

Are your **MRI contrast agents** cost-effective?

Learn more about generic **Gadolinium-Based Contrast Agents**.



AJNR

Correlation of the National Institutes of Health Patient Reported Outcomes Measurement Information System Scales and Standard Pain and Functional Outcomes in Spine Augmentation

L. Shahgholi, K.J. Yost and D.F. Kallmes

This information is current as of April 19, 2024.

AJNR Am J Neuroradiol 2012, 33 (11) 2186-2190

doi: <https://doi.org/10.3174/ajnr.A3145>

<http://www.ajnr.org/content/33/11/2186>

**ORIGINAL
RESEARCH**

L. Shahgholi
K.J. Yost
D.F. Kallmes



Correlation of the National Institutes of Health Patient Reported Outcomes Measurement Information System Scales and Standard Pain and Functional Outcomes in Spine Augmentation

BACKGROUND AND PURPOSE: The recently developed National Institutes of Health PROMIS initiative provides reliable and valid measures across many health domains. We correlated changes in pain-related PROMIS measures and changes in both an NRS and the RMDI in patients undergoing spine augmentation.

MATERIALS AND METHODS: Fifty patients, composed of 26 women (40–91 years of age; mean, 72.6 years) and 24 men (42–78 years of age, mean, 67.5 years) were enrolled in the study. They were asked at initial presentation and at 30 days to rate the intensity of their pain in the past 24 hours by using a 0–10 pain NRS as well as the 23-item RMDI. Study subjects also completed 3 different PROMIS short forms, including physical function, pain behavior, and pain interference. The Spearman correlation was used to assess the correlation between the scales. The RCI \times 1.96 was calculated for each measurement tool as an indicator of change.

RESULTS: All instruments were responsive to detection of change during 1 month (all, $P < .0001$). Correlations between changes in physical function, pain interference, and pain behavior PROMIS scores and changes in RMDI scores were 0.37, 0.44, and 0.42, respectively. Direction of changes (declines versus improvements) in RMDI and other scales were the same in approximately 60% of patients.

CONCLUSIONS: All measures evaluated had adequate and comparable psychometric properties. The choice of which measure to use depends on the clinical intent of the intervention.

ABBREVIATIONS: IRT = item response theory; MID = minimally important differences; NRS = numerical rating scale; PROMIS = Patient Reported Outcomes Measurement Information System; RCI = Reliable Changes Index; RMDI = Roland-Morris Disability Index; SEM = standard error of measurement

Patient-reported pain severity represents the prime outcome in most spine augmentation studies. Various permutations of pain-reporting techniques, including the visual analog scale, ordinal NRSs, and Box Score-11 Scales have been applied to this patient population.¹ In most previous studies, substantial improvement in pain as measured by these scales following spine augmentation was found in 70%–80% of patients.

Notwithstanding the consistent positive impact on patient-reported pain severity noted in most studies, serious shortcomings remain in the assessment of pain in patients treated with vertebroplasty. One recent study demonstrated markedly different pain-severity responses in a single cohort of patients treated with vertebroplasty based simply on different qualifi-

ers used for pain questions. These qualifiers, including “worst or best pain” or “pain at rest or with activity,” yielded mean preprocedural pain severities as great as 8/10 and as low as 2/10 in the same patient population at the same interview session. Furthermore, pain-severity responses may be impacted by psychosocial and behavioral factors that vary among patients.

The NRS and RMDI have several possible shortcomings compared with PROMIS measures. While the NRS is familiar to most patients, a single item such as an NRS measuring a complex construct like pain likely will have more measurement error than multi-item scales like the RMDI or PROMIS measures; as such, the NRS may be less effective than PROMIS at detecting clinically meaningful changes. The 23-item RMDI is longer than the PROMIS measures. Furthermore, the dichotomous response scale used in the RMDI is less discriminating than the 5- or 6-level response scale in the PROMIS measures.


The PROMIS is a National Institutes of Health infrastructure project charged with using the IRT²⁻⁴ to develop assessments of patient-reported outcomes that are brief and maximally reliable and valid in several health domains, including those related to pain. IRT refers to a family of statistical models that describe the relationship between a person’s response to an item on a questionnaire and his or her level of the construct (eg, pain, functioning) that is being measured by the questionnaire.⁵ For example, a person experiencing severe pain should

Received January 31, 2012; accepted after revision March 31.

From the Departments of Neuroradiology (L.S., D.F.K.) and Epidemiology (K.J.Y.), Mayo Clinic, Rochester, Minnesota.

We certify that no party supporting this article, having a direct interest in the results of the research, has or will confer a benefit on us or on any organization with which we are associated; and if applicable, we certify that all financial and material support for this research (eg, NIH or National Health Service grants) and work is clearly identified in the article.

Please address correspondence to David Kallmes, MD, 200 First St SW, Rochester, MN 55905; e-mail: Kallmes.david@mayo.edu.

 Indicates article with supplemental on-line appendix.

<http://dx.doi.org/10.3174/ajnr.A3145>

Table 1: Demographic characteristic of patients

Characteristic	No. of Patients (%)
Sex	
Female	26 (52%)
Male	24 (48%)
Ethnicity	
White	47 (94%)
African American	3 (6%)
Relationship status	
Never married	5 (10%)
Married	35 (70%)
Committed partner	1 (2%)
Divorced	5 (10%)
Widowed	6 (12%)
Avocational status	
Retired	26 (52%)
Employed	10 (20%)
Disabled/unable to work	3 (6%)
Unemployed	11 (24%)
Average annual income	
<\$20,000	5 (10%)
\$20,000–\$50,000	18 (36%)
>\$50,000	20 (40%)
Educational status	
Refused to answer	7 (14%)
Less than high school level	7 (14%)
High school graduate	18 (36%)
College or university degree	25 (50%)

respond in a predictable way to questions asking about the severity of the pain. IRT models can be used to explain the relationship between the level of pain experienced by a person and the probability that he or she will answer pain questions a certain way.

As such, PROMIS measures may offer alternatives to common pain scales currently in widespread clinical use.⁶ The PROMIS scales most relevant to patients receiving an intervention for back pain include “physical function,” “pain interference,” and “pain behavior.” The scores derived from PROMIS scales have been normalized on the basis of the general population.

PROMIS measures may be useful outcome measures for patients treated with vertebroplasty. However, these measures have never been validated in this patient population. The purpose of this study was to compare the psychometric properties of selected PROMIS measures with 2 commonly used outcomes measures: the pain NRS and RMDI. We hypothesized that the psychometric properties would be comparable among

measures and that PROMIS measures will be suitable options for assessing outcomes in this patient population.

Materials and Methods

Our institutional review board approved this prospective Health Insurance Portability and Accountability Act–compliant study, and written consent of all study participants was obtained. Between November 2008 and November 2011, fifty-nine consecutive patients referred to the radiology department for consideration of spine augmentation were enrolled. Potential participants were considered for enrollment in the study if they had a vertebral fracture due to osteoporosis or multiple myeloma (even with metastasis) and were able to answer the questions in English. Exclusion criteria were participation in another spine augmentation trial or not being considered appropriate for spine augmentation. Nine (11.8%) of 59 could not be contacted for the follow-up assessment and were excluded from the analysis.

In the first evaluation, after a physician visit in the evaluation room and after explaining how to answer the questions, patients were asked to complete PROMIS short forms in 3 different domains relevant to pain, including physical function, pain behavior, and pain interference. Patients who were able to complete the forms themselves did so. Patients who could not complete the forms were asked the questions in an interview style. In our practice, the average time to obtain scales is about 5 minutes for all 3 PROMIS forms and 3 minutes for the RMDI. The 10-question physical function short form is focused on the ability to perform various daily activities from self-care (eg, bathing and dressing) to vigorous physical activities (eg, running, strenuous sports; On-line Physical Functioning Form). The 6-question pain interference short form is focused on pain interfering with mental, physical, and social aspects of daily living (On-line Pain Impact Form). The 7-question pain behavior short form focuses on verbal, facial, and bodily expressions of pain (On-line Pain Behavior Form).

Responses to PROMIS questions on a given short form are combined⁷ and can be reported as raw scores and/or *t*-scores. The *t*-score scale has a mean score of 50 and an SD of 10 in the general population of the United States. For example, a person who has a PROMIS pain interference score of 70 is reporting adverse pain interference 2 SDs worse than the general population average.⁸ Higher *t*-scores indicate greater levels of the construct being measured. Thus, for pain behavior and pain interference, higher scores reflect worse pain, whereas for physical function, higher scores indicate better functioning.

Participants also rated the intensity of their pain in the past 24 hours by using a pain NRS from 0 to 10, with 0 indicating no pain and 10 indicating the worst imaginable pain. The 23-item modified RMDI, a widely used functional outcome scale, was also adminis-

Table 2: Pain instruments, baseline, 1 month, and change scores

	No. of Items	No. of Response Options	Baseline (mean) (median, range)	1 Month (mean) (median, range)	Change Score (1 month, baseline) (mean)	Change Score (1 month, baseline) (SD) (95% CI for SD)	<i>P</i> Value ^a
RMDI	23	2	17.6 ± 4.8 (19, 1–23)	14.6 ± 5.5 (16, 0–23)	–3.1	5.2 (1.6–4.6)	<.0001
PROMIS PF <i>t</i> -score	10	5	29.7 ± 6.8 (30, 14–41)	33.4 ± 6.1 (32.7, 18.7–49)	3.6	5.4 (2.1–5.1)	<.0001
PROMIS PI <i>t</i> -score	6	5	67.2 ± 7.7 (67.4, 32.7–78.3)	60.8 ± 9.6 (61.8, 41–78)	–6.3	10.1 (2.3–3.5)	<.0001
PROMIS PB <i>t</i> -score	7	6	62 ± 3.6 (62.5, 52–72)	59 ± 5 (59.2, 36–65)	–3.1	5.2 (1.6–4.5)	<.0001
Average pain in past 24 hr, NRS	1	11	5.5 ± 2 (6, 0–9)	3.7 ± 2.45 (3, 1–8)	–1.9	2.9 (–1 to –2.7)	<.0001

Note:—CI indicates confidence interval; PB, pain behavior; PF, physical function; PI, pain interference.

^a Pair-wise *t* test.

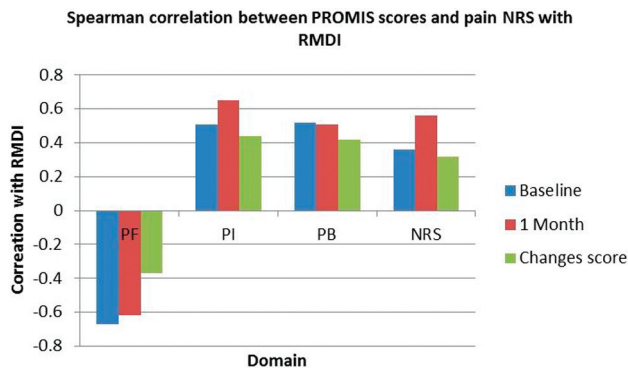


Fig 1. Spearman correlation coefficients (ρ) between PROMIS scores, pain NRS, and RMDI in the first-day and 30-day evaluation in the vertebroplasty group. PB indicates pain behavior; PF, physical function; PI, pain interference.

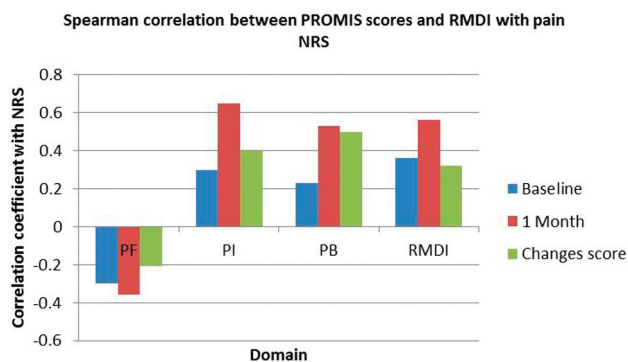


Fig 2. Spearman correlation coefficients (ρ) between PROMIS domains and average pain in the past 24-hour NRS on the first-day and 30-day evaluation in the vertebroplasty group. PB indicates pain behavior; PF, physical function; PI, pain interference.

tered. Study participants were telephoned 30 days after the initial assessment, at which time they were asked the same sets of questions.

Statistical Analysis

The Spearman correlation was determined to assess the correlation between RMDI outcome measures and PROMIS and NRS scores. The $RCI \times 1.96$ was calculated for the multi-item measurement tools (ie, PROMIS and RMDI) as an indicator of change that is greater than the measurement error of the scale.^{9,10} RCI is calculated as $\sqrt{2} \times SEM$. We used the internal consistency reliability (Cronbach α) of each tool¹¹⁻¹⁴ to derive the SEM. Changes exceeding 2 points on the pain NRS were considered clinically meaningful.¹⁵ Patients were categorized as having experienced meaningful improvement if their scores improved by at least $1.96 \times RCI$ for the RMDI and PROMIS scales or by at least 2 points for the NRS from baseline to month 1. Similarly, patients were categorized as having experienced meaningful decline if

their scores worsened by at least $1.96 \times RCI$ or 2 points. Criteria for an adequate cross-sectional and longitudinal Spearman correlation (ρ) were set at >0.5 and >0.3 , respectively.^{16,17} P values $< .05$ indicated a significant difference. Statistical analyses were performed by using JMP software, Version 7.0 (SAS Institute, Cary, North Carolina).

Results

Study Participants

Fifty patients met enrollment criteria and were available for follow-up at 1 month (Table 1). The sample consisted of 24 (48%) men and 26 (52%) women. The men ranged in age from 42 to 87 years (mean, 67.5 ± 15 years). The women ranged in age from 40 to 91 years (mean, 72.6 ± 12.4 years). Among the 26 women, 24 (92%) had a prior diagnosis of osteoporosis or osteopenia. Two women had history of multiple myeloma. Among the 24 men, 20 (83%) had a prior diagnosis of osteoporosis, and the rest had multiple myeloma. The patients underwent spine augmentation in 72 different levels: 5 (7%) midthoracic, 25 (34.7%) lower thoracic, 26 (36.1%) upper lumbar, 15 (21%) lower lumbar, and 1 (1.4%) sacroplasty.

Rating Pain, PROMIS T-Score, and RMDI

Table 2 presents baseline and 1-month scores, change scores, and significance levels for the paired t test comparing baseline and 1-month scores. All instruments were responsive to the detection of change during 1 month (all $P < .0001$). Mean scores for all domains demonstrated improvement during 1 month (eg, less pain, better physical function). SDs of the mean difference were reduced with some instruments compared with others. A decrease in the SD was seen in PROMIS physical function (6.8–6.1) when others had increases in this factor. The highest increment in SD was seen in the PROMIS pain interference (7.7–9.6).

Correlations between either RMDI or NRS and PROMIS scores were significant in cross-sectional measurements (all $P < .01$). Correlations between changes with time in PROMIS scores and changes with time both in pain NRS scores and RMDI scores were significant ($P < .05$), except for the cross-sectional correlation between NRS and pain behavior at the baseline assessment ($P = .10$) and the longitudinal correlation between changes in NRS and changes in physical function ($P = .15$) (Figs 1 and 2).

Direction of changes (declines versus improvements) in RMDI and PROMIS physical function, pain interference, pain behavior, and pain NRS was the same in 30 (60%), 32 (64%), 31 (62%), and 26 (52%) patients, respectively. Similarly, in the direction of changes in pain, NRS and other measurements

Table 3: Change outcome after 1 month based on the specific RCI of different measurement tools

	Reliability (Cronbach α)	Criterion for Meaningful Change ^a	Patients Experiencing Meaningful Improvement (No.) (%)	Patients Experiencing No Change (No.) (%)	Patients Experiencing Meaningful Decline (No.) (%)
RMDI	0.90	4.3	16 (32%)	32 (64%)	2 (4%)
PF	0.90	5.9	16 (32%)	33 (66%)	1 (2%)
PI	0.89	6.4	22 (44%)	26 (52%)	2 (4%)
PB	0.77	4.7	15 (30%)	32 (64%)	3 (6%)
NRS	NA	2 ^a	26 (52%)	20 (40%)	4 (8%)

Note:—NA indicates not applicable; PB, pain behavior; PF, physical function; PI, pain interference.

^aMeaningful change for RMDI and PROMIS measures is defined as a change score with an absolute value $\geq 1.96 \times$ Reliable Change Index. For NRS, meaningful change is defined as a change score with an absolute value ≥ 2 points, based on Farrar et al.¹⁵

ranged from 22 (44%) patients with physical function to 27 (54%) patients with PROMIS pain behavior.

On the basis of $1.96 \times \text{RCI}$, patients who had changes of >4.3 points for the RMDI, 5.9 points for physical function, 6.4 points for pain interference, and 4.7 points for pain behavior were considered to have experienced meaningful improvement. Across all measures, approximately 30%–50% of patients achieved clinically significant improvement (Table 3).

Discussion

Our current study demonstrated that 3 PROMIS measures relevant to patients with fracture-related back pain were correlated with measures commonly used in clinical practice, including the pain NRS and RMDI, and that these PROMIS measures very likely are appropriate tools for assessing patient experiences of pain related to vertebral compression fractures. All measurement tools had good responsiveness in patients treated with vertebroplasty. According to this study, with strong correlation between the RMDI and PROMIS scale, nearly equivalent results between the 2 measures can be obtained with PROMIS short forms. The benefit of the PROMIS short forms is that they are able to attain good reliability and responsiveness to change with fewer questions (7–10 compared with 23 in the RMDI).

The correlation between PROMIS pain behavior and RMDI scores was weaker than that for the other 2 PROMIS scales. This finding was expected, given the content of the scales. Unlike the PROMIS behavior scale, RMDI mainly measures functional disability and does not address the emotional and behavioral effects of pain. In analyzing the direction of changes for the different scales, we determined that they were in the same direction in approximately two-thirds of patients.

Several previous studies have evaluated the PROMIS scales in focused clinical populations. Fries et al² studied 451 patients with chronic rheumatoid arthritis and showed a strong correlation between PROMIS physical function and Health Assessment Questionnaire or Health Assessment Questionnaire Disability Index in patients with chronic rheumatoid arthritis. A recently published article by Baja et al¹⁸ reported significant correlation between PROMIS computerized adaptive testing tools and their corresponding legacy instruments for assessing Health-Related Quality of Life in patients with cirrhosis. In our current study, correlations between RMDI and PROMIS scales were stronger than those of Fries et al. We also observed good longitudinal correlation among change scores.

Exact determination of clinically meaningful changes in patient-reported outcomes remains difficult to assess. Yost et al¹⁷ defined MID in 6 PROMIS-Cancer scales, including physical function. Their recommended MID was 4–6 points for physical function and pain interference short forms (an MID for the pain behavior short form was not determined). An MID of 2–3 points for low back pain had already been established for the RMDI.¹⁹ However, because the methodology for establishing the MID differed between the PROMIS and RMDI scales, we used $1.96 \times \text{RCI}$ as a standard metric for evaluating change scores. Changes exceeding $1.96 \times \text{RCI}$ would be unlikely to occur at $P < .05$ in the absence of actual change.⁹ This frequently used indicator of change⁹ was not designed to determine clinically significant cutoff points for

deterioration, focusing only on improvement as a goal of therapy, which is well-accepted as a standard and liberal way of assessing longitudinal changes.

There are several limiting factors that affected our study. The instruments are not anchored to the same recall period. Specifically, when answering the RMDI and NRS, patients are instructed to report on their pain “today,” whereas the context for the PROMIS pain behavior and pain interference items is the past 7 days. The PROMIS physical function scale does not specify a recall period for the items. In addition, there is no true criterion standard for detecting meaningful improvement in pain following spine augmentation; as such, our data alone cannot indicate which measure is superior for clinical use. Finally, patients were not asked about medication use, which could have affected their pain ratings.

Conclusions

All measures used in this study offered adequate and comparable psychometric properties. The choice of which measure to use depends on the clinical intent of the intervention. If alleviating pain intensity is paramount, then the pain NRS may be the ideal choice for an outcome measure, whereas for improvement of functioning and decrease in the impact of pain on other aspects of well-being, RMDI, PROMIS physical function, or PROMIS pain interference may be better suited outcome measures.

Acknowledgments

We thank Roanna Vine and Leigh Gray for data gathering and supportive help.

Disclosures: David F. Kallmes—UNRELATED: Grants/Grants Pending: Sequent,* MicroVention,* Benvenue Medical,* Penumbra,* Payment for Lectures (including service on Speakers Bureaus): MicroVention, Royalties: University of Virginia Patent Foundation, Payment for Development of Educational Presentations: ev3,* CareFusion,* Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: MicroVention.* *Money paid to the institution.

References

1. Iskedjian M, Jaszewski B, Desjardins O. Comparison of scores derived from the Box Score-11 Scale and the pain attribute of the Health Utilities Index-3. *Patient* 2011;4:259–65
2. Fries JF, Krishnan E, Rose M, et al. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther* 2011;13:R147
3. Hambleton RK, Saminathan H, Rogers HJ. *Fundamentals of Item Response Theory*. Newbury Park, California: Sage Publishing; 1991
4. Kleyweg RP, van der Meche FG, Schmitz PI. Interobserver agreement in the assessment of muscle strength and functional abilities in Guillain-Barré syndrome. *Muscle Nerve* 1991;14:1103–09
5. 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: S22–31
6. Ader DN. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care* 2007;45(suppl 1):S1–S2
7. PROMIS. <http://www.nihpromis.org>. Accessed February 6, 2012
8. Yost KJ, Eton DT, Garcia SF, et al. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol* 2011;64: 507–16
9. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59: 12–19
10. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res* 2011;11: 171–84
11. Hahn EA, Cella D, Chassany O, et al. Precision of health-related quality-

- of-life data compared with other clinical measures. *Mayo Clin Proc* 2007; 82:1244–54
12. The Research Methods Knowledge Base. October 20, 2006. <http://www.socialresearchmethods.net/kb>. Accessed January 9, 2012
 13. 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: 2005–08.** *J Clin Epidemiol* 2010;63: 1179–94
 14. Martinez-Martin P, Carod-Artal FJ, da Silveira Ribeiro L, et al. **Longitudinal psychometric attributes, responsiveness, and importance of change: an approach using the SCOPA-Psychosocial questionnaire.** *Mov Disord* 2008;23: 1516–23
 15. Farrar JT, Young JP Jr, LaMoreaux L, et al. **Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale.** *Pain* 2001;94:149–58
 16. Revicki DA, Chen WH, Harnam N, et al. **Development and psychometric analysis of the PROMIS pain behavior item bank.** *Pain* 2009;146:158–69
 17. Yost KJ, Cella D, Chawla A, et al. **Minimally important differences were estimated for the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) instrument using a combination of distribution- and anchor-based approaches.** *J Clin Epidemiol* 2005;58:1241–51
 18. Bajaj JS, Thacker LR, Wade JB, et al. **PROMIS computerised adaptive tests are dynamic instruments to measure health-related quality of life in patients with cirrhosis.** *Aliment Pharmacol Ther* 2011;34:1123–32
 19. Bombardier C, Hayden J, Beaton DE. **Minimal clinically important difference: low back pain—outcome measures.** *J Rheumatol* 2001;28:431–38