Evaluating the National Institutes of Health Impact Stratification Score in a Sample of Active Duty U.S. Military Personnel with Low Back Pain

Ron D. Hays, Maria Orlando Edelen, Anthony Rodriquez, Patricia Herman

Hays, R. D., Edelen, M. O., Rodriquez, A., & Herman, P. (2020). Evaluating the National Institutes of Health Impact Stratification Score in a sample of active duty U.S. military personnel with low back pain. Preprint at https://labs.dgsom.ucla.edu/hays/pages/favorite_links

The National Institutes of Health (NIH) Pain Consortium established a steering committee and convened a research task force that proposed an Impact Stratification Score (ISS) consisting of the 9 items in the PROMIS-29 v2.0 physical function, pain interference and pain intensity measures (Deyo, Dworkin et al., 2014). Physical function (4 items that that each go from *without any difficulty* = 1 to *unable to do* = 5) and pain interference (4 items that each go from *not at all* = 1 to *very much* = 5) each contribute from 4 to 20 points, and pain intensity contributes from 0-10 points. The ISS has a possible range of 8 (least impact) to 50 (greatest impact). The task force suggested three categories of severity for the ISS: 8-27 (mild), 28-34 (moderate), and 35-50 (severe).

The ISS has received limited evaluation to date. Internal consistency reliability (coefficient alpha) of 0.91, an intraclass correlation coefficient of 0.73 over 3 months among those who reported that their pain was "about the same" at a 3-month follow-up assessment (test-retest reliability), and the ISS at 3 months was monotonically associated with patient reports of how much worse their pain was at 3 months compared to baseline in a study of 198 patients with chronic musculoskeletal pain (Deyo et al. 2015). Cronbach's alpha and kappa were similar for those with chronic low back pain and those with other musculoskeletal pain conditions. In a sample of 218 patients with low back pain who received epidural steroid injections, Spearman's rho of the ISS with the Roland-Morris Disability Questionnaire (RMDQ) and the Oswestry

Disability Index was 0.81 and 0.66, respectively, and the ISS was more responsive to change than was the RMDQ to epidural steroid injections (Deyo et al., 2014).

This paper expands upon prior work to assess the reliability and validity of the ISS using data collected in a prospective, multisite, parallel-group comparison effectiveness clinical trial of active duty U.S. military personnel (Goertz et al., 2016, 2018).

Methods

Sample

The study was conducted at one small hospital at a military training site (Naval Hospital in Pensacola, Florida) and two large military medical centers in major metropolitan areas: 1) Walter Reed National Military Medical Center in Bethesda, Maryland; and 2) Naval Medical Center in San Diego, California. The sample characteristics were reported in Goertz et al. (2018). Average age was 31; 76% were males and 67% white. Most of the participants reported low back pain for more than 3 months (chronic low back pain, 51%), but the sample also included those with acute (38%) and subacute (11%) low back pain.

Measures

Study participants were administered the Patient-Reported Outcomes Measurement and Information System (PROMIS®)-29 profile survey (Cella, Choi et al.,2019) and the RMDQ at baseline, 6-week later, and 12-weeks later. At the 6-week post-baseline assessment, participants were asked: Compared to your first visit, your low back pain is: *much worse, a little worse, about the same, a little better, moderately better, much better*, or *completely gone.*

The study also included questions aing about being bothered by low back pain in the past week (*BTHLBP*), recency of the start of the current episode of low back pain (*LBPDur*; <1 month; 1-3 months; >3 months), average low back pain in the past week (*PRSAvg*), and worst low back pain in the past 24 hours (*PRSWorst*),

Analysis Plan

We report the percentages of the sample classified on the ISS as having mild, moderate and severe impact and estimate internal consistency reliability for the ISS at baseline, 6-weeks post-baseline, and 12-weeks post-baseline. We estimate test-retest reliability from baseline to 6-weeks later by correlations for the subset of subjects reporting they are "about the same" on the retrospectively self-reported change item administered at the 6-week assessment.

We estimate product-moment correlations between the ISS and other variables hypothesized to be associated with it: Roland-Morris Disability Questionnaire (RMDQ), age, bothered by low back pain in the past week, recency of the start of the current episode of low back pain, average low back pain in the past week, worst low back pain in the past 24 hours), and the PROMIS-29 ability to participate in social roles and activities scale.

We estimate the effect size for change in the ISS from baseline to 6-months later and responsiveness to change of the ISS and its components (physical function, pain inference, pain intensity) from baseline to 6 weeks later using multiple anchors: 1) Retrospective rating of change; 2) How bothersome was your low back pain; and 3) Average low back pain. We estimate Spearman correlations between change in the ISS and the anchors and assess whether the amount of change on the ISS is monotonically related to change implied by the levels of

change suggested by the anchors. We also report F-statistics from general linear models with the ISS as the dependent variable and the anchors as independent variables.

We estimate the area under the curves with the ISS and its components as independent variables predicting improvement on the retrospective rating of change in low back pain item coded as *moderately better, much better* or *completely gone* (all other categories coded as not improved). We also evaluate different cut points for change in the ISS associated with improvement in the retrospective rating of change item. We identify cut points that optimize the Youden index 1950): (sensitivity + specificity)-1.

All analyses were conducted using SAS. ROC analyses were performed using SAS PROC Logistic, SAS PROC Gplot, and the SAS rocplot macro.

Results

Table 1 provides ISS severity categories by duration of low back pain and for the overall sample. Most of the sample (65%) had mild, followed by moderate (21%), and severe (14%) impact. Mild impact was more common among those with chronic low back pain and severe ISS scores were more common among those with acute low back pain.

The internal consistency reliability of the ISS in the ACT dataset was 0.92, 0.93 and 0.93 at baseline, 6 weeks post-baseline and 12 weeks post-baseline, respectively. Test-retest reliability of the ISS from baseline to 6-weeks later was 0.77; test-retest reliability of physical function was 0.70, pain interference was 0.68, and pain intensity was 0.72.

The ISS was correlated strongly with the RMDQ, correlations ranged from 0.75 (baseline) to 0.84 (6-weeks later). Correlations were also noteworthy with being bothered by back pain in the past week, worst back pain in the past 24 hours, average back pain in the last

week, and PROMIS-29 ability to participate in social roles and activities scale (Table 2). Correlations for the overall sample versus those with chronic back pain (shown at baseline) were similar.

The larger correlations post-baseline may be due to increased variance. The standard deviation of the ISS was 8.4, 8.8 and 9.1 at baseline, 6-weeks post-baseline, and 2-weeks post-baseline, respectively. Similarly, the standard deviation of the RMDQ was 5.5, 6.1, and 6.0 at these three timepoints.

The correlations of the components of the ISS (physical function, pain interference, and pain intensity) with the other variables are shown in Tables 3-5. Correlations for the overall sample were similar to those with chronic back pain (show within parentheses at baseline). The components with the strongest correlations varied by the variable and wave of data collection. Physical function correlated most strongly with the RMDQ at baseline and 6-weeks post-baseline, but at the 12-week post-baseline time point pain interference correlated slightly higher ($r = 0.78$ versus 0.75). Pain interference and pain intensity were more strongly associated with bothered by low back pain than was physical function. Not surprisingly, pain intensity was more highly correlated than physical function and pain interference with average and worst pain. Pain interference was more strongly associated than physical function and pain intensity with ability to participate in social roles and activities.

Effect size for change in the ISS from baseline to 6-weeks later was 0.21. Effect size for the physical function component was 0.17, pain interference component was 0.22 and pain intensity component was 0.24. The proportion of study participants in the mild, moderate, and severe ISS categories, respectively, were 65%, 21% and 14% at baseline, 82%, 12% and 6% at 6 weeks, and 84%, 9% and 6% at 12 weeks. From baseline to 6 weeks later, 7% of the sample

changed from severe to mild, 17% got better by one level (severe to moderate, or moderate to mild), 67% stayed the same, 6% got worse by one level (mild to moderate, or moderate to severe), and 1% went from mild to severe. The most common of 634 pattern of responses was mild at each of the three data collection waves (n = 324), followed by moderate at baseline and mild at 6-weeks and 12-weeks post-baseline (n = 60). There were also 48 people who were mild at the first two waves and missing at the third wave. Forty people were severe at baseline and mild at the two post-baseline assessments.

Spearman correlations between change on the ISS and the anchors were 0.63 (retrospective ratings of change in low back pain), 0.68 (bothered by low back pain), and 0.57 (average low back pain) exceeding the 0.371 threshold for an acceptable anchor (Hays, Farivar, & Liu, 2005). The second rows in Tables 6-8 provide the mean ISS change (6 weeks-baseline) for each anchor level. Means that do not differ significantly (Duncan's multiple range test) share a superscript. Change in ISS scores were monotonically associated with the retrospective rating of change item (Table 6). ISS change did not differ significantly between those who reported on the retrospective change item 6-weeks post-baseline that they were *about the same* or *a little better* compared to the first visit. The F-statistic for the change in ISS was F (6,626) =75.36. The F-statistics (all p's <.0001) for the ISS components were: physical function [F (6, 627) = 39.72], pain interference [F(6, 628)=47.11], and pain intensity [F(6, 628)=85.34]. Thus, pain intensity was most sensitive to change suggested by the retrospective change anchor.

Change in ISS scores were generally monotonically associated with the change in bothered by low back pain item except for the one case that changed from not at all to extremely bothered (Table 7). ISS change did not differ significantly between those whose change on the bother item was *about the same* or *a little better* or declined by one, two, or three categories.

Those who stayed the same did not differ significantly from those who declined one or two categories or improved by one category on the bother item. The F-statistic for the change in ISS was F (8,625) =91.10. The F-statistics (all p's <.0001) for the ISS components were: physical function [F (8, 627) = 53.44], pain interference [F(8, 628)=51.20], and pain intensity [F(8, 627)=71.95]. Thus, the ISS was more sensitive than its components to change suggested by bothered by low back pain.

Change in ISS scores were generally monotonically associated with the change in average low back pain item except for the two cases that declined three categories subgroup (Table 8). ISS change was similar for those who stayed the same or declined on the average low back pain item. The F-statistic for the change in ISS was F (6,627) =58.39. The F-statistics (all p's <.0001) for the ISS components were: physical function [F (6, 630) = 25.75], pain interference [F(6, 630)=31.30], and pain intensity [F(6, 629)=133.81]. Hence, pain intensity was most sensitive to change suggested by the average low back pain item (not surprisingly).

As shown in Figures 1-4, area under the curve estimates for retrospective rating of change predicted by change in ISS were as follows: 1) ISS (0.83), physical function (0.75), pain interference (0.78), and pain intensity (0.83). The optimal cut point for decrease in the ISS was about 7 (Figure 5). The sensitivity at this cut point = 66%, specificity = 85%, negative predictive value = 77%, and positive predictive value = 76%.

Discussion

This study provides substantial information about the psychometric properties of the ISS. Internal consistency reliability and test-retest reliability estimates in this study were similar to that reported by Deyo et al. (2015). We also found similar magnitude of associations of the ISS

with the RMDQ as previously reported by Deyo et al. (2014). In addition, we found support for construct validity of the ISS based on its correlations with measures of bother by back pain, worst back pain and ability to participate in social roles and activities. We found that these associations were similar for the overall sample (including those with acute, subacute and chronic low back pain) and limited to those with chronic low back pain.

Our study also extends prior work by examining correlations of the ISS components (physical function, pain interference, pain intensity) as well as the ISS overall. The physical function component was more strongly associated with the RMDQ than were the pain components. In contrast, the pain components were more strongly associated with bother by low back pain than was physical function. Pain interference was more strongly related to ability to participate in social roles and activities than was pain intensity or physical function.

Change in ISS scores in this study were largely monotonically associated with multiple independent anchors of change (retrospective rating of change, change in bother by low back pain, change in average low back pain). Pain intensity was the component of the ISS most sensitive to the retrospective rating of change item and the change in average pain item. The ISS was most sensitive to change in the bothered by low back pain item. Area under the curve analysis demonstrated that the ISS and pain intensity were best able to capture retrospective rating of change when change was defined by those who reported at the 6-week assessment that they were moderately better, much better, or their low back pain was completely gone. The optimal cut-point on the ISS for identifying this definition of change was 7 points. Thus, a 7-point change on the ISS is able to identify those who feel their low back pain has improved substantially who might be considered responders to treatment.

Study limitations are worth noting. As is true with all studies of low back pain of musculoskeletal origin, the specific diagnosis was difficult to determine or confirm. In addition, caution is warranted in generalizing from a sample is active duty members of the U.S. military. Further, a challenge of conducting research in the military is following patients who are transient, especially in times of war. When this study was designed, the active-duty population of interest was likely to be deployed. So, participants were followed for 12 weeks and were excluded if they were scheduled to leave the country within that period. Thus, the relatively short follow-up is a limitation of the study.

Despite the limitations, the study contributes to the limited literature on the ISS and provides additional support for its reliability and validity. Administering the entire PROMIS-29 rather than just the 9 ISS items has the advantage of more comprehensive assessment of health-related quality of life because it includes 4-item sleep disturbance, depression, anxiety, and ability to participate in social roles and activities scales, but the ISS represents the core physical health outcomes associated of low back pain with 20 fewer items. Further evaluation is needed to assess the psychometric properties of the ISS and evaluate the tentative severity classification levels proposed by the NIH Pain Consortium research task force.

References

Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS® adult health profiles: Efficient short-form measures of seven health domains. Value in Health, 22(5), 537-544.

Deyo RA, Ramsey K, Buckley DI, et al. Performance of a Patient Reported Outcomes Measurement Information System (PROMIS) short form in older adults with chronic musculoskeletal pain. *Pain Med.* 2015;17(2):314-324.

Deyo RA, Dworkin SF, Amtmann D, et al. Report of the NIH Task Force on research standards for chronic low back pain. *Pain Med.* 2014;15(8):1249-1267.

Goertz CM, Long CR, Vining RD, Pohlam, K. A., Kane, B, Corber, L., Walters, J., & Coulter, I. Assessment of chiropractic treatment for active duty, U.S. military personnel with low back pain: study protocol for a randomized controlled trial. *Trials*. 2016: 17: 70.

Goertz CM, Long CR, Vining RD, Pohlman KA, Walter J, Coulter I. Effect of Usual Medical Care Plus Chiropractic Care vs Usual Medical Care Alone on Pain and Disability Among US Service Members with Low Back Pain: A Comparative Effectiveness Clinical Trial. *JAMA Netw Open*. 2018;1(1):e180105.

Hays, R. D., Farivar, S. S., & Liu, H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 2005; 2, 63-67.

Von Korff, M., DeBar, L. L., Krebs, E. E., Kerns, R. D., Deyo, R A., & Keefe, F. J. (2020). Graded chronic pain scale revised: Mild, bothersome, and high-impact chronic pain. Pain, 2020; 161(3), 651-661.

Von Korff, M., Ormel, J., Keele, F., & Dworkin, S. F. Grading the severity of chronic pain. *Pain*, 1992; 50, 133-149.

Youden WJ. Cancer. Vol. 3. 1950. Index for rating diagnostic tests; pp. 32–35

Table 1: Impact Stratification Score Severity at Baseline by Duration of Low Back Pain

Severity	Acute (n = 286)	Subacute (n = 79)	Chronic (n = 384)	Overall (n = 749)
Mild	56%	61%	72%	65%
Moderate	23%	29%	18%	21%
Severe	21%	10%	10%	14%
Column Percent	38%	11%	51%	100%

Table 2: Product-moment Correlations of Impact Stratification Score with Other Variables

	Baseline (chronic back pain only)	6-weeks later	12-weeks later
RMDQ	0.75 (0.76)	0.84	0.83
Age	0.02 (0.09)	0.15	0.16
BTHLBP	0.65 (0.66)	0.78	0.78
PRSAvg	0.51 (0.56)	0.74	0.75
PRSWorst	0.56 (0.56)	0.77	0.77
SRTScore	-.64 (-.66)	-.68	-.71

RMDQ = Roland-Morris Disability Questionnaire, BHTLBP = bothered by low back pain in the past week, LBPDur = recency of the start of the current episode of low back pain: <1 month; 1-3 months; >3 months, PRSAvg = average low back pain in the past week, PRSWorst = worst low back pain in the past 24 hours, SRTScore = PROMIS-29 ability to participate in social roles and activities scale.

Table 3. Product-moment Correlations of Physical Function, Pain Interference, and Pain Intensity with Other Variables at Baseline

	Physical Function (chronic back pain only)	Pain Interference (chronic back pain only)	Pain Intensity (chronic back pain only)
RMDQ	0.71 (0.72)	0.68 (0.68)	0.48 (0.49)
Age	0.06 (0.15)	0.00 (0.05)	-.01 (0.03)
BTHLBP	0.52 (0.51)	0.60 (0.60)	0.59 (0.59)
PRSAvg	0.33 (0.38)	0.40 (0.42)	0.76 (0.80)
PRSWorst	0.45 (0.38)	0.45 (0.48)	0.62 (0.66)
SRTScore	-.56 (-.60)	-.63 (-.64)	-.40 (-.40)

RMDQ = Roland-Morris Disability Questionnaire, BHTLBP = bothered by low back pain in the past week, LBPDur = recency of the start of the current episode of low back pain: <1 month; 1-3 months; >3 months, PRSAvg = average low back pain in the past week, PRSWorst = worst low back pain in the past 24 hours, SRTScore = PROMIS-29 ability to participate in social roles and activities scale.

Table 4. Product-moment Correlations of Physical Function, Pain Interference, and Pain Intensity with Other Variables at 6-weeks post-Baseline

	Physical Function	Pain Interference	Pain Intensity
RMDQ	0.81	0.77	0.68
Age	0.13	0.17	0.09
BTHLBP	0.67	0.70	0.75
LBPDur	0.15	0.19	0.26
PRSAvg	0.58	0.62	0.91
PRSWorst	0.62	0.67	0.87
SRTScore	-.63	-.65	-.53

RMDQ = Roland-Morris Disability Questionnaire, BHTLBP = bothered by low back pain in the past week, LBPDur = recency of the start of the current episode of low back pain: <1 month; 1-3 months; >3 months, PRSAvg = average low back pain in the past week, PRSWorst = worst low back pain in the past 24 hours, SRTScore = PROMIS-29 ability to participate in social roles and activities scale.

Table 5. Product-moment Correlations of Physical Function, Pain Interference, and Pain Intensity with Other Variables at 12-weeks Post-Baseline

	Physical Function	Pain Interference	Pain Intensity
RMDQ	0.75	0.78	0.70
Age	0.13	0.16	0.13
BTHLBP	0.62	0.74	0.79
LBPDur	0.15	0.21	0.31
PRSAvg	0.54	0.66	0.92
PRSWorst	0.58	0.69	0.87
SRTScore	-0.62	-0.68	-0.60

RMDQ = Roland-Morris Disability Questionnaire, BHTLBP = bothersome of low back pain in the past week, LBPDur = recency of the start of the current episode of low back pain: <1 month; 1-3 months; >3 months, PRSAvg = average low back pain in the past week, PRSWorst = worst low back pain in the past 24 hours, SRTScore = PROMIS-29 ability to participate in social roles and activities scale.

Table 6. Change in ISS by: Compared to your first visit, your low back pain is: *much worse, a little worse, about the same, a little better, moderately better, much better* or *completely gone*

Much worse	A little worse	About the same	A little better	Moderately better	Much better	Completely gone
8[a]	3[b]	-2[c]	-3[c]	-6[d]	-12[e]	-17[f]
n = 16	n = 64	n = 178	n = 104	n = 89	n = 146	n = 36

F (6, 626) = 75.36, p <.0001

Table 7. Change in ISS by: How bothersome was your low back pain in the past week? *Not at all bothersome, Slightly bothersome, Moderately bothersome, Very bothersome, Extremely bothersome* [6 weeks-baseline]

From not at all to Extremely bothersome	Declined three categories	Declined two categories	Declined one category	Stayed the same	Improved one category	Improved two categories	Improved three categories	From Extremely to Not at all bothersome
-9[e,d]	10[a]	8[a,b]	2[a,b,c]	0[b,c,d]	-5[c,d,e]	-11[e,f]	-19[f]	-29[g]
n = 1	n = 1	n = 8	n = 62	n = 217	n = 176	n = 114	n = 45	n = 10

F (8, 625) = 91.10, p<.0001

Table 8. Change in ISS by: Select the number that best describes your average low back during the past week. 0 = No pain -> 10= Worst possible pain (Recoded so that 10=5, 7-9 = 4, 4-6 = 3, 1-3 = 2, and 0 = 1; based on Sheehan Disability Scale and the Flushing Questionnaire) [6 weeks-baseline]

Declined 3 categories	Declined 2 categories	Declined 1 category	Stayed the same	Improved 1 category	Improved 2 categories	Improved 3 categores
$-6^{b,c}$	6^{a}	$1^{a,b}$	$-2^{b,c}$	-8^{c}	-16^{d}	-24^{e}
n = 2	n = 10	n = 85	n = 273	n = 203	n = 51	n = 10

F (6, 627) = 58.39, p<.0001

Figure 1. Area Under the Curve Predicting Improvement on Retrospective Change by Change in ISS (Baseline to 6-Weeks Later)

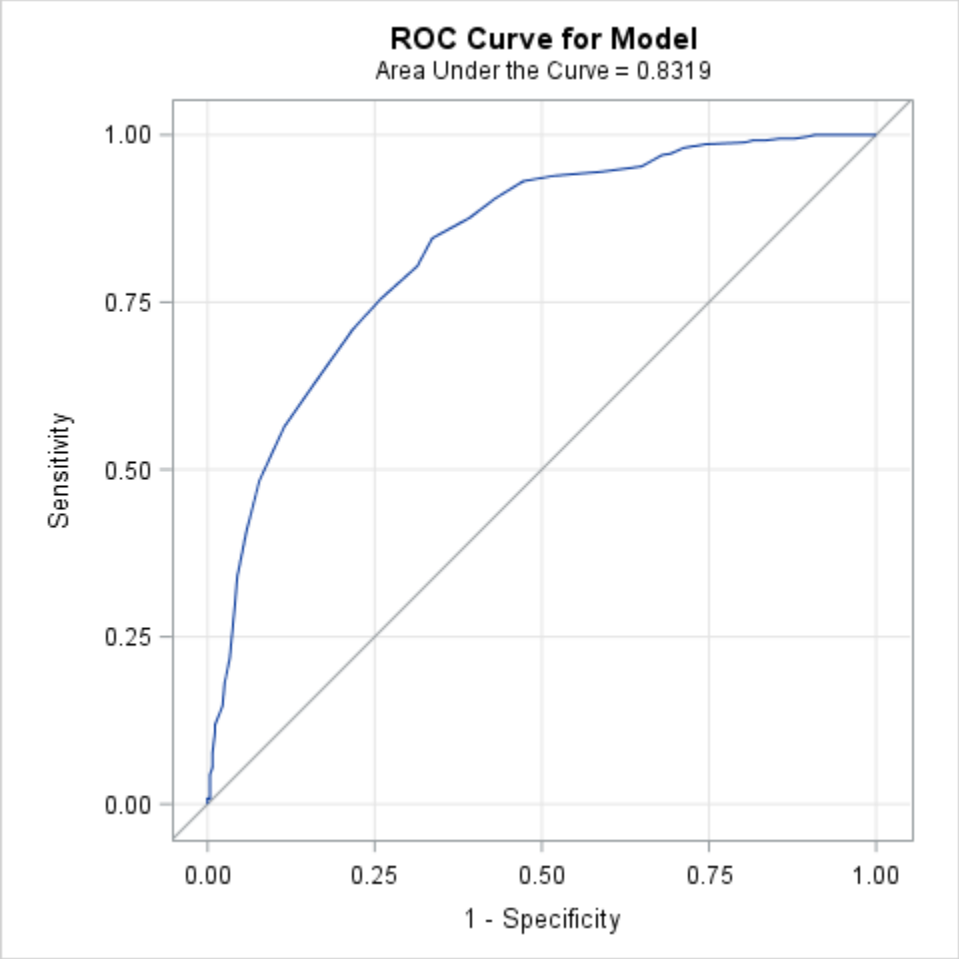

Figure 2. Area Under the Curve Predicting Improvement on Retrospective Change by Change in Physical Function (Baseline to 6-Weeks Later)

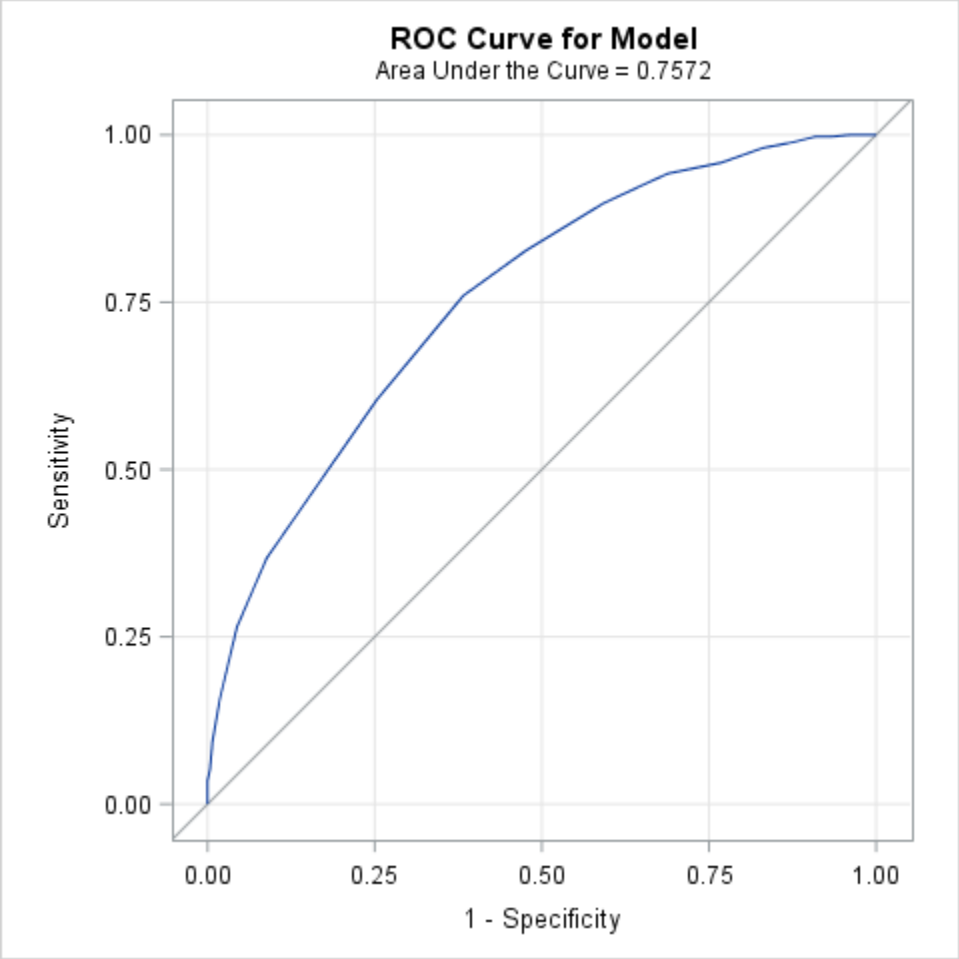

Figure 3. Area Under the Curve Predicting Improvement on Retrospective Change by Change in Pain Interference (Baseline to 6-Weeks Later)

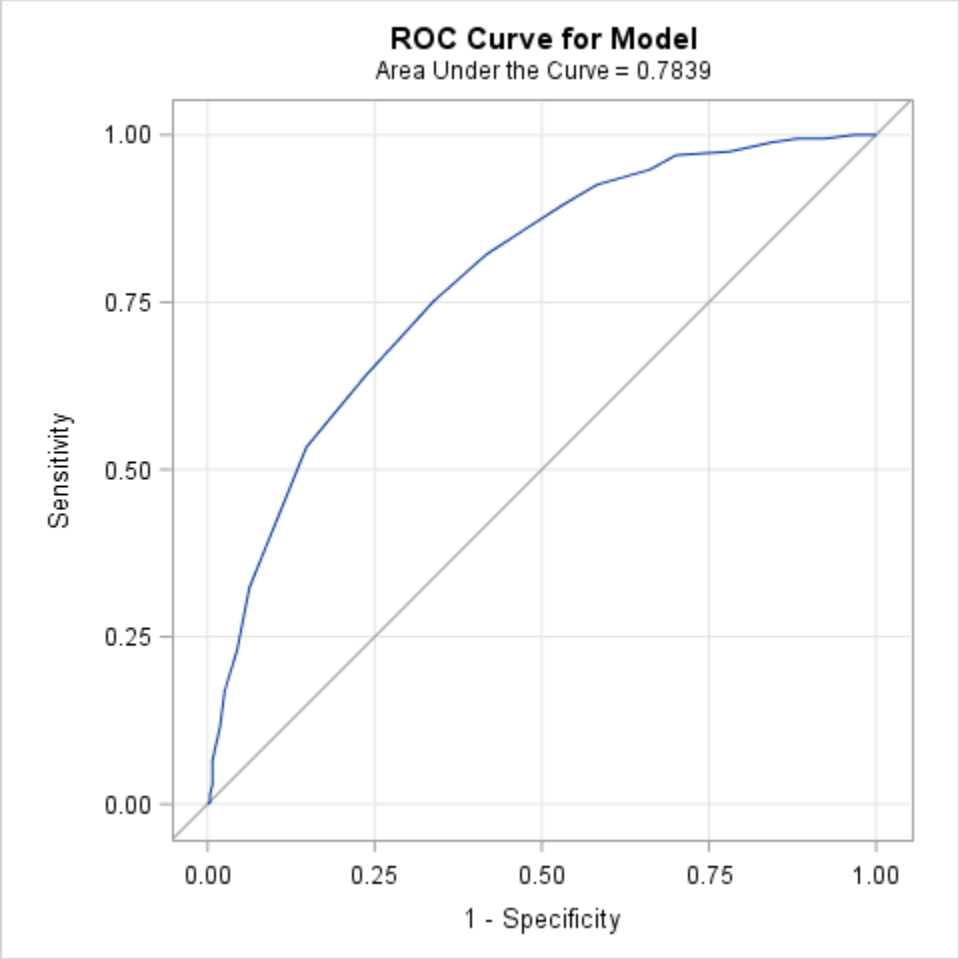

Figure 4. Area Under the Curve Predicting Improvement on Retrospective Change by Change in Pain Intensity (Baseline to 6-Weeks Later)

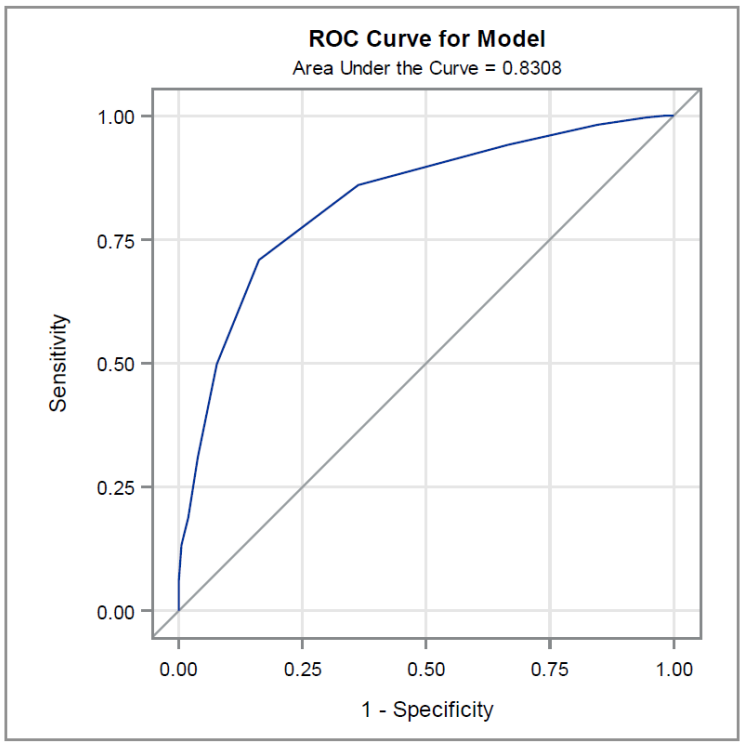

Figure 5. Optimal Cut points for Change in ISS (Baseline to 6-Weeks Later)