HSSJ (2020) 16 (Suppl 2):S443–S451
DOI 10.1007/s11420-020-09817-w

HSS Journal®

ORIGINAL ARTICLE

Validating the VR-12 Physical Function Instrument After Anterior Cervical Discectomy and Fusion with SF-12, PROMIS, and NDI

Nathaniel W. Jenkins, MS · James M. Parrish, MPH · Michael T. Nolte, MD · Nadia M. Hrynewycz, BS · Thomas S. Brundage, BS · Kern Singh, MD

Received: 15 April 2020/Accepted: 29 September 2020 /Published online: 9 November 2020

Abstract *Background:* Development and validation of Veterans RAND 12-item (VR-12) physical component survey (PCS) has been established among civilian and veteran populations but it has not been examined among anterior cervical discectomy and fusion (ACDF) patients. *Purposes/Questions:* We sought to validate legacy patient-reported outcome measures (PROMs) with VR-12 PCS among patients undergoing ACDF procedures. *Methods:* A prospectively collected surgical registry was retrospectively evaluated for elective single or multi-level ACDFs performed for degenerative spinal pathologies from January 2014 to August 2019. Exclusion criteria included missing pre-operative surveys and surgery for trauma, metastasis, or infection. Demographic variables, baseline pathologies, and peri-operative variables were collected. A paired *t* test evaluated the change from the pre-operative score to each post-operative timepoint for VR-12 PCS, the 12-item Short-Form Survey (SF-12) PCS, Patient-Reported Outcomes Measurement Information System physical function (PROMIS-PF), and Neck Disability Index (NDI). Minimal clinically important difference (MCID) achievement was calculated at each timepoint. Correlation was evaluated with a Pearson's correlation coefficient and time-independent partial correlation. *Results:* Of the 202 patients who underwent ACDF, 41.1% were female and the average age was 49.5 years. All PROMs had statistically significantly increased from baseline when compared with post-operative timepoints (12 weeks, 6 months, 1 year, and 2 years). MCID achievement rates increased through 2 years. All timepoints revealed strong VR-12 PCS correlations with SF-12 PCS, PROMIS-PF, and NDI scores. *Conclusion:* VR-12 PCS was strongly correlated with the well-validated SF-12 PCS and NDI metrics as well as with the more recent PROMIS-PF. All PROMs demonstrated statistically significant improvement in patients post-operatively. VR-12 PCS is a valid measure of physical function among patients undergoing ACDF.

Keywords ACDF · anterior cervical discectomy and fusion · cervical surgery · physical function · patient-reported outcomes

Level of Evidence: Level III: Retrospective Cohort Study

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11420-020-09817-w) contains supplementary material, which is available to authorized users.

N. W. Jenkins, MS · J. M. Parrish, MPH · M. T. Nolte, MD · N. M. Hrynewycz, BS · T. S. Brundage, BS · K. Singh, MD (✉)
Department of Orthopaedic Surgery, Rush University Medical Center,
1611 W. Harrison St. Suite #300,
Chicago, IL 60612, USA
e-mail: kern.singh@rushortho.com

Introduction

Over the past decade in the USA, there has been a rapidly growing emphasis on providing value-based care. This value is determined by measuring the benefit to the patient relative to the costs of the intervention. Often, this benefit is framed in terms of patient-reported outcome measures (PROMs), as these metrics have been strongly linked to overall patient satisfaction [7, 35]. PROMs also serve important roles in the prediction of peri-operative outcomes [33] and in the risk stratification of patients [25, 29, 40]. Finally, PROMs may help to determine not only the cost-effectiveness of an intervention but also provider reimbursement [6, 40]. In light of these points, establishing the validity of these measures and determining their clinical relevance are of the utmost importance.

PROMs have been widely employed in the spine population, especially for patients undergoing cervical spine surgery. The Neck Disability Index (NDI), for example, which measures functional disability due to cervical spine pathology, has served as a keystone PROM in spine surgery since

1991 [39, 41]. Similarly, the RAND Corporation developed generalizable health outcome assessments in the 36-item Short-Form Survey (SF-36) and the abbreviated 12-item Survey (SF-12) that have been used not only in spine surgery but also in myriad other specialties [16, 42]. The Veterans RAND 12-item Survey (VR-12) represents a modification of the SF-12 that aims to increase the precision of the instrument and has been administered and validated in over 5 million questionnaires nationwide [23, 37]. Furthermore, the VR-36 and VR-12 improved on the earlier SF instrument by implementing 5-point response scales to address role limitation due to physical and emotional function. In addition to these changes, the VR series adds two items that change how mental and physical health are assessed over time [19, 21, 22]. More recent efforts include the Patient-Reported Outcomes Measurement Information System (PROMIS), which can utilize computer-adaptive testing (CAT) to evaluate health outcomes. The CAT technology employs an algorithmically based question system such that the outcome measure scores can be determined with high specificity and fewer questions relative to comparable metrics.

With numerous PROMs to choose from, the VR-12 remains one of the most widely used in patients undergoing cervical spine surgery [40]. Its ease of use for both patient and provider, in addition to the breadth of previous research and available data, has rendered the VR-12 a powerful legacy tool. Furthermore, in light of the growing emphasis on PROMIS, early research has suggested that VR-12 scores can be easily linked and converted into the PROMIS framework [36].

Minimal clinically important difference (MCID) is a threshold value that may be used to quantify the smallest change in PROM score that corresponds to a patient's perceived benefit from treatment. It was first defined in 1989 by Jaeschke et al., although over the years it has had many minor variations in definition [18]. It should be noted that the methods for calculating MCID and the threshold values themselves vary throughout literature. Anterior cervical discectomy and fusion (ACDF) is a reliable treatment modality that remains to be the most commonly performed surgical procedure for degenerative pathology of the cervical spine [1, 31]. However, the VR-12 PCS has yet to be formally validated among any spine surgery procedures. Given the commonality of the ACDF procedure, the legacy data of the VR-12, and the adaptability of VR-12 to the PROMIS framework, having a comprehensive understanding of VR-12 scores in patients undergoing ACDF is of critical importance. Thus, the aim of the present study was to validate the VR-12 health survey with both legacy PROMs and the novel PROMIS metric for patients undergoing ACDF surgery.

Methods

Institutional review board approval (ORA #14051301) was obtained before reviewing a prospectively recorded surgical registry for eligible patients between January 2014 and August 2019. Included patients were required to have undergone an elective single- or multi-level ACDF for diagnosed degenerative spinal pathology. Exclusion criteria included patients that underwent surgery for infectious, traumatic, or malignant reasons and those that did not complete baseline pre-operative physical function surveys. Patients were recruited into the surgical registry at their initial clinic visit after a surgical intervention was determined. Patients were provided reminders via text and/or email to complete their PROM surveys. Administration of said surveys was via an online outcomes data collection software which patients accessed through their electronic device. Patients completed the following surveys: VR-12, SF-12, NDI, and PROMIS-PF in no preset order.

Patient charts were reviewed for demographic variables including gender, age, tobacco use, body mass index (BMI, < 30 kg/m^2 or ≥ 30 kg/m^2), comorbidity burden evaluated by Charlson comorbidity index (CCI), and diagnosis of pre-operative spinal pathology.

Peri-operative characteristics were recorded including operative duration (from skin incision to closure), estimated blood loss (EBL in mL), and hospital length of stay following surgery. Baseline and post-operative (e.g., 6 weeks, 12 weeks, 6 months, 1 year, and 2 years) survey scores were recorded at all timepoints for the following PROMs, VR-12 Physical composite score (PCS), PROMIS physical function (PF), SF-12 PCS, and NDI.

During the evaluated time period 244 potential candidates were identified. After the removal of 42 subjects who did not complete baseline VR-12 surveys, we were left with a total of 202 patients that underwent ACDF who were included in our cohort. The majority were male (58.9%), non-smokers (86.1%), non-obese (60.9%), and had a mean age of 49.5 years (Table 1). The most common pre-operative spinal pathology diagnosis was a herniated nucleus pulposus (HNP) (81.7%). The majority of ACDFs were single-level (56.4%) followed by two-level (35.2%) fusions. The mean time for procedures was 61.2 min, with a mean EBL of 31.8 mL, and a mean post-operative length of stay of 13.8 h.

The analysis was conducted utilizing Stata SE 16.1 (College Station, TX, USA). All of our overall cohort calculations were completed at pre-operative evaluation, and at 6 weeks, 12 weeks, 6 months, 1 year, and 2 years. We computed each timepoint mean, mean change (from pre-operative to each post-operative assessment), median, interquartile range, floor, and ceiling effects. A paired t test evaluated score improvement for within-patient change in scores from baseline to each post-operative timepoint at 6 weeks, 12 weeks, 6 months, 1 year, and 2 years. MCID was evaluated for all PROMs at all post-operative timepoints. The following MCID threshold values were used VR-12 PCS (2.5) [22], SF-12 PCS (8.1) [23], PROMIS-PF (4.5) [24], and NDI (10.0) [2, 41]. Scatterplots were used to visualize the relationship between VR-12 PCS with SF-12 PCS and PROMIS-PF at each timepoint. The strength of correlation between VR-12 PCS with SF-12 PCS, PROMIS-PF, and NDI was performed using a Pearson's correlation coefficient. The correlation strength was assessed by the following value categories for $|r|$: 0.00–0.20: very weak;

Table 1 Baseline characteristics of study population

	Total %, (*n*)
Age (mean ± SD, years)	49.5 ± 10.0
Gender	
Female	41.1% (83)
Male	58.9% (119)
Smoking Status (*n*)	
Non-smoker	86.1% (174)
Smoker	13.9% (28)
Body mass index (BMI)	
< 30 kg/m^2 - non-obese	60.9% (123)
≥ 30 kg/m^2 - obese	39.1% (79)
Charlson Comorbidity Index	1.4 ± 1.5
Spinal pathology	
Degenerative spondylolisthesis	4.8% (3)
Herniated nucleus pulposus	81.7% (165)
Degenerative disc disease	7.4% (15)
Stenosis	59.4% (120)
Foraminal stenosis	8.4% (17)
Operative levels	
1-level	56.4% (114)
2-level	35.2% (71)
3-level	7.4% (15)
4-level	1.0% (2)
Peri-operative characteristics	
Operative time (mean ± SD, min)	61.2 ± 19.3
Estimated blood loss (mean ± SD, mL)	31.8 ± 16.3
Hospital length of stay (mean ± SD, hours)	13.8 ± 12.2

SD standard deviation

0.21–0.40: weak; 0.41–0.60: moderate; 0.61–0.80: strong; and 0.81–1.00: very strong [9, 12]. Significance was set at an alpha = 0.05. Given the multiple correlation comparisons, we applied a Bonferroni correction factor. VR-12 PCS scores were correlated with 3 PROMs (NDI, SF-12 PCS, and PROMIS-PF) at six timepoints, and given an alpha of 0.05, the Bonferroni correction factor was (0.05/18 = 0.0028). *p* values less than this critical value suggest overall significance.

Results

The average baseline score for the VR-12 PCS was 36.2 ± 8.9. Mean baseline scores for SF-12 PCS, PROMIS-PF, and NDI were 34.4 ± 8.0, 39.6 ± 6.6, and 39.8 ± 19.6, respectively (Table 1). At the 6-week post-operative timepoint, NDI was the only PROM that had a significant improvement from baseline (− 8.2, $p < 0.001$, Table 2). All PROMs (VR-12 PCS, SF-12 PCS, PROMIS-PF, and NDI) demonstrated a statistically significant improvement from baseline from 12 weeks onward ($p \leq 0.001$ for 12 week, 6 month, 1 year, and 2 year points) when evaluated by paired *t* tests. Although NDI had the largest floor and ceiling effects, all were less than 15% at any single evaluation interval. Throughout the post-operative period, patients consistently increased the rate of achieving MCID with rates of 94.1% (VR-12 PCS), 84.7% (SF-12 PCS), 92.1% (PROMIS-PF), and 97.0% (NDI) by the 2-year timepoint.

When classifying all of our statistically significant correlation comparisons, most were very strong ($n = 53$, $p < 0.001$) or strong ($n = 58$, $p \leq 0.029$, Table 3, Figs. 1, 2, 3, 4, 5, 6). Overall correlations between VR-12 PCS and SF-12 PCS, PROMIS-PF and NDI had a *p* value less than the Bonferroni correction factor. A minority were medium ($n = 12$, $p \leq 0.027$) or weak ($n = 2$, $p \leq 0.026$), and none was very weak. When assessing instrument Pearson correlations among all patient subgroups, VR-12 PCS had the most "very strong" ($n = 41$) or "strong" ($n = 1$) correlation strengths with SF-12 PCS ($p < 0.001$), followed by PROMIS ("very strong," $n = 11$; "strong," $n = 29$; "medium," $n = 2$, $p \leq 0.009$) and NDI ("very strong," $n = 11$; "strong," $n = 29$; "medium," $n = 2$; "weak," $n = 2$, $p \leq 0.026$) scores. The VR-12 PCS correlation strength with PROMIS-PF generally increased in each subgroup, reaching its strongest correlation strength within each subgroup at the 2-year follow-up. VR-12 PCS correlation strength with NDI was the highest among single-level patients. The lowest statistically significant correlation strength values occurred between VR-12 PCS and NDI (Check for accuracy. Table shows 1 as strong (-0.704), 2 as moderatem and 2 as weak) among patients with severe disability or greater (NDI score ≥ 40).

Discussion

In our study of 202 patients, we observed that for each of the evaluated PROMs, patients attained statistically significant improvement by 12 weeks that sustained through 2 years following ACDF. In addition, a similar proportion of patients attained MCID for all PROMs out to 2 years. Our analysis determined that VR-12 PCS has a strong correlation with physical function metrics (SF-12 PCS, PROMIS-PF) and disability metrics (NDI) at both pre-operative and post-operative timepoints following ACDF. Even though the performance of these PROMs appears in line with one another, there are logistical and application-based differences between VR-12 and the other instruments.

Our study has limitations that need to be considered before drawing an ultimate conclusion. First, selection bias may be present given the retrospective nature in which our cohort was assembled. Additionally, the generalization of our results may be restricted due to the fact that all patients were treated at a single institution by a single surgeon. Third, patients who were lost to follow-up produced missing data points. From baseline, 126 patients who completed the pre-operative VR-12 survey were lost to follow-up at 1 year, and an additional 26 patients were lost to follow-up at 2 years. Subgroup analysis of these missing data points revealed no significant differences in the evaluated variables. Patients may not have returned to clinic for various reasons. If a patient had experienced tremendous improvement in their symptoms, they may not see a reason to spend their time for further post-operative evaluations. With the administration of multiple PROMs, there is a risk of patient-survey fatigue. Reducing the number of questions on an online survey has been reported to increase the odds of response (OR 1.73; CI: 1.40 to 2.13; $p = 0.08$) [10]. The questions at the end of a lengthy survey may be more prone to misclassification or measurement error [11]. Because we did not directly

Table 2 Post-operative changes in physical function scores

	Mean (mean ± SD)	Change (mean ± SD)	†*p* value*	*n*	Median (IQR)	Floor effect	Ceiling effect	% MCID[a]
VR-12 PCS								
Pre-operative	36.2 ± 8.9			202	35.5 (29.2–42.0)			
6 weeks	36.9 ± 9.5	0.8 ± 9.1	0.313	142	36.5 (31.0–43.3)	< 1%	< 1%	51.5%
12 weeks	41.0 ± 10.2	4.9 ± 8.7	*< 0.001*	111	41.1 (33.3–50.1)	< 1%	< 1%	66.3%
6 months	43.4 ± 10.6	7.9 ± 10.1	*< 0.001*	108	45.3 (34.6–52.9)	< 1%	< 1%	75.3%
1 year	46.2 ± 9.4	9.5 ± 9.9	*< 0.001*	76	49.2 (40.9–53.0)	< 1%	< 1%	86.6%
2 years	41.8 ± 12.0	6.2 ± 11.0	*< 0.001*	50	42.1 (34.2–53.0)	< 1%	< 1%	94.1%
SF-12 PCS								
Pre-operative	34.4 ± 8.0			202	33.0 (28.5–39.4)			
6 weeks	34.9 ± 8.4	0.6 ± 8.2	0.389	142	33.8 (29.2–39.7)	< 1%	< 1%	38.1%
12 weeks	38.5 ± 10.0	3.9 ± 8.8	*< 0.001*	111	36.1 (31.4–47.3)	< 1%	< 1%	59.9%
6 months	40.9 ± 10.6	6.9 ± 9.9	*< 0.001*	108	40.7 (31.6–50.6)	< 1%	< 1%	70.8%
1 year	43.8 ± 10.3	8.7 ± 10.0	*< 0.001*	76	45.8 (35.9–51.7)	< 1%	< 1%	83.7%
2 years	41.3 ± 11.6	7.4 ± 11.1	*0.001*	50	39.5 (32.3–51.7)	< 1%	< 1%	84.7%
PROMIS-PF								
Pre-operative	39.6 ± 6.6 (82)			115	38.6 (35.4–43.7)			
6 weeks	41.1 ± 7.1 (82)	1.5 ± 7.3	*0.065*	93	42.0 (34.7–47.2)	< 1%	< 1%	67.3%
12 weeks	44.9 ± 9.5 (67)	5.5 ± 8.0	*< 0.001*	80	45.2 (38.2–48.8)	< 1%	< 1%	79.7%
6 months	47.0 ± 9.5 (60)	7.4 ± 7.0	*< 0.001*	69	46.1 (39.5–51.0)	< 1%	< 1%	84.7%
1 year	47.8 ± 7.7 (50)	7.7 ± 7.4	*< 0.001*	56	48.0 (41.7–51.3)	< 1%	< 1%	90.1%
2 years	46.6 ± 9.2 (43)	7.3 ± 8.2	*< 0.001*	49	47.4 (40.2–51.2)	< 1%	< 1%	92.1%
NDI								
Pre-operative	39.8 ± 19.6			191	40.0 (26.0–54.0)			
6 weeks	31.6 ± 19.2	− 8.2 ± 17.3	*< 0.001*	177	30.0 (18.0–44.4)	< 1%	< 1%	53.5%
12 weeks	28.1 ± 20.4	− 13.4 ± 17.8	*< 0.001*	161	26.0 (12.0–40.0)	5%	< 1%	71.3%
6 months	22.4 ± 20.1	− 17.7 ± 19.5	*< 0.001*	146	18.0 (8.0–32.0)	10%	< 1%	84.7%
1 year	21.9 ± 19.2	− 15.4 ± 19.5	*< 0.001*	78	16.0 (6.0–30.0)	5%	< 1%	93.6%
2 years	27.6 ± 20.9	− 16.0 ± 18.1	*< 0.001*	38	22.0 (8.1–44.0)	4%	< 1%	97.0%

IQR interquartile range; *MCID*, minimal clinically important difference; *NDI*, Neck Disability Index; *PROMIS-PF*, Patient-Reported Outcomes Measurement Information System physical function; *SF-12*, Short-Form 12-item; *VR-12*, Veterans RAND 12-item Physical Component Summary

* Values in italics indicate statistical significance

† *p* value calculated using paired *t* test comparing scores at each timepoint to pre-operative values

[a] MCID threshold values were used VR-12 PCS (2.5) [23], SF-12 PCS (8.1) [24], PROMIS-PF (4.5) [25], and NDI (10.0) [2, 43]

evaluate the magnitude of VR-12 survey fatigue in the context of the other administered PROMs, future investigations should consider this and quantify the impact fatigue has on misclassification and subject responses. Another limitation of our study cohort was that we did not ensure a comprehensive distribution of baseline PROM scores and an equal range in disease severities. A wide range of PROM scores and symptoms would improve the validity of the VR-12 PCS performance. Therefore, to ameliorate these limitations, future investigations should aim to be prospective and multi-centered with cohort sizes large enough to account for inevitable patient dropouts.

PROMIS-PF and VR-12 PCS represent two of the most frequently used physical function measures. In general, VR-12 has numerous significant advantages stemming from its open availability, ease of administration, and nearly 2 decades of clinical use [20, 27]. One often cited advantage of both the VR and PROMIS surveys is that they are not disease-specific and are able to assess global health domains. While this is theoretically helpful to compare patients from differing populations or across different disease conditions, it can be difficult to relate two global metrics due to the vast nature of assessing global health, along with nuanced differences in question item structure, syntax, semantics, and pragmatics [27]. Like the VR-12, the short-form versions of PROMIS are available online free of cost [21]. Also similar to the VR-12, PROMIS short-form surveys require respondents to answer all questions, though the accuracy of PROMIS short forms has been observed to vary based on the number of items used (e.g., 4-, 6- or 8-item surveys). While information is scarce regarding guidance for when investigators might select specific PROMIS survey lengths, the VR-12 is always administered according to the same item number standard.

In order to maintain comparability with the most recent literature, MCID values were utilized from current spine and quality of life outcome research in an effort to apply the most applicable patient populations. Our patient population achieved a statistically significant improvement in all of the assessed PROMs by 12 weeks, and with NDI by 6 weeks. When comparing our pre- and post-operative lumbar VR-12 PCS mean scores to other investigations, Gornet et al. reported VR-12 PCS pre-operative scores of 25.63 ± 6.45 (imputed), with post-operative scores assessed at a 1- or 2-year follow-up of 35.37 ± 11.94 [13]. The overall score change observed by Gornet et al. was 9.73 ± 11.97, well within one standard deviation of our VR-12 PCS score improvements at both 1- and 2-year follow-ups. Our measurements of SF-12 PCS were observed to have the lowest overall MCID achievement at each timepoint, which is similar to the

Table 3 Association of VR-12 PCS with SF-12 PCS, PROMIS-PF, and NDI

	Overall		Single-level		Multi-level		HNP		Stenosis		NDI < 40		NDI ≥ 40	
	r	*p* value	*r*	*p* value	*r*	*p* value	*r*	*p* value	*r*	*p* value	*r*	*p* value	*r*	*p* value
VR-12 vs. SF-12														
Pre-operative	*0.950*	*< 0.001*	*0.960*	*< 0.001*	*0.937*	*< 0.001*	*0.947*	*< 0.001*	*0.951*	*< 0.001*	*0.937*	*< 0.001*	*0.958*	*< 0.001*
6-week	*0.886*	*< 0.001*	*0.841*	*< 0.001*	*0.961*	*< 0.001*	*0.952*	*< 0.001*	*0.826*	*< 0.001*	*0.828*	*< 0.001*	*0.960*	*< 0.001*
12-week	*0.965*	*< 0.001*	*0.970*	*< 0.001*	*0.956*	*< 0.001*	*0.967*	*< 0.001*	*0.960*	*< 0.001*	*0.968*	*< 0.001*	*0.956*	*< 0.001*
6-month	*0.973*	*< 0.001*	*0.971*	*< 0.001*	*0.974*	*< 0.001*	*0.976*	*< 0.001*	*0.975*	*< 0.001*	*0.967*	*< 0.001*	*0.978*	*< 0.001*
1-year	*0.886*	*< 0.001*	*0.826*	*< 0.001*	*0.975*	*< 0.001*	*0.850*	*< 0.001*	*0.976*	*< 0.001*	*0.970*	*< 0.001*	*0.774*	*< 0.001*
2-year	*0.986*	*< 0.001*	*0.992*	*< 0.001*	*0.981*	*< 0.001*	*0.982*	*< 0.001*	*0.982*	*< 0.001*	*0.985*	*< 0.001*	*0.986*	*< 0.001*
VR-12 vs. PROMIS-PF														
Pre-operative	*0.729*	*< 0.001*	*0.730*	*< 0.001*	*0.722*	*< 0.001*	*0.742*	*< 0.001*	*0.719*	*< 0.001*	*0.667*	*< 0.001*	*0.623*	*< 0.001*
6-week	*0.740*	*< 0.001*	*0.699*	*< 0.001*	*0.819*	*< 0.001*	*0.719*	*< 0.001*	*0.736*	*< 0.001*	*0.647*	*< 0.001*	*0.765*	*< 0.001*
12-week	*0.714*	*< 0.001*	*0.799*	*< 0.001*	*0.568*	*0.002*	*0.676*	*< 0.001*	*0.619*	*< 0.001*	*0.627*	*< 0.001*	*0.653*	*< 0.001*
6-month	*0.804*	*< 0.001*	*0.816*	*< 0.001*	*0.791*	*< 0.001*	*0.794*	*< 0.001*	*0.802*	*< 0.001*	*0.760*	*< 0.001*	*0.585*	*0.009*
1-year	*0.770*	*< 0.001*	*0.792*	*< 0.001*	*0.771*	*< 0.001*	*0.769*	*< 0.001*	*0.761*	*< 0.001*	*0.665*	*< 0.001*	*0.842*	*< 0.001*
2-year	*0.858*	*< 0.001*	*0.855*	*< 0.001*	*0.892*	*< 0.001*	*0.878*	*< 0.001*	*0.882*	*< 0.001*	*0.789*	*< 0.001*	*0.957*	*< 0.001*
VR-12 vs. NDI														
Pre-operative	*− 0.659*	*< 0.001*	*− 0.638*	*< 0.001*	*− 0.692*	*< 0.001*	*− 0.655*	*< 0.001*	*− 0.671*	*< 0.001*	*− 0.454*	*< 0.001*	*− 0.336*	*0.005*
6-week	*− 0.611*	*< 0.001*	*− 0.620*	*< 0.001*	*− 0.609*	*< 0.001*	*− 0.626*	*< 0.001*	*− 0.552*	*< 0.001*	*− 0.531*	*< 0.001*	*− 0.489*	*0.001*
12-week	*− 0.725*	*< 0.001*	*− 0.758*	*< 0.001*	*− 0.669*	*< 0.001*	*− 0.699*	*< 0.001*	*− 0.713*	*< 0.001*	*− 0.689*	*< 0.001*	*− 0.575*	*< 0.001*
6-month	*− 0.642*	*< 0.001*	*− 0.720*	*< 0.001*	*− 0.534*	*0.001*	*− 0.654*	*< 0.001*	*− 0.611*	*< 0.001*	*− 0.708*	*< 0.001*	*− 0.393*	*0.026*
1-year	*− 0.731*	*< 0.001*	*− 0.839*	*< 0.001*	*− 0.546*	*0.003*	*− 0.792*	*< 0.001*	*− 0.651*	*< 0.001*	*− 0.720*	*< 0.001*	*− 0.704*	*0.001*
2-year	*− 0.593*	*0.001*	*− 0.613*	*0.012*	*− 0.602*	*0.029*	*− 0.482*	*0.027*	*− 0.592*	*0.006*	*− 0.749*	*< 0.001*	− 0.364	0.271

HNP, herniated nucleus pulposus; *NDI*, Neck Disability Index; *PROMIS-PF*, Patient-Reported Outcomes Measurement Information System physical function; *SF-12*, Short-Form 12-item; *VR-12*, Veterans RAND 12-item Physical Component Summary

[*] Values in italics indicate statistical significance

[†] *p* value calculated using Pearson's correlation

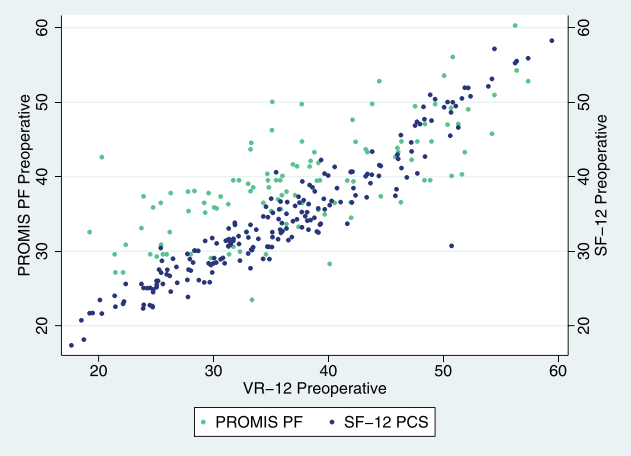

Fig. 1. ACDF pre-operative association of VR-12 with SF-12 and PROMIS

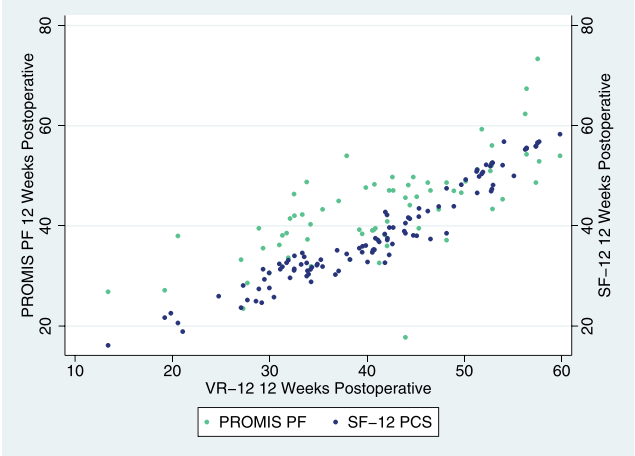

Fig. 3. ACDF 12-week post-operative association of VR-12 with SF-12 and PROMIS.

findings of others among cervical surgery patients [40]. It is also important to note that the VR-36 [21], VR-12 [37], SF-12 [43], SF-36 [28], and PROMIS [15] are all centered with a mean of 50 and a standard deviation of 10 points. Despite these similarities, the commonly used SF-12 PCS MCID value of 8.1 can be twice the magnitude of MCID thresholds that are established for the SF-36 questionnaire [3, 5, 8]. Several MCID values have been reported for VR-12 PCS outcomes. In one large cohort ($n = 7902$) of elective patients, researchers established a VR-12 MCID threshold for the PCS of 2.5 [26], which is nearly one-third that of the SF-12 PCS MCID.

When comparing the MCID achievement rates between PROMIS-PF and NDI, our results are similar to the findings of others in that early post-operative MCID rates were higher among the PROMIS-PF evaluations, though NDI evaluations had the highest cumulative achievement rate by the final 2-year follow-up [40]. VR-12 PCS MCID achievement was lower than both PROMIS-PF and NDI at all timepoints until the 2-year follow-up. At 2-year follow-up, it was 94.1% which was between the analogous values of 92.1% and 97.0% for PROMIS-PF and NDI, respectively.

Estimation of utility scores for the use in costs per quality adjusted life years (QALYs) from the VR-12 instrument has been demonstrated using a derivation known as the VR-6D [38]. VR-6D provides users of VR-12 an ability to calculate QALYs and has been reported to have comparable performance with the more established Brazier SF-6D [38]. There is a scarcity of research evaluating the use of VR-6D in spine surgery patients, thus providing room to analyze its usefulness in cervical surgery QALY investigations.

It is difficult to draw direct comparisons of our correlation values to those of other validation studies because few studies have correlated VR-12 with PROMIS, NDI, or SF-12 PROMs. In one similar comparison among spine clinic patients, Lapin et al. observed correlation strengths of 0.73 between VR-12 and PROMIS-10 scores [27]. This study differed from ours in that it utilized only clinic patients, compared "cross-walk" PROM conversion tools, only examined one clinical visit, and utilized a 10-item short-form

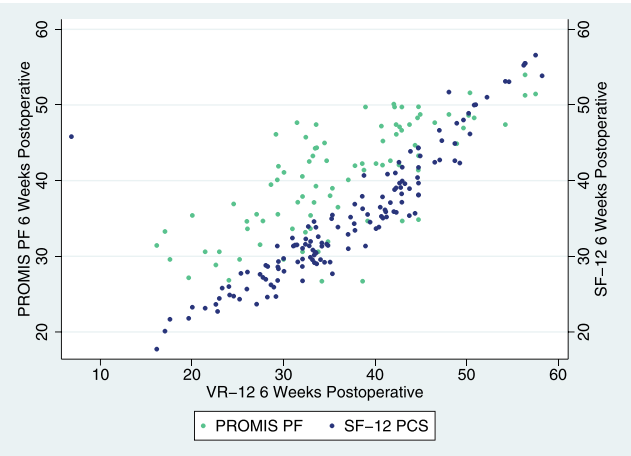

Fig. 2. ACDF 6-week post-operative association of VR-12 with SF-12 and PROMIS.

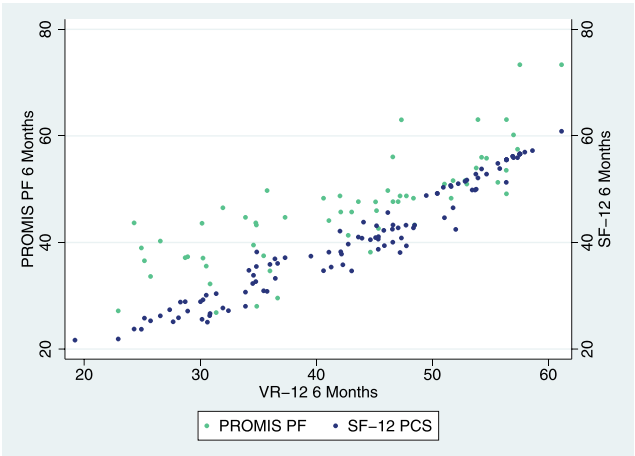

Fig. 4. ACDF 6-month post-operative association of VR-12 with SF-12 and PROMIS.

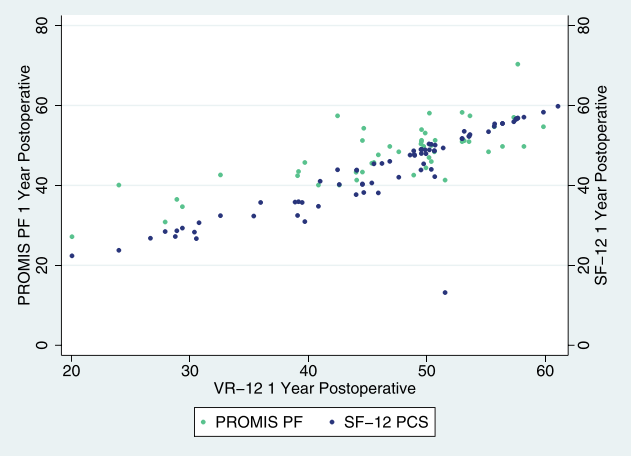

Fig. 5. ACDF 1-year post-operative association of VR-12 with SF-12 and PROMIS.

of the PROMIS-PF questionnaire. With the exception of our 12-week post-operative correlation ($r = 0.72$), all of our correlations were approximately equal or greater than this value. Although spine surgery population VR-12 PCS scores have not been previously correlated with PROMIS, NDI, or SF-12, numerous studies have reported correlations between PROMIS-PF, NDI, SF-12, and SF-36 [17, 24, 30, 32, 34, 40].

Correlations between NDI and PROMIS are typically reported to be "moderate" or "strong" [4, 17, 24, 30, 32, 34, 40]. Caution should be used in interpreting correlation classifications. For example, different investigations can use varied cutoffs for correlations that are labeled as "strong," e.g., $r \geq 0.5$ [14, 24], $r \geq 0.6$ [34], $r \geq 0.7$ [32], or $r \geq 0.8$ [4, 17]. We observed the lowest VR-12 PCS correlations to occur with NDI during the pre-operative evaluation ($|r| = 0.659$), at 6 weeks ($|r| = 0.611$), and at 2 years ($|r| = 0.593$). It is difficult to ascertain exactly why this occurred, as this has been hypothesized to occur for several reasons. First, global assessments may have limited detection of disease-specific symptoms encountered with degenerative cervical spine pathology. Many of these are most severe during the pre-operative or early follow-up time periods. Second, disease-specific instruments, such as NDI, may be limited in their ability to detect a broader spectrum of symptoms such as difficulties with ambulation, radiculopathy, or overall functional disturbances.

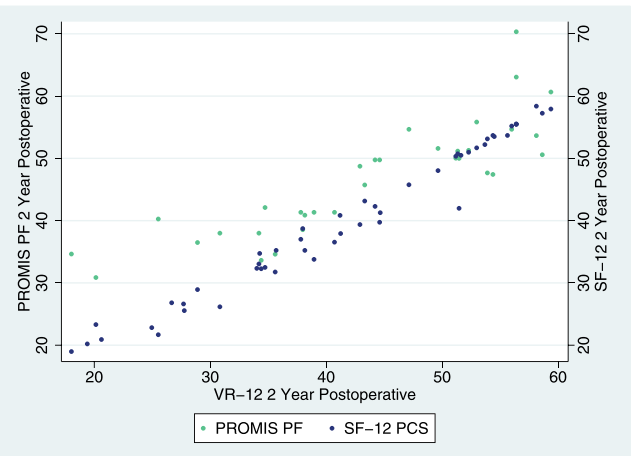

Fig. 6. ACDF 2-year post-operative association of VR-12 with SF-12 and PROMIS.

Following subanalysis, VR-12 PCS and SF-12 PCS correlations for single-level and multi-level procedures demonstrated very strong correlations, HNP and stenosis had very strong correlations, and NDI < 40 had very strong correlations, while NDI ≥ 40 had very strong correlations for all but the 1-year timepoint which was strong. VR-12 PCS and PROMIS-PF had similar correlations although more strong than very strong, with a few moderate correlations reported for multi-level and NDI ≥ 40. VR-12 PCS and NDI had lower strength correlations compared with the other two PROMs, ranging from moderate to very strong, with NDI ≥ 40 having weak correlations at the pre-operative, 6 month, and 2 year timepoints.

In conclusion, VR-12 PCS was strongly correlated with physical function (SF-12 PCS and PROMIS-PF) and disability (NDI) metrics following ACDF out to 2 years. Additionally, all evaluated PROMs demonstrated significant improvement from pre-operative baseline at all post-operative timepoints. Our study provides evidence for utilizing the VR-12 PCS instrument as a valid measure of physical function both pre- and post-operatively among patients undergoing ACDF. We recommend the VR-12 PCS instrument as a suitable alternative to PROMIS-PF or SF-12 PCS in cervical surgery patients undergoing single- or multi-level procedures, with a range of possible spinal pathologies and symptom severity.

Compliance with Ethical Standards

Conflict of Interest: Nathaniel W. Jenkins, MS, James M. Parrish, MPH, Michael T. Nolte, MD, Nadia M. Hrynewycz, BS, and Thomas S. Brundage, BS, declare that they have no conflicts of interest. Kern Singh, MD, reports personal fees and royalties from Zimmer Biomet, royalties from Stryker, RTI Surgical, and Lippincott Williams and Wilkins, stock ownership from Avaz Surgical LLC, stock ownership and board membership from Vital 5 LLC, personal fees from K2M, non-financial support and board membership from TDi LLC, non-financial support from Minimally Invasive Spine Study Group, Contemporary Spine Surgery, Orthopedics Today, and Vertebral Columns, and grants from Cervical Spine Research Society, outside the submitted work.

Human/Animal Rights: All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2013.

Informed Consent: Informed consent was waived] from all patients for being included in this study.

Required Author Forms Disclosure forms provided by the authors are available with the online version of this article.

References

1. Andresen AK, Paulsen RT, Busch F, Isenberg-Jørgensen A, Carreon LY, Andersen MØ. Patient-reported outcomes and patient-reported satisfaction after surgical treatment for cervical radiculopathy. *Global Spine J*. 2018;8:703–708.
2. Anon. *AAOS File Uploads—NDI*. Available at: https://www.aaos.org/uploadedFiles/NDI.pdf. Accessed 6 Apr 2020.
3. Auffinger B, Lam S, Shen J, Thaci B, Roitberg BZ. Usefulness of minimum clinically important difference for assessing patients with subaxial degenerative cervical spine disease: statistical versus substantial clinical benefit. *Acta Neurochir.* 2013;155:2345–2355.
4. Boody BS, Bhatt S, Mazmudar AS, Hsu WK, Rothrock NE, Patel AA. Validation of Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests in cervical spine surgery. *J Neurosurg Spine*. 2018;28:268–279.
5. Carreon LY, Glassman SD, Campbell MJ, Anderson PA. Neck Disability Index, short form-36 physical component summary, and pain scales for neck and arm pain: the minimum clinically important difference and substantial clinical benefit after cervical spine fusion. *Spine J.* 2010;10:469–474.
6. Carter Clement R, Bhat SB, Clement ME, Krieg JC. Medicare reimbursement and orthopedic surgery: past, present, and future. *Curr Rev Musculoskelet Med.* 2017;10:224–232.
7. Chow A, Mayer EK, Darzi AW, Athanasiou T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery*. 2009;146:435–443.
8. Chung AS, Copay AG, Olmscheid N, Campbell D, Walker JB, Chutkan N. Minimum clinically important difference: current trends in the spine literature. *Spine*. 2017;42:1096–1105.
9. Cohen J. CHAPTER 4 - Differences between Correlation Coefficients. In: Cohen J, ed. Statistical Power Analysis for the Behavioral Sciences. Academic Press; 1977:109–143.
10. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst. Rev.* 2009:MR000008.
11. Egleston BL, Miller SM, Meropol NJ. The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects. *Stat Med.* 2011;30:3560–3572.
12. Evans JD. Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co; 1996.
13. Gornet MF, Copay AG, Sorensen KM, Schranck FW. Assessment of health-related quality of life in spine treatment: conversion from SF-36 to VR-12. *Spine J.* 2018;18:1292–1297.
14. Haws BE, Khechen B, Bawa MS, et al. The Patient-Reported Outcomes Measurement Information System in spine surgery: a systematic review. *J. Neurosurg. Spine*. 2019;30:405–413.
15. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res.* 2009;18:873–880.
16. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. *Health Econ.* 1993;2:217–227.
17. Hung M, Saltzman CL, Voss MW, et al. Responsiveness of the Patient-Reported Outcomes Measurement Information System (PROMIS), Neck Disability Index (NDI) and Oswestry Disability Index (ODI) instruments in patients with spinal disorders. *Spine J.* 2019;19:34–40.
18. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. *Controlled Clinical Trials*. 1989;10:407–415.
19. Jones D, Kazis L, Lee A, Rogers W, Skinner K, Cassar L, Wilson N, Hendricks A. Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. *J Ambul Care Manage.* 2001;24:68.
20. Kazis LE, Lee A, Spiro A 3rd, et al. Measurement comparisons of the medical outcomes study and veterans SF-36 health survey. *Health Care Financ Rev.* 2004;25:43–58.
21. Kazis LE, Miller DR, Clark JA, et al. Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. *J Ambul Care Manage.* 2004;27:263–280.
22. Kazis LE, Ren XS, Lee A, Skinner K, Rogers W, Clark J, Miller DR. Health status in VA patients: results from the Veterans Health Study. *Am J Med Qual.* 1999;14:28–38.
23. Kazis LE, Selim A, Rogers W, Ren XS, Lee A, Miller DR. Dissemination of methods and results from the Veterans health study: final comments and implications for future monitoring strategies within and outside the Veterans healthcare system. *J Ambul Care Manage.* 2006;29:310–319.
24. Khechen B, Patel DV, Haws BE, et al. Evaluating the concurrent validity of PROMIS physical function in anterior cervical discectomy and fusion. *Clin Spine Surg*. 2019;32(10):449–453. doi: https://doi.org/10.1097/BSD.0000000000000786.
25. Khor S, Lavallee D, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surgery*. 2018;153:634.
26. Kronzer VL, Jerry MR, Ben Abdallah A, et al. Changes in quality of life after elective surgery: an observational study comparing two measures. *Qual Life Res.* 2017;26:2093–2102.
27. Lapin BR, Kinzy TG, Thompson NR, Krishnaney A, Katzan IL. Accuracy of linking VR-12 and PROMIS global health scores in clinical practice. *Value Health*. 2018;21:1226–1233.
28. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Joint Surg Am.* 2015;97:1628–1634.
29. McGirt MJ, Speroff T, Dittus RS, Harrell FE Jr, Asher AL. The National Neurosurgery Quality and Outcomes Database (N2QOD): general overview and pilot-year project description. *Neurosurg Focus*. 2013;34:E6.
30. Moses MJ, Tishelman JC, Stekas N, et al. Comparison of Patient Reported Outcome Measurement Information System With Neck Disability Index and visual analog scale in patients with neck pain. *Spine* . 2019;44:E162–E167.
31. Oglesby M, Fineberg SJ, Patel AA, Pelton MA, Singh K. Epidemiological trends in cervical spine surgery for degenerative diseases between 2002 and 2009. *Spine*. 2013;38:1226–1232.
32. Paulino Pereira NR, Janssen SJ, et al. Most efficient questionnaires to measure quality of life, physical function, and pain in patients with metastatic spine disease: a cross-sectional prospective survey study. *Spine J.* 2017;17:953–961.
33. Peolsson A, Peolsson M. Predictive factors for long-term outcome of anterior cervical decompression and fusion: a multivariate data analysis. *Eur Spine J.* 2008;17:406–414.
34. Purvis TE, Andreou E, Neuman BJ, Riley LH 3rd, Skolasky RL. Concurrent validity and responsiveness of PROMIS health domains among patients presenting for anterior cervical spine surgery. *Spine*. 2017;42:E1357–E1365.
35. Recinos PF, Dunphy CJ, Thompson N, Schuschu J, Urchek JL 3rd, Katzan IL. Patient satisfaction with collection of patient-reported outcome measures in routine care. *Adv Ther*. 2017;34(2):452–465. doi: https://doi.org/10.1007/s12325-016-0463-x.
36. Schalet BD, Rothrock NE, Hays RD, et al. Linking physical and mental health summary scores from the veterans RAND 12-Item Health Survey (VR-12) to the PROMIS Global Health Scale. *J Gen Intern Med.* 2015;30:1524–1530.
37. Selim AJ, Rogers W, Fleishman JA, et al. Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Qual Life Res.* 2009;18:43–52.
38. Selim AJ, Rogers W, Qian SX, Brazier J, Kazis LE. A preference-based measure of health: the VR-6D derived from

the veterans RAND 12-Item Health Survey. *Qual Life Res.* 2011;20:1337–1347.
39. Steinhaus ME, Iyer S, Lovecchio F, et al. Which NDI domains best predict change in physical function in patients undergoing cervical spine surgery? *Spine J.* 2019;19:1698–1705.
40. Vaishnav AS, Gang CH, Iyer S, McAnany S, Albert T, Qureshi SA. Correlation between NDI, PROMIS and SF-12 in cervical spine surgery. *Spine J.* 2020;20:409–416.
41. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther.* 1991;14:409–415.
42. Ware JE Jr. SF-36 health survey update. *Spine*. 2000;25:3130–3139.
43. Ware JE, Kosinski M, Keller SD. How to score the SF-12 physical and mental health summary scales. Lincoln: QualityMetric. 1998.