



# Comparing Quantification of Pain Severity by Verbal Rating and Numeric Rating Scales

Marcel Dijkers, PhD

Mount Sinai School of Medicine, New York, New York

Received August 19, 2008; accepted April 9, 2010

## Abstract

**Background:** Researchers have reported widely varying correlations among the 3 main instruments used to quantify pain severity, Visual Analog Scale (VAS), Verbal Rating Scale (VRS), and Numeric Rating Scale (NRS), both at the level of groups and at the level of individuals.

**Objective:** To assess the comparability of reports of pain severity using a VRS and a NRS in a spinal cord injury (SCI) sample.

**Methods:** Data were taken from a longitudinal observational study. Patients were 168 individuals with new traumatic SCI admitted for inpatient rehabilitation who completed the VRS and NRS multiple times, each time for multiple pains as appropriate.

**Results:** For 1,114 ratings of pain, VRS and corresponding NRS ratings were correlated weakly (Spearman correlation,  $\rho = 0.38$ ). For 36 individuals with at least 10 completions of paired VRS and NRS,  $\rho$  ranged from  $-0.55$  to  $0.76$ . Variation in NRS rating for each VRS adjective was reduced by about 25% when between-patient variation was eliminated. Mean NRS ratings by VRS adjective, for patients who had used each of at least 2 adjectives at least 5 times each, showed large differences in mean NRS scores between individuals using the same VRS adjective.

**Conclusion:** There are considerable differences between individuals in how NRS and VRS are used; there also seem to be individuals whose understanding of the meaning of the VRS adjectives is completely different from what was assumed by the creators of this VRS. Both VRS and NRS data must be used with extreme caution by SCI clinicians and researchers.

*J Spinal Cord Med. June 2010;33(3):232–242*

**Key Words:** Spinal cord injuries; Pain, neuropathic, musculoskeletal, central; Reproducibility; Pain measurement; Visual Analog Scale; Verbal Rating Scale; Numeric Rating Scale

## INTRODUCTION

Pain is a significant problem for many people with spinal cord injury (SCI), often starting soon after injury (1–3) and continuing throughout the person's life (4). In addition to central pain due to injury to the central nervous system itself, musculoskeletal pain resulting from unusual demands on the body (transfers, wheeling) is commonly reported (5–7). Because it is unclear what

percentage of the SCI population is affected by chronic pain (reports vary from 26% to 96%) (8) and how function and quality of life are affected by pain, SCI pain has increasingly become the subject of scholarly attention of SCI and pain specialists. This is evidenced by the recent publication of 2 reviews of measures for the classification and quantification of pain in the SCI population (9,10) and of a proposal for a "Basic SCI Pain Data Set" offering guidelines on how specifics of the pain experience are to be recorded in research and clinical records (11).

Key among pain characteristics is severity. Several types of instruments are used to convert the internal pain severity experience into a number that can be manipulated in research (12,13). The 3 most commonly used are the Visual Analog Scale (VAS), the Numeric Rating Scale (NRS) (selected for the Basic SCI Pain Data Set) (11), and the Verbal Rating Scale (VRS). A VAS is a horizontal or vertical line, most often 10 cm long, and marked at the extremes with "no pain" and "worst pain imaginable" or

---

Please address correspondence to Marcel Dijkers, PhD, Rehabilitation Medicine, Mount Sinai School of Medicine, One Gustave Levy Place, Box 1240, New York, NY 10029-6574; p: 212 659 8587; f: 212 348 5901 (e-mail: marcel.dijkers@msnyuhealth.org).

This work was supported in part by grants H133N000027 and H133N060027 from the National Institute on Disability and Rehabilitation Research (NIDRR), Office of Special Education Services, US Department of Education to Mount Sinai School of Medicine, New York, New York.

similar phrases. The patient is asked to place a mark on the line that represents his/her pain level. A ruler is used to translate the information to a score ranging from 1 to 100. VAS scales are presumed to produce ratio-level data.

A NRS has similar anchors at the extremes but offers numbers from 0 through 10 (sometimes 0–20 or 0–100) to represent pain severity levels from none to “most intense pain imaginable.” Patients circle the number corresponding to their pain severity. Traditionally, a NRS has been assumed to offer measurement at the ratio level. A VRS consists of a series of adjectives reflecting degrees of pain severity, arranged from “no pain” to whatever word or phrase is used to designate the most extreme pain. A pain severity VRS may have 4 or more gradations; the specific adjectives used vary from version to version. Patients circle or put a checkmark next to the adjective that best describes how severe their pain is. The numbers that generally are provided along with the adjectives (0–3, 1–5, or something similar) are used in processing the information. VRSs offer measurement at the ordinal level, although not a few investigators have treated them as providing interval level data. [More detail on the 3 types of measures, as well as their metrologic properties, may be found in a review by Bryce and colleagues (9).] The only study in SCI to contribute to the metrologic evidence is by Lund and colleagues (14).

Over the years, a number of studies have investigated the relationship between patients’ rating of their pain on 2 or more of these pain severity operationalizations (14–24). The relationship between the VRS and VAS has had the most attention.

Woodforde and Merskey reported a high correlation (type unknown) between a VAS and a 5-point VRS: 0.87 for 14 men and 0.83 for 13 women with a variety of pain syndromes (15). However, Reading found Pearson correlations of 0.29 and 0.26, respectively, between a 5-point VRS and the VAS and NRS for women who had given birth after an episiotomy. On a second day, the 26 women involved reported less pain, but correlations improved to 0.71 and 0.57, respectively (16). Jones and colleagues reported correlations (type unknown) of at least 0.80 between a 4-point VRS and a NRS for 2 samples of elderly nursing home residents (N = 135 each) (17). For 113 Mexican patients with arthritis, Clark et al reported a Spearman correlation of 0.79 between a 5-point VRS and a VAS; upon retest a week later, the correlation was 0.74 (19). For 206 patients with osteoarthritis, Averbuch and Katzper reported a Pearson correlation of 0.71 between a VAS and a 5-point VRS (20).

Interesting results were reported from 2 studies that asked patients to rate their pain on 2 pain intensity instruments multiple times. Linton and Götestam had 15 patients with chronic back or joint pain complete a 6-point “behaviorally defined” VRS and a VAS daily for 1 to 2 weeks. Pearson correlations for individual patients

ranged from  $-0.34$  to  $0.96$ , with a mean of  $0.44$  (18). Low or negative correlations were noted for patients with little day-to-day variation in pain level on the VRS, which is not surprising because in the presence of restriction in range, correlations tend to be small. Breivik et al reported Spearman correlations ranging from  $0.48$  to  $0.97$  (median:  $0.89$ ) for 35 oral surgery patients who rated their pain using both a 4-point VRS and a VAS (21).

From these reports, it is unclear why there is so much variation in the reported correlations between VRS and NRS/VAS. The causes may be the type of correlation used (Pearson vs Spearman), the number of categories on the VRS and the specific anchoring adjectives used, the range in pain levels present in the sample, or other sample or study methodologic characteristics.

The lack of correspondence between NRS (or VAS, for that matter) and VRS has generally been explained by the fact that the VRS adjectives have different meanings for different people (14,24). If that is the case, one would expect a higher correspondence between NRS and VRS for repeated pain ratings by one and the same person than is found for a single set of ratings by many different persons. The goal of this study was to test, in a sample of patients with SCI pain, the hypothesis that within-individual correlations are higher than between-individual correlations. Because the increase in the size of the correlation was found to be fairly modest, the use of VRS adjectives and NRS ratings by the participants was further explored.

## METHODS

As part of a prospective study on the development of various types of chronic pain after SCI, detailed data were collected on the nature and severity of pain(s), if any, experienced by 185 newly injured rehabilitation inpatients with SCI. All patients who gave informed consent completed a battery of pain-related measures on admission to the rehabilitation unit and at discharge from the unit. If they reported having one or more pain components, they completed a simpler pain questionnaire approximately weekly during their stay. They were followed for a year after discharge with about quarterly interviews (the majority done by telephone) that elicited the same information. Each time, for every reported pain component, a form was completed that included the Short-Form McGill pain questionnaire (25) as modified for SCI by Turner and Cardenas (26), a VRS and a NRS. A pain component was defined to the patients as pain in a different anatomic locale, which by severity, nature, etc, appeared to be separate from any other pain or pains they experienced. When a patient reported more than one pain, all ratings for the first pain component were made and all questions about it answered before the same set of ratings and questions was completed for the next pain.

After obtaining ratings of the Short-Form McGill adjectives, the research assistant who collected the data

provided the patient with a copy of the NRS and stated, "Now I would like to know how severe **this pain** is at the moment. Use a scale that runs from 0: no pain at all, to 10: the worst possible pain you can imagine. How bad would you say this pain is on this scale from 0 to 10?" The NRS presented the numbers 0 through 10 arranged horizontally and labeled at the extremes with "no pain" and "worst possible pain." When patients insisted that their pain level could be expressed on the NRS only by using a half-value (for instance, 3.5) in spite of suggestions that such a level of detail was unnecessary, the decimal value was recorded; in the present analysis, all of these have been rounded up to the nearest integer.

The form incorporated the McGill Pain Questionnaire's (MPQ) VRS, which consists of 6 pain severity descriptors: "no pain," "mild," "discomforting," "distressing," "horrible," and "excruciating" (27). After obtaining the NRS rating and asking 4 interpolated questions aimed at determining to what degree the pain in question was neuropathic in nature, the research assistant would present the patient with a card listing the adjectives in the order given and state, "I am going to read you a number of words that people sometimes use to describe how distressing their pain is. Please tell me how distressing **this pain** is by selecting the word that best describes the distress you feel." (During interviews conducted after discharge from the unit, patients were asked to refer to 2 cards that had been mailed to them.)

Because a patient could rate multiple pains at the same time, a single pain continuously or several times during successive weeks, or both, many patients provided multiple combinations of the VRS and the NRS. Their pains might vary from one another or over time in severity, offering an opportunity to investigate covariation of VRS and NRS at the level of the individual, similar to the studies referred to previously (18,21). The responses of patients who completed at least 10 combinations were selected for further analysis of the relationship between VRS and NRS ratings.

The statistical analysis applied to the data used counts for the VRS, means, and SDs for the NRS, as well as Spearman correlations ( $\rho$ ) for the association between NRS and VRS. Counts, means, SDs, and  $\rho$  were calculated at the group and the individual level. No inferential statistical tests are used, because in all crucial statistical tests to be applied, the assumption of independence of observations would be violated. In the exploratory part, statistical testing would be inappropriate.

## RESULTS

Altogether 164 persons (87% of 185) reported pain at least once. Of these, 36 completed the form 10 or more times for different pains (the highest number reported for a single day was 6; the median was 1) and/or subsequent weeks. This subgroup produced 586 pairs of ratings, or 16 pairs on average. The 128 patients with fewer than 10 pairs of ratings completed 529 forms combined, or 4

pairs on average. Demographic and injury information on these 2 groups and those who never reported pain is provided in Table 1. The 3 groups are fairly comparable with one another, except, not surprisingly, on length of stay: the average length of stay was shorter for the "no pain ever reported" group than for the "under 10 pain reports" group, the average length of stay for which was shorter than that for the "10+ pain reports" group.

Of the more than 2,000 pairs of pain ratings, 9% were completed at admission, 7% at discharge, 59% during the interim, and 25% during follow up. There was an association between the VRS and NRS ratings for the sample overall and for the subgroups of people who completed fewer than 10 vs 10 or more sets of ratings.  $\rho$  was 0.28 for the former group and 0.46 for the latter. It was 0.38 for the entire sample. Figure 1 shows considerable overlap in terms of the NRS values that the patients associated with each VRS category.

For individuals in the group of 36, the  $\rho$  value for the correlation between VRS and NRS ranged from a low of  $-0.55$  to a high of  $0.76$  (Table 2). The median was  $0.23$ , and the mean of the absolute values of  $\rho$  was  $0.32$ . Both of these numbers are less than the  $\rho$  value of  $0.46$  for the entire group of 36, but that presumably is because the average patient used the extremes of the 6 VRS categories and the 11 NRS scale points less often than did the group as a whole. (All other things being equal, a restriction of the range of one or both variables involved in a correlation results in a smaller correlation value.)

A better measure of group vs individual NRS variability by VRS category is provided by the SD of the NRS for each VRS category and its reduction when individual patients rather than the group as a whole are considered. Relevant data are provided in Table 3. Column 1 provides the mean of the NRS ratings by VRS adjective, and column 2 shows the corresponding SD across all completions by all 36 patients with 10 or more pairs. These are the data plotted by the dashed line in Figure 1. If NRS mean and SD is calculated for each patient and these parameters are averaged across all 36 patients, the values shown in columns 4 and 6, respectively, are obtained. (The number of cases in column 5 or 7 is not equal to 36, because not all patients used all VRS adjectives; for instance, "discomforting," the most popular VRS answer category, was used by 34 of 36. The number of cases in column 7 is smaller than the number in column 5, because patients who have only one NRS rating within a VRS category do not contribute to the mean SD, as their ratings do not have a SD.) The ratio of the mean SD thus calculated to the SD based on the ratings of the group as a whole (column 6 compared with column 2) is provided in column 8; these values suggest that for the average patient, the SD for NRS ratings corresponding to a chosen VRS adjective is about 75% of the SD that is obtained without paying attention to differences in rating tendencies between individuals. In columns 4 through 8, all individuals have the same

**Table 1.** Demographic and Injury Characteristics by Pain-Reporting Category

	Pain Never Reported	<10 Pain Ratings	≥10 Pain Ratings	Total
No. of patients <sup>a</sup>	21	128	36	185
Sex				
Male	92%	81%	74%	81%
Female	8%	19%	26%	20%
Age (y)				
10–19	25%	10%	19%	13%
20–29	8%	29%	29%	28%
30–39	17%	17%	16%	17%
40–49	8%	13%	26%	15%
50–69	33%	24%	10%	21%
≥70	8%	7%	0%	5%
Mean age (SD)	43 (22)	39 (18)	34 (14)	38 (18)
Racial/ethnic group				
White, non-Hispanic	8%	23%	30%	23%
White, Hispanic	17%	17%	27%	19%
African American	50%	42%	23%	39%
Other, combination	25%	18%	20%	19%
Education				
Less than high school	30%	30%	32%	31%
Less than bachelor’s degree	50%	51%	52%	51%
Bachelor’s degree or more	20%	18%	16%	18%
Etiology of spinal cord injury				
Traffic	33%	38%	32%	37%
Interpersonal violence	8%	19%	26%	20%
Sports	8%	8%	13%	9%
Fall	25%	27%	23%	26%
Other	25%	8%	6%	9%
Neurologic impairment				
Paraplegia incomplete	9%	11%	7%	10%
Paraplegia complete	27%	26%	26%	26%
Tetraplegia incomplete	55%	47%	36%	45%
Tetraplegia complete	9%	16%	32%	19%
Rehabilitation LOS (d)				
<20	17%	14%	6%	13%
20–39	25%	29%	3%	24%
40–59	17%	14%	13%	14%
60–79	17%	13%	16%	14%
80–99	25%	13%	26%	17%
≥100	0%	16%	35%	19%
Mean LOS (SD)	52 (31)	68 (76)	93 (52)	72 (70)

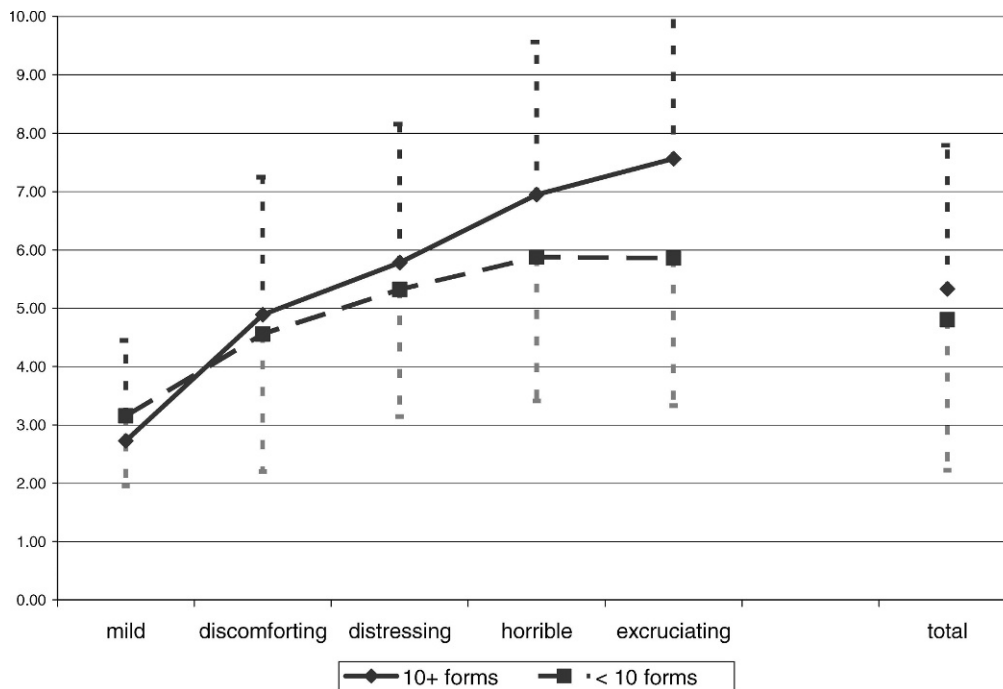
LOS, length of stay.

<sup>a</sup> The actual number of cases for which means and percentages are calculated is smaller for every variable because of missing information.

weight; the fact that some contributed 2 NRS ratings in the VRS “mild” category, and others 10 or more is not taken into account. Columns 9 through 11 provide parallel information when each individual’s mean and SD are weighted by the number of paired ratings on which

their mean and SD was actually based. The ratios (column 11) are slightly different from those in column 8, but the result is the same: a reduction in SD by about 25%, with the largest reduction occurring in the 2 extreme categories, “mild” and “excruciating.”





**Figure 1.** Mean and SD on NRS by VRS and group (all pain components, all occasions).

Information at the level of the individual is provided in Table 2 under the headings from “mild” through “excruciating.” Included in each of the VRS adjectives columns is the mean NRS rating of those of the 36 patients who made at least 5 NRS ratings corresponding to the VRS category in question. This number was taken as an arbitrary minimum, under the assumption that 5 is large enough to counteract any upward or downward effect of 1 erroneous or otherwise atypical NRS rating on an individual’s tendency to associate a single NRS rating (or a narrow range of NRS ratings) with a particular VRS category. Reading down columns, it is easy to see that the VRS adjectives are paired with mean NRS ratings that vary considerably between patients, which reflects in another way the findings presented in Table 3. However, what is surprising are the sizes of the discrepancies between patients. Bolding identifies the lowest and highest means in each column. These extremes are also shown at the bottom of Table 2, along with the difference between the 2 extremes. For example, “mild” corresponds with mean values between 1.4 and 3.3, a difference of 1.9 NRS points, and “discomforting” with values between 2.4 and 7.0, a discrepancy of 4.6 NRS points. The smaller difference values for the 2 extreme adjectives (“mild” and “excruciating”) presumably result from the limits imposed by the extreme values of the NRS, 0 and 10. However, the number of patients selecting “excruciating” at least 5 times (3) is too small to be certain of that finding. In general, compared with the 36 patients with 10 or more rating pairs or compared with the means for the entire sample of 164 patients, individuals may have aberrant mean NRS values for any VRS adjective they select frequently.

In addition, for an individual, the difference in mean NRS values between successive VRS adjectives may be unreasonably small. For instance, patient 95 has a mean NRS value of 9.3 for 8 ratings corresponding to “discomforting” and 1 of 10.0 for 7 ratings corresponding to “horrible.” He used the intermediate adjective “distressing” only once (with a corresponding NRS rating of 4, which is much lower than the mean rating for either “discomforting” or “horrible.”) Four patients in Table 2 (patients 19, 39, 85, and 105) deserve special attention, because their mean NRS ratings seem disordered vis-à-vis their VRS selections. Detailed information on these 4 is provided in Table 4. Patient 19 used “mild” 10 times, pairing it with NRS ratings between 2 and 5, most often selecting a severity rating of 3. He used “discomforting” 14 times, also matching it with NRS ratings between 2 and 5, with a preference for the lower end of the NRS scale. Patient 105 paired “discomforting” with values between 5 and 8, with a preference for the high end of this range; “distressing” he coupled with NRS values between 4 and 8, with a preference for the low end. Compared with these 2 patients, patient 39 seems to have randomly paired both “discomforting” and “distressing” with about any NRS value between 1 and 8. Patient 85 associated “horrible” both with a NRS rating of 6 and one of 10, and “excruciating” selections typically appeared next to NRS ratings of 7 or 8.

## DISCUSSION

Poor correspondence between VRS scores and VAS or NRS ratings traditionally has been explained by the fact that the meaning of the VRS adjectives may differ from

**Table 2.** Patients With 10 or More Completed Forms: Rho Between VRS and NRS and Mean NRS for Each VRS Category<sup>a</sup>

Patient No.	Rho	N	VRS Adjective				
			Mild	Discomforting	Distressing	Horrible	Excruciating
1	0.73	14		3.4			
4	-0.55	10		<b>7.0</b>			
7	0.58	38		3.7	7.0		
11	0.20	23		5.8			
13	0.76	12				8.2	
14	0.68	15				7.1	
16	0.03	14		5.2			
19	-0.01	27	3.2	3.1			
39	-0.36	15		4.8	2.8		
42	0.04	11			4.7		
52	0.33	23	<b>1.4</b>	<b>2.4</b>		<b>5.4</b>	
56	0.25	14					7.2
58	0.33	11	2.2				
65	0.29	12	<b>3.3</b>				
67	-0.03	15		6.0			
69	0.54	11		2.5			
70	0.14	10				6.8	
73	0.18	38		6.3	<b>7.2</b>		
78	0.41	11				7.2	
81	0.46	11		6.0			
85	0.10	17				8.1	7.9
86	-0.03	25		3.7			
91	0.36	10				7.2	
95	0.27	16		9.3		<b>10.0</b>	
99 <sup>b</sup>	0.60	11					
100	-0.20	11		3.6			
105	-0.51	13		6.9	5.5		
106 <sup>b</sup>	0.49	10					
107	0.61	17		5.4			
108	0.00	11		5.0			
115	0.17	11		4.8			
140	0.13	27		5.5			
144	0.20	14		4.0			
146	0.29	26		4.0			
153	-0.37	16		4.8			
158	0.40	16		6.3			7.8
Lowest	-0.55		1.4	2.4	2.8	5.4	7.2
Highest	0.76		3.3	7.0	7.2	10.0	7.9
Difference	1.31		1.9	4.6	4.4	4.6	0.7
Mean for 36 patients	—		2.7	4.9	5.8	7.0	7.6
Mean for 128 patients	—		3.1	4.6	5.3	5.9	5.9
Mean for 164 patients	—		2.9	4.7	5.6	6.5	6.8

<sup>a</sup> This applies if 5 or more ratings corresponding to the adjective in question were made.

<sup>b</sup> Patients completed 10 or more forms but did not complete 5 or more NRS ratings for any individual VRS adjective.

one person to the next (14,24). The data in Table 2 indeed suggest that in this study there is not much agreement between individuals as to what NRS value (or range of NRS values) corresponds to each of the VRS

adjectives available for selection. Although consistency within individuals can reduce the spread of NRS values corresponding to a VRS category by about 25% compared with the cross-individuals situation addressed

**Table 3.** Mean and SD of NRS Within Categories of VRS for Patients With 10 or More Completed Forms<sup>a</sup>

VRS Adjective	Average of the Means and SDs of Individual Patients										
	Mean (1)	SD (2)	NC (3)	Unweighted					Weighted		
				Mean (4)	Mean (5)	NP (6)	SD (7)	Ratio of SDs (8)	Mean (9)	SD (10)	Ratio of SDs (11)
1 = Mild	2.73	1.20	51	2.86	13	0.70	9	0.58	2.73	0.75	0.63
2 = Discomforting	4.89	2.36	316	5.07	34	1.77	32	0.75	4.89	1.85	0.79
3 = Distressing	5.78	2.18	96	5.59	28	1.70	23	0.78	5.78	1.68	0.77
4 = Horrible	6.95	2.47	75	6.50	22	2.09	15	0.85	6.95	1.84	0.75
5 = Excruciating	7.56	2.53	48	7.32	15	1.68	11	0.66	7.56	1.57	0.62
Total across VRS levels	5.33	2.58	586	5.42	36	1.94	36	0.75	5.33	1.97	0.76

<sup>a</sup> NC, number of NRS completions; NP, number of patients.

in most prior research (Table 3), there still is a worrisome discrepancy between successive pairings completed by the same individuals, as well as (sometimes) apparent reversals of the VRS adjectives, resulting in negative correlation values.

Differences between individuals are not surprising, if one believes with Ohnhaus and Adler that “the intervals between the word categories do not represent identical steps in pain intensity. A continuous sensation is artificially transferred into a digital system (24).” How-

**Table 4.** Ratings of Four Patients With Disordered VRS Adjectives

Patient No.	NRS	VRS Adjective					Total
		Mild	Discomforting	Distressing	Horrible	Excruciating	
19	2	1	5				6
	3	6	4	2			12
	4	3	3	1			7
	5		2				2
	Total	10	14	3			27
39	1		2	1			3
	2			2			2
	3		1	1			2
	4		1				1
	5		2				2
	6		1	1			2
	7		1				1
	8		2				2
Total		10	5			15	
85	5		1				1
	6				2		2
	7		1		1	3	5
	8				1	3	4
	9						
	10		1		3	1	5
Total		3		7	7	17	
105	4			1			1
	5		1	3			4
	6		1	1			2
	7		3				3
	8		2	1			3
	Total		7	6			13

**Table 5.** Mean NRS Ratings for VRS Adjectives (by Subgroup and Total) and Differences Between Category Means

Group	None	Δ	Mild	Δ	Discomforting	Δ	Distressing	Δ	Horrible	Δ	Excruciating
Patients with <10 forms											
Mean	–		3.16		4.56		5.32		5.88		5.86
Difference		–		1.4		0.76		0.56		–0.02	
N	0		45		303		81		57		42
Patients with ≥10 forms											
Mean	–		2.73		4.89		5.78		6.95		7.56
Difference		–		2.16		0.89		1.17		0.61	
N	0		51		316		96		75		48
All patients											
Mean	–		2.93		4.73		5.57		6.48		6.77
Difference		–		1.8		0.84		0.91		0.29	
N	0		96		619		177		132		90

ever, Melzack included the VRS terms used here in the MPQ *exactly because* they were equidistant on a 1-to-5 scale, according to his samples (27). Applications of the Pearson correlation coefficient to quantify the relationship between VRS and NRS (or VAS) and other uses of parametric statistical tests assume that the VRS offers an interval or ratio scale when successive cardinal numbers are assigned to adjacent adjectives. The data in Figure 1 and Table 5 strongly suggest that this is not the case. For all patients combined, the difference between the mean NRS scores for “mild” and “discomforting” is more than twice as large as the difference between the latter and “distressing” or than the “distressing” and “horrible” difference. Two phenomena are noteworthy: (a) for some reason, the spread of values for the “10 or more forms” group is larger than the one for the patients who completed a smaller number of forms; (b) major areas of the 0 to 10 NRS range do not seem to be covered by an appropriate VRS adjective.

The 4 patients shown in Table 4 seem to have been fairly consistent in ordering the VRS adjectives in ways the McGill developers may never have considered. Even if one believes that the meaning of the various adjectives in a VRS are fairly well established through consensual usage by persons in a language community, it is unlikely that transfer into a “digital system” proceeds according to identical psychological processes. The affective, evaluative, and other aspects of the pain experience may impinge on the sensory dimension and affect the selection of adjectives on a VRS (16) and likely the selection of numbers on a NRS. “Simple Pain Rating Scales Hide Complex Idiosyncratic Meanings” was the title of the paper by de C. Williams and colleagues, who investigated how and why people use pain self-report terms and scales (28). Differences in familiarity with numbers and in understanding of the 0-to-10 scale may cause their own distortions. A number of studies have shown that patients in general, and especially those of an

older age, tend to prefer the VRS over the VAS and the NRS and make fewer errors using the former (13,17). Some research has claimed that individuals with more education make fewer errors in completing pain severity measures and produce smaller correlations between VRS and NRS/VAS (19,29–31), but that is contradicted by other investigations (30,32,33), which may have been underpowered.

Presumably, all these factors played a role in previous studies that investigated the association between the VRS and the NRS/VAS. The correlations found in the present study are among the lowest reported. That may be due to the special circumstances of the patients involved (only months if not weeks from the sudden onset of a paralyzing injury with life-shattering implications that led to new nociceptive and neuropathic pains) and to the specific VRS selected. Much prior research in this area has used a VRS with fewer and simpler adjectives (eg, no pain-mild-moderate-severe). Upon reflection, the adjectives that are part of the McGill VRS strike one as pompous terms whose relative severity, outside of the fixed order of the ordinal scale, is not clear. Is a distressing pain expected to be more severe than a discomforting one? Is a horrible pain the same in severity as an excruciating one? Or worse? It may be that the linear relationship between VAS and VRS reported by Averbuch and Katzper, for instance, was found because they used much simpler VRS terms. Their adjectives were numbered 0 = none, 2 = mild, 3 = moderate, 4 = severe, and 5 = extreme, and the regression line between levels 2 and 4 was indeed straight; the number of patients selecting “none” or “extreme” was too small to determine whether linearity extended to the 2 scale extremes (20).

If patients have a fairly clear image in mind of what the rank-order of the McGill VRS adjectives is and of the amount of pain each one represents, there are possible explanations why in the present study the within-patient



correlations were much lower than in the earlier research that used a similar design. In the Breivik et al study (21), all ratings of pain were made within hours after oral surgery; in the Linton and Götestam investigation (18), pain ratings were completed daily over 1 or 2 weeks. In contrast, patients in the study reported here made ratings at approximately 1-week intervals during the inpatient phase, and in some instances (pain disappeared temporarily; patient or research assistant was not available for some reason), the interval was more than a week. Post discharge, intervals were about 3 months. In addition, patients were asked to make separate ratings of the multiple pains they experienced. It is unlikely that they completed the VRS and NRS in terms of “right this second,” in spite of a reference in the NRS question (but not the VRS one!) to the pain’s severity “at this moment.” Thus, if multiple pains were not present continuously, the patients could have referred to the severity of their distinct pains “this morning” or “earlier during this interview” and separated out the impact of the 2 to 4 or 5 pains some said they had. But even in that situation, one may wonder how feasible it was for them to keep apart the severity of the separate pains, especially on a VRS that, given the nature of the adjectives used, has a strong admixture from the affective pain dimension.

The weak correlation between NRS and VRS reported here may be blamed on the nature of the VRS used. As mentioned, the nature of the adjectives suggests an ordinal scale of the evaluative rather than the intensity or severity aspect of pain. In fact, in Figure 1 of his paper, Melzack arranges the adjectives under the column heading “evaluative” but specifies that the distances between them reflect an intensity scale (27). On the MPQ itself, the VRS is presented under the heading “How strong is your pain?” and the instructions refer explicitly to intensity (27). Among pain researchers, the MPQ VRS is known as PPI, which stands for Present Pain Intensity. Thus, a direct or indirect link of the MPQ VRS to severity is assumed by the MPQ originator and others. Even if one assumes that the MPQ VRS reflects pain evaluation rather than sensation, the findings reported here as to the weak and sometimes inconsistent link between VRS and NRS are problematic. Although a linear relationship between the 2 would not be assumed at all, one would presume that there is always a positive correlation between NRS and VRS and that the NRS scores corresponding to a particular VRS adjective would be fairly tightly grouped. Assumptions about mood states and other distractors might be adduced to explain weak correlations, but negative correlations (eg,  $-0.55$  for patient 4 in Table 2) are harder to explain.

Lund et al have suggested that the meanings of VRS adjectives differ by pain type; they found that optimal VAS cut-points corresponding to the successive VRS adjectives were not the same for people with different pain etiologies (14). These authors used a fairly small

number of patients, and to date no cross-validation of their results has been published. The patients in the current study had a variety of pain types related to their SCI (34), and some had pain resulting from associated injuries that they incurred at the time of their spinal injury. If a clear assignment of pains to pain problem etiologic categories is feasible, it may be possible to further investigate this suggestion, although the number of VRS and NRS completions per pain type within patients may be too small to do this analysis. It seems unlikely that the “pain type effect” is strong enough to completely counteract the “individual use of adjectives effect” shown here.

Thus far, the focus in this discussion of the weak and variable correlations between the VRS and NRS in this investigation has been on the presumed weaknesses of the VRS that was used. However, it is possible that problems with the NRS contributed to the lack of correspondence that was noted. The NRS is assumed to measure pain severity on a ratio scale, but that depends on 2 assumptions: (a) that there is such a thing as the “worst possible pain you can imagine” and (b) (when data for various patients are combined) that this anchoring level of extreme pain is the same for all patients. These may be erroneous assumptions. It is difficult to remember pain in terms of its severity level and perhaps impossible to imagine the severity level of a pain one has never had. Presumably, patients in setting the high-end boundary when completing a NRS (or VAS) (the pain sensation corresponding to 10 or 100, the “worst pain imaginable”) refer to or compare with pains they have actually experienced in their lives. The worst possible pain may be much more severe for a woman who has gone through a difficult childbirth than for someone who has never experienced anything worse than a bad headache. For those of our patients who before their SCI never had experienced much more than a toothache or a broken leg, the “worst possible pain” reference point may have been their post-SCI pain, and that pain may have been a moving target. Post-SCI pains and especially neuropathic pains may develop weeks, months, and even years after SCI and can increase in severity over time (2,35). Thus, the *same* momentary pain level experienced may have been given a lower NRS rating from one week to the next, not just because a chronic pain became familiar, but because a flare-up of the same or another pain moved higher, to the scale endpoint corresponding to “worst possible pain.”

Another assumption underlying use of a NRS as a ratio scale is that patients use the points intermediate between 0 and 10 in a fashion that translates the internal sensation linearly into a score. There is research to suggest that this may not be the case (36).

The correlation between VRS responses and NRS ratings may have been especially poor in this sample, because the patients’ recent injuries distressed them. Traumatic brain injuries (diagnosed or undiagnosed),

medications, confusion resulting from the hustle and bustle of a busy rehabilitation program, and new bodily sensations (including pain) all may have contributed to random error in rating pain severity. Quite likely, in a sample of outpatients with pains that have been fairly stable, a higher correlation would be found. However, the possible inconsistent and unreliable use of pain severity measures is the situation clinicians and clinical researchers of pain phenomena in acute rehabilitation must manage. They cannot assume that the pain severity scores their patients provide are highly reliable, and they may need to routinely use 2 or more measures to confirm that their pain treatment measures are effective.

## CONCLUSIONS

Pain is a common human experience that is difficult to communicate to others, whether words or numbers or linear analogs are used. Because pain is so common among individuals with SCI, it deserves ongoing research by SCI and pain specialists. The 3 major pain severity assessment instruments in use, VRS, NRS, and VAS, all have been declared to be valid ways of operationalizing the pain sensory experience, mostly based on comparisons with one another. All have been recommended for use in SCI, with some reservations relevant to the VAS because individuals with SCI may lack the hand function needed to draw their mark on the analog line (9). However, strong correlations are not equivalent to substitutability, and additional research on how scores on the 3 instruments relate to one another may be helpful in better understanding the shortcomings of each and their limitations in use as an outcome measure in clinical and research efforts to reduce pain. Continued examination of how people use or fail to “correctly” use the instruments may help in modifying them such that they produce data that contain less bias, or at least less variability. SCI researchers and clinicians should be aware of the limitations of the various instruments used for operationalizing pain severity and make decisions accordingly.

## ACKNOWLEDGMENT

Thanks to Ava Dorfman, Ayana Jones, Salma Akter, Veronica Toribio, and Vishali Saldi for data collection.

## REFERENCES

- Nepomuceno C, Fine PR, Richards JS, et al. Pain in patients with spinal cord injury. *Arch Phys Med Rehabil.* 1979; 60(12):605–609.
- Stormer S, Gerner HJ, Gruninger W, et al. Chronic pain/dysaesthesiae in spinal cord injury patients: results of a multicentre study. *Spinal Cord.* 1997;35(7):446–455.
- Siddall PJ, Taylor DA, McClelland JM, Rutkowski SB, Cousins MJ. Pain report and the relationship of pain to physical factors in the first 6 months following spinal cord injury. *Pain.* 1999;81(1–2):187–197.
- Mariano AJ. Chronic pain and spinal cord injury. *Clin J Pain.* 1992;8(2):87–92.
- Nichols PJ, Norman PA, Ennis JR. Wheelchair user's shoulder? Shoulder pain in patients with spinal cord lesions. *Scand J Rehabil Med.* 1979;11(1):29–32.
- Sie IH, Waters RL, Adkins RH, Gellman H. Upper extremity pain in the post rehabilitation spinal cord injured patient. *Arch Phys Med Rehabil.* 1992;73(1):44–48.
- Waters RL, Sie IH. Upper extremity changes with SCI contrasted to common aging in the musculoskeletal system. *Top Spinal Cord Inj Rehabil.* 2001;6(3):61–68.
- Dijkers MP, Bryce TN, Zanca J. Prevalence of chronic pain after traumatic spinal cord injury: a systematic review. *J Rehabil Res Dev.* 2009;46(1):13–30.
- Bryce TN, Budh CN, Cardenas DD, et al. Pain after spinal cord injury: an evidence-based review for clinical practice and research. Report of the National Institute on Disability and Rehabilitation Research Spinal Cord Injury Measures Meeting. *J Spinal Cord Med.* 2007;30(5):421–440.
- Sawatzky B, Bishop CM, Miller WC, SCIRE Research Team. Classification and measurement of pain in the spinal cord-injured population. *Spinal Cord.* 2008;46(1):2–10.
- Widerstrom-Noga E, Biering-Sorensen F, Bryce T, et al. The international spinal cord injury pain basic data set. *Spinal Cord.* 2008;46(12):818–823.
- Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain.* 1986; 27(1):117–126.
- Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J Clin Nurs.* 2005;14(7):798–804.
- Lund I, Lundeberg T, Sandberg L, Budh CN, Kowalski J, Svensson E. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross-sectional description of pain etiology groups. *BMC Med Res Methodol.* 2005;4(5):31.
- Woodforde JM, Merskey H. Some relationships between subjective measures of pain. *J Psychosom Res.* 1972;16(3): 173–178.
- Reading AE. A comparison of pain rating scales. *J Psychosom Res.* 1980;24(3–4):119–124.
- Jones KR, Vojir CP, Hutt E, Fink R. Determining mild, moderate, and severe pain equivalency across pain-intensity tools in nursing home residents. *J Rehabil Res Dev.* 2007;44(2):305–314.
- Linton SJ, Gotestam KG. A clinical comparison of two pain scales: correlation, remembering chronic pain, and a measure of compliance. *Pain.* 1983;17(1):57–65.
- Clark P, Lavielle P, Martinez H. Learning from pain scales: patient perspective. *J Rheumatol.* 2003;30(7):1584–1588.
- Averbuch M, Katzper M. Assessment of visual analog versus categorical scale for measurement of osteoarthritis pain. *J Clin Pharmacol.* 2004;44(4):368–372.
- Breivik EK, Bjornsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. *Clin J Pain.* 2000;16(1):22–28.
- Berntson L, Svensson E. Pain assessment in children with juvenile chronic arthritis: a matter of scaling and rater. *Acta Paediatr.* 2001;90(10):1131–1136.
- Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthritis Cartilage.* 2003;11(7):499–507.
- Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal

- rating scale and the visual analogue scale. *Pain*. 1975;1(4): 379–384.
25. Melzack R. The short-form McGill pain questionnaire. *Pain*. 1987;30(2):191–197.
26. Turner JA, Cardenas DD. Chronic pain problems in individuals with spinal cord injuries. *Semin Clin Neuropsychiatry*. 1999;4(3):186–194.
27. Melzack R. The McGill pain questionnaire: major properties and scoring methods. *Pain*. 1975;1(3):277–299.
28. de C Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. *Pain*. 2000;85(3):457–463.
29. Jelsma JM, Machiri G, Madzivire DM. The use of pain measurement scales in the Zimbabwean context. *Cent Afr J Med*. 1997;43(9):256–259.
30. Li L, Liu X, Herr K. Postoperative pain intensity assessment: a comparison of four scales in Chinese adults. *Pain Med*. 2007;8(3):223–234.
31. Fadaizadeh L, Emami H, Samii K. Comparison of visual analogue scale and faces rating scale in measuring acute postoperative pain. *Arch Iran Med*. 2009;12(1): 73–75.
32. Herr K, Spratt KF, Garand L, Li L. Evaluation of the lowa pain thermometer and other selected pain intensity scales in younger and older adult cohorts using controlled clinical pain: a preliminary study. *Pain Med*. 2007;8(7): 585–600.
33. Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. *Clin J Pain*. 2004;20(4): 207–219.
34. Bryce TM, Dijkers MP, Ragnarsson KT, Stein A, Chen B. Reliability of the Bryce/Ragnarsson SCI pain taxonomy. *J Spinal Cord Med*. 2006;29:118–132.
35. Siddall PJ, McClelland JM, Rutkowski SB, Cousins MJ. A longitudinal study of the prevalence and characteristics of pain in the first 5 years following spinal cord injury. *Pain*. 2003;103(3):249–257.
36. Pesudovs K, Noble BA. Improving subjective scaling of pain using Rasch analysis. *J Pain*. 2005;6(9):630–636.