) What number best describes your pain on average in the past week? (2) What number best describes how, during the past week, pain has interfered with your enjoyment of life? (3) What number best describes, how, during the past week, pain has interfered with your general activity? PEG response options range from 0 to 10, with 10 indicating the most severe pain. The PEG scale score is the mean of the 3 items and has a possible range of 0-10.

Analysis plan

We summarize demographic and health characteristics of the sample. Next, we report product-moment correlations of the PEG with the PROMIS-29+2 v2.1 measures. Then we estimate item-scale correlations (corrected for overlap of each item with the scale score) and internal consistency reliability[11] for the PEG and the ISS and report their means and standard deviations (SDs). A minimum bivariate correlation of 0.87 between the ISS and PEG is considered necessary for use of optimal methods such as item response theory co-calibration.[12] For correlations less than that, ordinary least squares (OLS) regression models have been used.[13] OLS models were evaluated in terms of R^2 and the normalized mean absolute error (NMAE). The NMAE statistic indicates the average deviation between the observed and predicted scores divided by the SD of the observed score. Lower NMAE values indicate better performance. There is no absolute rule of thumb for an acceptable NMAE but close to 0.50 is what was previously found for associations of PROMIS-29 scales with targeted disability measures.[13] We examined the PEG in predicting the ISS and

vice versa. Finally, we report correlations between the PEG and ISS by sex (female, male), ethnicity (non-Hispanic, Hispanic), race (non-White, White), and education (high school or less, more than high school).

Results

As seen in table 1, the sample of 1931 adults with back pain had an average age of 41 (range, 19-77). Forty-eight percent were female; 16% were Hispanic, 7% non-Hispanic Black, 5% non-Hispanic Asian, and 71% non-Hispanic White; 90% had more than high school education; 69% were married or living with a partner; and 69% were working full-time. The most common conditions reported were depression (49%), hypertension (41%), and anxiety (38%).

The mean PEG score was 4.02 (SD=2.12). Internal consistency reliability of the PEG was 0.89 and item-scale correlations corrected for overlap of each item with the scale ranged from 0.71 to 0.84. Table 2 reports PROMIS-29+2 v2.1 score means and SDs. The sample of respondents had worse physical function and cognitive function and more pain interference, pain intensity, fatigue, sleep disturbance, anxiety, depression, and worse health overall (physical health summary, mental health summary, and PROPr) than the general US population. The largest differences were medium (pain intensity, pain interference, anxiety, depression) or large (PROPr) effect sizes.[14] The mean ISS score was 20.68 (SD=8.06), falling within the "mild" range of severity.[2] Internal consistency reliability of the ISS was 0.79 and item-scale correlations (corrected for overlap) ranged from 0.59 to 0.75.

Table 3 shows that the correlations of the PEG with PROMIS-29+2 v2.1 scales ranged from −0.27 (cognitive distress) to 0.74 (ISS). Because they are all less than 0.80, these correlations support an OLS approach to crosswalking. The ISS was chosen for the crosswalk with the PEG because it had the largest correlation with it.

Table 1 Characteristic of the sample (n=1931)

Variable	Estimate
Age means, y (range)	41 (19-77)
Female (%)	48
Race/ethnicity (%)	
Hispanic	16
Non-Hispanic	
White	71
Black	7
Asian	5
Other	1
Education (%)	
<High school	0.2
High school graduate	10
Some college	16
Associate degree	8
Bachelor's degree	49
Master's degree	15
PhD or professional degree	2
Working full-time	69
Marital status (%)	
Married or living with partner	69
Never married	22
Separated, divorced, or widowed	9
Hypertension (%)	41
Arthritis (%)	23
Depression (%)	49
Anxiety (%)	38
Cancer (%)	7
Asthma (%)	22
Diabetes (%)	17
Chronic obstructive pulmonary disease (%)	8
Angina (%)	7
Heart disease (%)	8
Myocardial infarction (%)	6

Table 2 PROMIS-29+2 v2.1 Scale scores for the sample (n=1931)

Variable	Mean±SD
Physical function (4 items)	46 (8)
Pain interference (4 items)	56 (8)
Pain intensity (1 item)	57 (9)
Ability to participate in social roles and responsibilities	50 (9)
Fatigue	53 (9)
Sleep disturbance	52 (9)
Cognitive function	48 (8)
Anxiety	57 (9)
Depression	56 (9)
Physical health summary score	46 (8)
Mental health summary score	46 (8)
PROPr	0.35 (0.21)

NOTE. Higher scores mean better physical function, ability to participate in social roles and activities, and cognitive function. Higher scores mean better health on the physical health and mental health summary scores and on the PROPr. Higher scores on the other measures indicate worse health. The PROPr is scored so that 0 is dead or as bad as being dead and 1 is perfect health. The general population mean of the PROPr=0.52 . The other scales are scored on a T score metric with a mean of 50 and SD of 10 in the US general population.

Table 3 Product-moment correlations of the PEG with the PROMIS-29+2 v 2.1 scales and the Impact Stratification Score (n=1931)

PROMIS-29+2 Measures	PEG
Impact Stratification Score (ISS)	0.74
Pain intensity	0.70
Pain interference	0.68
Physical health summary score	−0.62
PROPr	−0.59
Mental health summary score	−0.58
Physical function	−0.57
Ability to participate in social roles and activities	−0.56
Fatigue	0.43
Anxiety	0.42
Depression	0.39
Sleep disturbance	0.28
Cognitive function	−0.27

NOTE. Higher scores mean better physical function, ability to participate in social roles and activities, cognitive function, physical health summary score, mental health summary score, and PROPr. Higher scores on the other measures indicate worse health.

The ISS accounted for 55% of the variance (adjusted R^2) in the PEG (NMAE=0.53) and the regression equation was PEG =−0.043982+0.19620 × ISS. The PEG accounted for 55% of the variance (adjusted R^2) in the ISS (NMAE=0.52) and the regression equation was ISS=9.33684+2.82342 × PEG. The correlations between the PEG and ISS were similar for women (r=0.75) and men (r=0.73) and between those with more than high school education (r=0.74) vs high school or less (r=0.77). But the correlation differed for Hispanics (r=0.60) vs non-Hispanics (r=0.76) and those who were non-White (r=0.68) vs those who were White (r=0.77).

Discussion

That the ISS had the strongest correlation of all the PROMIS-29 v2.1 measures with the PEG provides further support for the value of the pain impact measure recommended by the NIH Pain Consortium's RTF on chronic low back pain. The study provides useful preliminary crosswalks between the PEG and ISS using regression to predict one from the other. The NMAE estimates indicate that the average deviation between the observed and predicted scores is about a half of a standard deviation. The NMAE values are like those obtained in a previous study predicting the Oswestry Disability Index from the PROMIS-29 physical function, pain interference, and pain intensity measures that are used to create the ISS.[13]

Study limitations

This study has limitations. The correlation between the ISS and the PEG of 0.74 was below the threshold to use item response theory equating. Moreover, the MTurk sample from which the respondents with back pain were selected differs in pain, mental health, age, education, and income from that of the US general population.[15] Hilton et al[16] found that an MTurk sample was more likely to report chronic low back pain than a clinic-based sample but less average and worst pain and lower Oswestry Disability Index scores (indicating less disability). It is unclear how well the back pain subgroup focused on in this study represents adults with back pain in general. The results of this study are based on only 1 sample and results may vary in other samples. Finally, the correlation between the PEG and the ISS was smaller for Hispanics and those who were non-White compared to non-Hispanics and those who were White. Hence, there is less accuracy in prediction from the crosswalks among these subgroups of the population.

Conclusions

Researchers can use the regression equations reported here to estimate one score (PEG or ISS) from the other. Regression to the mean can be accounted for using linear equating[17] that adjusts the regression model predictions to have the same means and standard deviation as the observed dependent variable scores (see Appendix). These estimates can be used to facilitate comparisons across interventions and enhance interpretation of study results.

The prediction equations can be used for group-level comparisons, but there is too much error for use in estimating individuals' scores. Further studies are needed to evaluate the generalizability of the prediction equations derived in this study given the unknown representativeness of the sample of adults with back pain to the overall population.

Keywords

Pain measurement; Patient reported outcome measures; Quality of life

Corresponding author

Ron D. Hays, PhD, Division of General Internal Medicine and Health Services Research, UCLA Department of Medicine, 1100 Glendon Avenue, Suite 850, Los Angeles 90024, CA. *E-mail address:* drhays@ucla.edu.

Appendix: Summary of Linear Equating

Linear equated prediction = a + (b/c) × (d − e)

a=observed dependent variable mean in the data set used in this article.
b=observed dependent variable standard deviation in the data used in this article.
c=predicted standard deviation in another data set.
d=predicted score for each individual in another data set.
e=predicted mean score in another data set.

Note: ISS a=20.68048 and b=8.06007 and PEG a=4.01769 and b=2.12471. Predicted scores <0 for PEG and <8 for ISS should be recoded to minimum possible score, and scores >10 for PEG and >50 for ISS recoded to the maximum possible score.

References

1. Herman PM, Edelen MO, Rodriquez A, Hilton LG, Hays RD. A protocol for chronic pain outcome measurement enhancement by linking PROMIS-29 scale to legacy measures and improving chronic pain stratification. BMC Musculoskelet Disord 2020;21:671.
2. Deyo RA, Dworkin SF, Amtmann D, et al. Focus article: report of the NIH Task Force on Research Standards for Chronic Low Back Pain. Eur Spine J 2014;23:2028–45.
3. Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on Research Standards for Chronic Low Back Pain. Spine 2014;39:1128–43.
4. Hays RD, Edelen MO, Rodriquez A, Herman P. Support for the reliability and validity of the National Institutes of Health Impact Stratification Score in a sample of active-duty US military personnel with low back pain. Pain Med 2021;22:2185–90.
5. Hays RD, Peipert J. Between-group minimally important change versus individual treatment responders. Qual Life Res 2021;30:2765–72.
6. Krebs EE, Lorenz KA, Bair MJ, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med 2009;24:733–8.
7. Kean J, Monahan PO, Kroenke K, et al. Comparative Responsiveness of the PROMIS Pain Interference Short Forms, Brief Pain Inventory, PEG, and SF-36 Bodily Pain Subscale. Med Care 2016;54:414–21.
8. Cella D, Choi SW, Condon DM, et al. PROMIS® adult health profiles: efficient short-form measures of seven health domains. Value Health 2019;22:537–44.
9. Hays RD, Spritzer KL, Schalet BD, Cella D. PROMS®-29 v2.0 profile physical and mental health summary scores. Qual Life Res 2018;27:1885–91.
10. DeWitt B, Feeny D, Fischoff B, et al. Estimation of a preference-based summary score for the Patient-Reported Outcomes Measurement Information System: the PROMIS®-Preference (PROPr) Scoring System. Medical Decision Making 2018;38:683–98.

11. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.
12. Dorans NJ. Equating, concordance, and expectation. Appl Psychol Meas 2004;28:227–46.
13. Edelen MO, Rodriguez A, Herman P, Hays RD. Crosswalking the Patient-Reported Outcomes Measurement Information System physical function, pain interference, and pain intensity scores to the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. Arch Phys Med Rehabil 2021;102:1317–23.
14. Cohen JA. A power primer. Psychol Bull 1992;112:155–9.
15. Qureshi N, Edelen M, Hilton L, Rodriquez A, Hays RD, Herman PM. Comparison of data collected using Amazon's Mechanical Turk to national surveys. Am J Health Behav. In Press.
16. Hilton LG, Coulter ID, Ryan G, Hays RD. Comparing the recruitment of research participants with chronic low back pain using Amazon Mechanical Turk versus recruitment of patients from chiropractic clinics: a quasi-experimental study. J Manipulative Physiol Ther 2021;21:601–11.
17. Fayers PM, Hays RD. Should linking replace regression when mapping from profile to preference-based measures? Value Health 2014;17:261–5.