Title

Crosswalking Four Pain Impact Measures in a Nationally Representative Sample of Adults with Back Pain

Authors

DiGuiseppi, Graham T
Edelen, Maria Orlando
Rodriguez, Anthony
et al.

Publication Date

2024-09-01

DOI

10.1016/j.apmr.2024.09.007

Peer reviewed

ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2024;000: 1−7

ORIGINAL RESEARCH

Crosswalking 4 Pain Impact Measures in a Nationally Representative Sample of Adults With Back Pain

Graham T. DiGuiseppi, PhD,[a] Maria Orlando Edelen, PhD,[b,c] Anthony Rodriguez, PhD,[b] Mary Slaughter, PhD,[d] Ron D. Hays, PhD,[e] Chengbo Zeng, PhD,[c] Ian D. Coulter, PhD,[d] Patricia M. Herman, PhD[d]

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

Abstract

Objective: To generate crosswalk equations and tables for 4 pain impact measures: the Impact Stratification Score (ISS), Oswestry Disability Index (ODI), Roland-Morris Disability Questionnaire (RMDQ), and the Pain, Enjoyment of Life and General Activity Scale (PEG).

Design: Cross-sectional survey assessing demographics and pain impact. Crosswalks were developed using item-response theory (IRT) cocalibrations and linear regressions between the ISS, ODI, RMDQ, and PEG.

Setting: Online panel.

Participants: Population-based sample of United States adults aged 18 and older. Eligibility criteria were reporting current back pain, not reporting 2 fake health conditions, and having data for 2 or more pain measures (N=1530; 37% of sample). Crosswalks were developed (n=1030) and cross-validated in a subsample of 500 participants (n=125 randomly sampled from each ISS quartile).

Interventions: Not applicable.

Main Outcome Measures: ISS, ODI, RMDQ, and the PEG.

Results: Associations of the ISS with the PEG and ODI met the criteria for IRT cocalibration. Other measure pairs were crosswalked using regression. Associations were strongest between the PEG and the ISS (r=0.87, normalized mean absolute error [NMAE]=0.38) and between the ODI and the ISS (r=0.85, NMAE=0.39). Associations were weakest between the PEG and the RMDQ (r=0.69, R^2=0.48, NMAE: 0.55-0.58). Regression equations and IRT accounted for 48%-64% of the variance (NMAE: 0.38-0.58) in corresponding pain measures in the cross-validation sample.

Conclusions: The crosswalks between the ISS and common legacy pain measures created in this study of a nationally representative sample of United States adults with back pain can be used to estimate 1 pain impact measure from another. Further evaluation in clinical samples is recommended.

Archives of Physical Medicine and Rehabilitation 2024;000:1−7

An estimated 39% of U.S. adults suffer from back pain,[1] and 20.5 million report chronic back pain—ie, lasting 3 months or longer.[2] Chronic pain is a leading cause of disability and is associated with poor quality of life and excessive healthcare expenditures.[2,3] Pain and its impact are often assessed using patient-reported outcome measures (PROMs).[4] In connection with developing standards for research on chronic low back pain (cLBP), a National Institutes of Health (NIH) research task force recommended using the Impact Stratification Score (ISS) to classify pain impact based on pain intensity, pain interference with activities, and physical function.[5] ISS items are a subset of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) 29-item profile measure (PROMIS-29).[6] However,

This work was funded by the National Center for Complementary and Integrative Health (grant number R01AT010402, Principal Investigator: Patricia Herman).

Disclosures: none.

0003-9993/$36 - see front matter © 2024 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.2024.09.007

measures such as the Oswestry Disability Index (ODI)[7-9] and the Roland-Morris Disability Questionnaire (RMDQ)[10] are commonly used to assess pain among back pain patients.[11] The variety of PROMs measured on different scales hinders the ability to assess the epidemiology of back pain, compare the effectiveness of interventions across studies, and pool results in meta-analyses.

To address this problem, recent efforts have been made to "link" (or crosswalk) PROMs of pain impact and other outcomes, allowing researchers to compare results across measures, samples, and settings.[12,13] Two common approaches include creating crosswalk tables to equate 2 measures, and using linear regression equations to produce expected scores.[14] Askew et al[15] created a crosswalk table to link pain interference scores from the Brief Pain Inventory Pain Interference Scale (BPI-PI) with the PROMIS Pain Interference (PROMIS-PI) in a community-dwelling sample of people with multiple sclerosis (MS) and cross-validated with MS participants in a study about aging with physical disabilities. The crosswalk produced predictions of pain interference that were invariant across subgroups (eg, gender, race, age).[15] Cook et al[16] linked 2 legacy measures (the BPI-PI and the Short Form 36 [SF-36] Bodily Pain Subscale) with the PROMIS-PI in a general population sample, and found predictive accuracy results consistent with those of Askew et al.[15] More recently, Tang et al[17] linked the ODI to the PROMIS-PI in an orthopedic back pain population, and Edelen et al[18] crosswalked PROMIS physical function, pain interference, and pain intensity scores with the RMDQ and ODI in 3 samples of adults with cLBP.[18]

To our knowledge, only 1 study has linked the ISS composite score recommended by the NIH Research Task Force on cLBP with an existing legacy measure. Hays et al[19] crosswalked the ISS with the Pain, Enjoyment of Life and General Activity Scale (PEG) using panel data from Amazon's Mechanical Turk (MTurk), finding that the ISS and PEG shared 55% of the variance. Furthermore, correlations were similar across subgroups, with some exceptions for race/ethnicity. Because the ISS is a recommended measure for cLBP, the current study crosswalked the ISS with 3 popular legacy PROMs, the ODI, RMDQ, and the PEG, and crosswalked each with each other. Crosswalks were developed and applied in a nationally representative sample of U.S. adults endorsing current back pain, extending prior efforts to link the ISS with existing legacy measures in a convenience sample.[19]

List of abbreviations:

BPI-PI	**Brief Pain Inventory Pain Interference Scale**
cLBP	**chronic low back pain**
ECV	**explained common variance**
IRT	**item-response theory**
ISS	**Impact Stratification Score**
NIH	**National Institutes of Health**
NMAE	**normalized mean absolute error**
ODI	**Oswestry Disability Index**
OLS	**ordinary least squares**
PEG	**Pain, Enjoyment of Life and General Activity Scale**
PROM	**patient-reported outcome measures**
PROMIS	**Patient-Reported Outcomes Measurement Information System**
PROMIS-29	**29-Item PROMIS Profile Measure**
PROMIS-PI	**PROMIS Pain Interference**
RMDQ	**Roland-Morris Disability Questionnaire**
SF-36	**Short Form 36**
U.S.	**United States**

Methods

Data sources

Data were collected from Ipsos' KnowledgePanel, the largest nationally representative online panel in the United States (U.S.)[20] Ipsos' probability-based sampling methodology uses address-based sampling from up-to-date delivery sequence files from the U.S. Postal Service. A random sample of households is sent invitations to join via mail and is followed up with mail contacts and phone calls. KnowledgePanel is comprised of over 55,000 members, with new members recruited quarterly to make up for attrition. All surveys are conducted online with respondents' informed consent. Randomly sampled panel members receive a modest incentive for participating in surveys for which they are eligible; on average, completing 2 to 3 surveys per month.

Data for the current study come from a larger study, approved by the RAND Human Subjects Protection Committee, 1 component of which analyzed data from 4149 KnowledgePanel members to evaluate PROMIS and legacy pain measures.[21] The analysis was restricted to 1533 individuals who: (1) endorsed "back pain" in response to an item asking "Do you currently have…?" followed by a list of health conditions; and (2) did not endorse either of 2 fake conditions ("chekalism" and "syndomitis") included in the list as attention checks.[22] Cases were further restricted to those that had complete data for at least 2 of the 4 pain measures (3 participants were excluded). The final analytic sample of 1530 individuals was primarily non-Hispanic White (73%), almost half (45%) were aged 60 years or older, and 52% identified as female.

For purposes of validating the empirical crosswalks, we randomly selected 125 cases from each quartile of the ISS, corresponding to the following observed scores: 8-13 (first quartile), 14-18 (second quartile), 19-25 (third quartile), and 26-49 (fourth quartile). This resulted in a "holdout" cross-validation sample of 500 respondents. The remaining KnowledgePanel sample (n = 1030) was retained for developing the crosswalks. There were no significant demographic differences between the development and validation samples (χ^2 tests of independence, all P values>.05). Results of 2-tailed independent samples t tests also revealed no significant differences between pain measures across the 2 samples (see appendix 1), with sufficient power (96%) to detect small differences between groups. Additional sample characteristics are displayed in table 1.

Measures

Impact stratification score

The ISS uses 9 PROMIS-29 items. Four items assess pain interference (eg, "How much did pain interfere with your day-to-day activities?"), with response items ranging from 1 ("Not at all") to 5 ("Very much"). Four items assess physical function (eg, "Are you able to go for a walk for at least 15 minutes?"); response options range from 1 ("Without any difficulty") to 5 ("Unable to do"). One item assesses pain intensity ("In the past 7 days, how would you rate your pain on average?"), with response options ranging from 0 ("no pain") to 10 ("worst imaginable pain"). The

Table 1 KnowledgePanel sample descriptive statistics

Variable	Complete Sample	Development Sample	Validation Sample
n	1530	1030	500
Gender			
Female	801 (52.5%)	545 (53.2%)	256 (51.2%)
Male	715 (46.9%)	476 (46.4%)	239 (47.8%)
Transgender/other identity	9 (0.6%)	4 (0.4%)	5 (0.8%)
Race/ethnicity			
White, non-Hispanic	1,122 (73.3%)	760 (73.8%)	362 (72.4%)
Black, non-Hispanic	125 (8.2%)	89 (8.6%)	36 (7.2%)
Multiracial, non-Hispanic	68 (4.4%)	42 (4.1%)	26 (5.2%)
Other, non-Hispanic	56 (3.7%)	33 (3.2%)	23 (4.6%)
Hispanic	159 (10.4%)	106 (10.3%)	53 (10.6%)
Age, y			
18-29	162 (10.6%)	110 (10.7%)	52 (10.4%)
30-44	317 (20.7%)	201 (19.5%)	116 (23.2%)
45-59	364 (23.8%)	253 (24.6%)	111 (22.2%)
≥60	687 (44.9%)	466 (45.2%)	221 (44.2%)
Education			
No high school diploma or GED	114 (7.5%)	81 (7.9%)	33 (6.6%)
High school or GED graduate	440 (28.8%)	304 (29.5%)	136 (27.2%)
Some college or Associate's degree	444 (29.0%)	286 (27.8%)	158 (31.6%)
Bachelor's degree	288 (18.8%)	199 (19.3%)	89 (17.8%)
Master's degree or higher	244 (16.0%)	160 (15.5%)	84 (16.8%)
Household income			
<$10,000	67 (4.4%)	47 (4.6%)	20 (4.0%)
$10,000-$49,999	443 (29.0%)	296 (28.7%)	147 (29.4%)
$50,000-$99,999	471 (30.8%)	317 (30.8%)	154 (30.8%)
$100,000 or more	549 (35.9%)	370 (35.9%)	179 (35.8%)
Pain measures, mean (SD)			
ISS	20.1 (9.4)	19.8 (9.5)	20.7 (9.3)
ODI	19.9 (16.9)	19.6 (16.9)	20.4 (16.8)
RMDQ	6.3 (6.7)	6.2 (6.6)	6.6 (6.8)
PEG	3.3 (2.5)	3.2 (2.5)	3.5 (2.5)

NOTE. Cases with missing data: *n*=5 (gender); *n*=31 (ISS); *n*=3 (ODI); *n*=9 (RMDQ). Results of χ^2 tests of independence and 2-tailed independent samples *t* tests revealed no statistically significant differences between the development and validation samples (all *P* values>.05).
Abbreviations: GED, General Equivalency Degree.

ISS results in a summative score ranging from 8 (least impact) to 50 (greatest impact).[5]

Oswestry Disability Index

The ODI is a measure of functional disability consisting of 10 items assessing pain intensity, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and traveling.[7] Response options range from 0 (eg, "I have no pain at the moment") to 5 (eg, "The pain is the worst imaginable at the moment"), with higher scores indicating greater disability. The scale is scored by summing all items, dividing by the maximum possible score, and then multiplying by 100. The ODI score can be classified as 0-20 (minimal disability), 21-40 (moderate), 41-60 (severe), 61-80 (crippled), and 81-100 (bedbound or exaggerating symptoms).[7]

Roland-Morris Disability Questionnaire

The RMDQ is a 24-item measure assessing physical disability due to lower back pain.[10,23] The items were chosen from the Sickness Impact Profile (SIP)[24] and reflect statements about physical function and impairment (eg, "I stay at home most of the time because of the pain in my back"). Response options are dichotomous (0 = does not describe me today; 1 = describes me today) and yield a total score ranging from 0 (no disability) to 24 (maximum disability).

Pain, Enjoyment of Life, and General Activity Scale

The PEG is a brief, 3-item subset of the Brief Pain Inventory. Support for its reliability and construct validity was reported in a sample of primary care and ambulatory clinic patients.[25] PEG items contain a single item assessing pain intensity, from 0 ("no pain") to 10 ("pain is as bad as you can imagine"); and 2 items assessing interference with enjoyment of life and interference with general activities, ranging from 0 ("does not interfere") to 10 ("completely interferes"). A PEG score is obtained by taking the average of the 3 items.

Data analysis

Pearson product-moment correlations between the ISS, ODI, RMDQ, and PEG were estimated. Following guidance from Dorans[14] on linking method selection according to the strength of the observed correlation between measures, for PROM pairs with correlations ≤0.80 in the development sample, ordinary least squares

(OLS) regression was used.[14,18] For PROM pairs with correlations >0.80 in the development sample, item-response theory (IRT) cocalibration was used. For the OLS crosswalk solutions, bivariate regressions, R^2, and the normalized mean absolute error (NMAE) were estimated in the development sample. Regression equations are provided for researchers to predict scores from 1 measure to another. For IRT cocalibration, we first ran bifactor models to establish "essential unidimensionality" of the combined item sets based on the explained common variance (ECV) and omegaH.[26,27] We used Samejima's[28] graded response model to cocalibrate scales, using the parameters for the ISS to set the scale, and used derived summed score to IRT-score conversion tables for each measure to establish a score-to-score linkage.[29] The number of respondents with missing data was minimal: 3 (ODI), 5 (gender), 9 (RMDQ), 31 (ISS). All bivariate analyses used pairwise deletion.

To evaluate crosswalk performance, regression equations and IRT-based translation tables derived from the development sample were used to generate corresponding scores among pain impact measures in the validation sample, with R^2 (for OLS solutions only) and NMAE used to evaluate predictive accuracy. NMAE, calculated by taking the average of the absolute value of the residuals (observed−predicted scores) divided by the standard deviation of the observed scores,[18,19] allows for a common metric to be compared between scales, with lower NMAE indicating better prediction. A cutoff of 0.50 or less has been used as an acceptable level of NMAE.[18] Finally, predictive performance was compared between demographic subgroups by stratifying the NMAE for each of the crosswalks in the validation sample and calculating absolute differences by gender (female vs male), race/ethnicity (non-Hispanic White vs non-White and/or Hispanic), and age (18-59 vs 60 and older). Positive values indicate higher NMAE in the first group and negative values indicate higher NMAE in the second group.

Results

Bivariate correlations among the pain measures in the 2 samples are displayed in table 2. The strongest correlations were observed between the PEG and the ISS (r_{range}: 0.86-0.87) and between the ODI and the ISS (r_{range}: 0.84-0.85). Correlations were not as strong between the PEG and RMDQ (r_{range}: 0.69-0.70). Correlations between the RMDQ and the ISS, and the RMDQ and ODI were slightly stronger in the validation sample (r_{range}: 0.77-0.80) than the development sample (r_{range}: 0.74-0.77), although these differences were minimal. Based on these bivariate correlations, OLS regression was performed for all crosswalks involving the RMDQ as well as the crosswalk of the PEG with the ODI. IRT cocalibration was performed to crosswalk the ISS with the PEG and the ODI. Bifactor models supported the essential unidimensionality of these 2 items sets for IRT cocalibration (ISS and PEG: ECV=0.77, omegaH=0.80; ISS and ODI: ECV=0.74, omegaH=0.83).

Table 2 Correlations among pain measures in development and validation samples

Pain Measure	Sample	ISS	ODI	RMDQ	PEG
ISS	KP development				
	KP validation				
ODI	KP development	0.85			
	KP validation	0.84			
RMDQ	KP development	0.74	0.77		
	KP validation	0.77	0.80		
PEG	KP development	0.87	0.75	0.69	
	KP validation	0.86	0.75	0.70	

NOTE. $P<.001$ for all bivariate correlations.

Development of OLS crosswalk equations in the development sample

Regression equations and performance metrics (R^2 and NMAE) of OLS crosswalks in the development sample are shown in columns 2-4 of table 3. All correlations and OLS regressions were statistically significant ($P<.001$). Consistent with correlations, a greater amount of variance was explained, and lower error was found in regressions between the RMDQ and the ODI (R^2=0.59, NMAE: 0.45-0.47), the PEG and the ODI (R^2=0.56, NMAE: 0.49-0.52), and the RMDQ and the ISS (R^2=0.54, NMAE: 0.50−0.51). This is compared to regressions involving the PEG and the RMDQ, which accounted for the lowest amount of variance explained and the greatest amount of error (R^2=0.48, NMAE: 0.53-0.57).

Performance of OLS and IRT crosswalks in the validation sample

The performance of crosswalks in the validation sample is shown in the 2 rightmost columns of table 3. The 8 crosswalks derived from OLS regression accounted for between 48%-64% of the variance in the validation sample (NMAE range: 0.44-0.58). As in the development sample, OLS regression crosswalks between the RMDQ and the ODI performed best, accounting for 63%-64% of the variance in each other (NMAE=0.44). OLS regression crosswalks did not perform as well when crosswalking the PEG to the RMDQ (R^2=0.49, NMAE=0.55) or the RMDQ to the PEG (R^2=0.48, NMAE=0.58). As shown in the bottom 4 rows of table 3, IRT-based predictions between the PEG and the ISS (NMAE: 0.38−0.39) and between the ODI and the ISS (NMAE: 0.42) had the lowest amount of error of all the crosswalks generated. This is consistent with the strong bivariate associations between these measures in the development sample. Score translation tables generated from IRT cocalibrations of the ISS with the PEG and ODI are displayed in appendices 2 and 3.

Crosswalk performance across demographic subgroups

The relative performance of crosswalks, when compared across demographic subgroups in the validation sample, is shown in table 4 (see appendix 4 for all NMAE values). For all comparisons, larger absolute values indicate greater differences in prediction between groups and can be interpreted as effect size metrics. In general, subgroup differences tended to be small across the 12 crosswalks and 3 grouping variables, although there were a few with absolute values of 0.10 or larger (2 for the race/ethnicity comparison and 2 for the age comparison). Crosswalks performed better for males than females, wherein all but one NMAE absolute difference were positive ($NMAE_{range:}$ −0.01 to 0.06). Crosswalks also tended to perform better for White individuals versus non-White and Hispanic individuals ($NMAE_{range:}$ −0.15 to −0.01) and better for individuals aged 18-59 versus individuals aged 60 and

Table 3 Predictive performance of OLS regression equations and IRT cocalibration

	KP Development (n=1030)			KP Validation (n=500)	
Crosswalk	Regression Equation	R^2	NMAE	R^2	NMAE
OLS regression					
RMDQ → ODI	7.43+1.96×RMDQ	0.59	0.47	0.64	0.44
ODI → RMDQ	0.28+0.30×ODI	0.59	0.45	0.63	0.44
PEG → ODI	3.11+5.13×PEG	0.56	0.49	0.56	0.51
ODI → PEG	1.09+0.11×ODI	0.56	0.52	0.56	0.53
RMDQ → ISS	13.31+1.05×RMDQ	0.54	0.51	0.58	0.50
ISS → RMDQ	−4.01+0.52×ISS	0.54	0.50	0.58	0.49
PEG → RMDQ	0.20+1.87×PEG	0.48	0.53	0.49	0.55
RMDQ → PEG	1.63+0.26×RMDQ	0.48	0.57	0.48	0.58
IRT Cocalibration					
PEG → ISS	n/a	n/a	n/a	n/a	0.38
ISS → PEG	n/a	n/a	n/a	n/a	0.39
ODI → ISS	n/a	n/a	n/a	n/a	0.42
ISS → ODI	n/a	n/a	n/a	n/a	0.42

NOTE. independent variable → dependent variable.

older, wherein all but 2 NMAE absolute difference values were negative ($NMAE_{range}$: −0.11 to 0.06).

Discussion

The present study crosswalked the ISS with 3 popular legacy measures of pain impact, and the 3 legacy measures with one another in a nationally representative sample of U.S. adults reporting current back pain, and results were cross-validated in a holdout sample of participants stratified by empirical ISS scores. In the development and validation samples, the ISS was most strongly associated with scores on the ODI and the PEG, and vice versa. Associations between the RMDQ and the PEG were the weakest and accounted for the least amount (slightly less than half) of the variance in each other in the validation sample. Overall, most crosswalks accounted for over half of the variance in respective pain measures, but this study shows that some pain measures may be more easily translated than others.

In previous work by Edelen et al,[18] 13%-56% of the variance in PROMIS physical function, pain interference, and pain intensity measures were explained by the ODI and RMDQ, with performance being relatively lower when predicting pain intensity. Although not directly comparable, the ODI and the RMDQ performed as well or better in predicting ISS composite scores in the present study. Edelen et al[18(p1323)] concluded that the weak performance of some crosswalks "may be due in part to predicting a single PROMIS score (eg, physical function) from multidimensional composites." We believe that multidimensionality may also play a role in the relatively weaker performance of the RMDQ in predicting other PROMs scores in this study. The RMDQ is distinct from other measures, as it is a count of the number of functional impairments due to low back pain, whereas the item content of the ISS, ODI, and PEG are broader in scope, including items, for example, assessing pain intensity.

Table 4 Subgroup comparisons of crosswalk performance (expressed as differences in normalized mean absolute error)

	Gender	Race/Ethnicity	Age
Crosswalk	Female vs Male	White vs Non-White/Hispanic	18-59 y vs 60+ y
OLS regression			
RMDQ → ODI	0.00	−0.06	−0.05
ODI → RMDQ	0.06	−0.04	−0.01
PEG → ODI	−0.01	−0.15	−0.02
ODI → PEG	0.05	−0.06	−0.01
RMDQ → ISS	0.01	−0.03	−0.02
ISS → RMDQ	0.00	−0.07	0.01
PEG → RMDQ	0.00	−0.08	0.06
RMDQ → PEG	0.03	−0.07	−0.02
IRT cocalibration			
PEG → ISS	0.02	−0.01	−0.11
ISS → PEG	0.02	−0.01	−0.10
ODI → ISS	0.04	−0.08	−0.07
ISS → ODI	0.02	−0.11	−0.07

NOTE. Values are differences in the NMAE. Higher absolute values indicate greater difference in NMAE between the 2 groups. NMAE differences were calculated for: gender (female vs male), race/ethnicity (White vs non-White/Hispanic), and age (18-59 years vs 60 years and older).

When compared to another recent study by Hays et al,[19] our results also showed better predictive performance of crosswalks between the ISS and the PEG (NMAE: 0.38-0.39). These different results may be due to sample composition. In the study by Hays,[19] the ISS and PEG shared 55% of the variance (NMAE=0.53) in a convenience sample of Amazon MTurk workers, who tend to be younger, more educated, and report a higher prevalence of depression, anxiety, and chronic back pain than nationally representative samples (Herman et al., unpublished data, 2024). Similar to our study, Hays et al[19] showed that correlations between the ISS and the PEG differed by race and ethnicity, with correlations between measures being weaker for non-White individuals (vs White individuals), and weaker for Hispanic individuals (vs non-Hispanic individuals). In contrast, we found differences in crosswalk performance by gender (better performance among males), but these differences were small in magnitude. In addition, we found that crosswalks performed better for younger individuals in our study (vs those aged 60 years and older), particularly between the PEG and the ISS. This suggests the continued need to evaluate the performance of linking PROMs for cLBP across demographic subgroups.

Overall, the observed and predicted associations between measures in the present study suggest that the ODI and the PEG may translate well to the ISS, although more research may be needed in diverse clinical back pain populations. Given the strong relationships between these measures, this study provides support for the use of the ISS as an alternative to legacy pain impact measures. This is in line with recommendations from an NIH task force on research standards for cLBP.[5] By classifying back pain by its "impact" (pain intensity, interference with activities, and physical functioning), the ISS may help standardize the measurement of back pain and cLBP in future research and clinical practice.[5]

Study limitations

Although this study sought to develop and cross-validate crosswalks in a nationally representative sample of U.S. adults, the ability to generalize these results to clinical samples of patients with back pain (and cLBP) deserves further study. We also cross-validated crosswalks in a random subsample of KnowledgePanel participants that were similar in composition to the development sample. Although use of "holdout" samples is a common method of cross-validation,[30] future research may benefit by cross-validating in other diverse samples.[31]

Conclusions

We provide crosswalk equations and IRT-derived crosswalk tables to link scores between a PROM recommended by an NIH Research Task Force on cLBP (the ISS) and 3 legacy measures (the ODI, RMDQ, and PEG). These equations and crosswalk tables may be used to translate scores across PROMs, estimate effect sizes in intervention studies, and to pool data for meta-analyses, although further validation is needed in diverse clinical samples. This work can help advance understanding of the relationship between common pain impact measures and aid the selection of measures for clinical and research purposes.

Keywords

Back pain measures; Back pain; crosswalks; Impact Stratification Score; IRT; linking; Oswestry Disability Index; Pain, Enjoyment of Life and General Activity Scale (PEG); Roland-Morris Disability Questionnaire

Corresponding author

Graham T. DiGuiseppi, PhD, RAND, 4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213. *E-mail address:* gdiguise@rand.org.

References

1. Lucas J, Connor E, Back Bose J. lower limb, and upper limb pain among U.S. adults, 2019. NCHS Data Brief 2021;415:1–8.
2. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain 2022;163:e328–32.
3. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021;397:2082–97.
4. Pogatzki-Zahn E, Schnabel K, Kaiser U. Patient-reported outcome measures for acute and chronic pain: current knowledge and future directions. Curr Opin Anaesthesiol 2019;32:616–22.
5. Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014;15:569–85.
6. 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.
7. Fairbank JC, Pynsent PB. The Oswestry disability index. Spine 2000;25:2940–52. discussion 2952.
8. Chen CX, Kroenke K, Stump T, et al. Comparative responsiveness of the PROMIS pain interference short forms with legacy pain measures: results from three randomized clinical trials. J Pain 2019;20:664–75.
9. Yee TJ, Smith BW, Joseph JR, et al. Correlation between the Oswestry Disability Index and the 4-item short forms for physical function and pain interference from PROMIS. J Neurosurg Spine 2019;31:691–6.
10. Roland MO, Morris RW. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8:141–4.
11. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. Spine 2011;36(Suppl 21):S54–68.
12. Schalet BD, Lim S, Cella D, Choi SW. Linking Scores with Patient-Reported Health Outcome Instruments: a validation study and comparison of three linking methods. Psychometrika 2021;86:717–46.
13. Choi SW, Lim S, Schalet BD, Kaat AJ, Cella D. PROsetta: an R package for linking patient-reported outcome measures. Appl Psychol Meas 2021;45:386–8.
14. Dorans NJ. Equating, concordance, and expectation. Appl Psychol Meas 2004;28:227–46.
15. Askew RL, Kim J, Chung H, Cook KF, Johnson KL, Amtmann D. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. Qual Life Res 2013;22:2769–76.
16. Cook KF, Schalet BD, Kallen MA, Rutsohn JP, Cella D. Establishing a common metric for self-reported pain: linking BPI Pain Interference and SF-36 Bodily Pain Subscale scores to the PROMIS Pain Interference metric. 2015;24(10):2305-18. doi:10.1007/s11136-015-0987-6.
17. Tang X, Schalet BD, Hung M, Brodke DS, Saltzman CL, Cella D. Linking Oswestry Disability Index to the PROMIS pain interference CAT with equipercentile methods. Spine J 2021;21:1185–92.
18. 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.
19. Hays RD, Qureshi N, Edelen M, Rodriguez A, Slaughter M, Herman PM. Crosswalking the National Institutes of Health impact stratification score to the PEG. Arch Phys Med Rehabil 2023;104:425–9.
20. Ipsos. KnowledgePanel. https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel. Accessed July 20, 2023.

21. Herman PM, Edelen MO, Rodriguez 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 Musculoskel Disord 2020; 21:671.
22. Hays RD, Qureshi N, Herman PM, Rodriguez A, Kapteyn A, Edelen MO. Effects of excluding those who report having "Syndomitis" or "Chekalism" on data quality: longitudinal health survey of a sample from Amazon's mechanical Turk. J Med Internet Res 2023;25:e46421.
23. Roland M, Fairbank J. The Roland-Morris disability questionnaire and the Oswestry disability questionnaire. Spine 2000;25:3115–24.
24. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981;19:787–805.
25. 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.
26. Rodriguez A, Reise SP, Haviland MG. Applying bifactor statistical indices in the evaluation of psychological measures. J Pers Assess 2016;98:223–37.
27. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: calculating and interpreting statistical indices. Psychol Methods 2016;21:137–50.
28. 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.
29. Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: application to depression measurement. Psychol Assess 2000;12:354–9.
30. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv 2010;4:40–79.
31. Tang X, Schalet BD, Janulis P, et al. Can a linking crosswalk table be applied to a different population? An independent validation study for a crosswalk between BSI depression and PROMIS depression scales. PLOS ONE 2022;17:e0278232.