

## Development and validation of the Pain Outcomes Questionnaire-VA

Michael E. Clark, PhD; Ronald J. Girona, PhD; Robert W. Young, PhD  
*James A. Haley Veterans Hospital and University of South Florida, Tampa, FL*

**Abstract**—The development of effective pain treatment strategies requires the availability of precise and practical measures of treatment outcomes, the importance of which has been noted in the Veterans Health Administration's (VHA's) National Pain Initiative. This paper presents the results of a 5-year collaborative effort to develop and validate a comprehensive and efficient self-report measure of pain treatment outcomes. Two samples of veterans (957 total subjects) undergoing inpatient

or outpatient pain treatment at six VHA facilities completed Pain Outcomes Questionnaire-VA (POQ-VA) items and several additional measures. We used a comprehensive, multistage analytic procedure to evaluate the psychometric properties of the instrument. Results provided strong support for the reliability, validity, and clinical use of the POQ-VA when used to evaluate the effectiveness of treatment for veterans experiencing chronic noncancer pain.

**Key words:** chronic pain, outcomes assessment, pain assessment, psychological test, rehabilitation outcome.

**Abbreviations:** AAPM = American Academy of Pain Management, ADL = activities of daily living, ANOVA = analyses of variance, ANX = anxiety, APSPOQ = American Pain Society's Patient Outcomes Questionnaire, BPI = Brief Pain Inventory, CFA = confirmatory factor analysis, CFI = Comparative Fit Index, CI = confidence interval, DEP = depression, FIM = Functional Independence Measure, HEA = health concerns, IRB = institutional review board, M = mean, MANOVA = multivariate analysis of variance, MLe = Maximum Likelihood Estimation, MMPI-2 = Minnesota Multiphasic Personality Inventory-2, MPI = Multidimensional Pain Inventory, NA = negative affect, NFI = Normed Fit Index, NPDB = National Pain Data Bank, NRS = numerical rating scale, PCE = Physical Capacities Evaluation, POQ-VA = Pain Outcomes Questionnaire-VA, PQFF = Pain Questionnaire Feedback Form, PTS = Pain Treatment Satisfaction, PVAS = Pain Visual Analog Scale, RCI = Reliable Change Index, RMSEA = root-mean-square error of approximation, SD = standard deviation, SEM = standard error of measurement, SIP = Sickness Impact Profile,

SPQ = Sleep Problems Questionnaire, TSK = Tampa Scale of Kinesophobia, VA = Department of Veterans Affairs, VHA = Veterans Health Administration, WHYMPI = West Haven-Yale Multidimensional Pain Inventory.

**This material was based on work supported by the Department of Veterans Affairs (VA) Rehabilitation Research and Development grant O2069R, VA Rehabilitation Research and Development Career Development Award/Eastern Paralyzed Veterans Association 2002 Scholar Award B2744V, VA Rehabilitation Research and Development Research Enhancement Award Program (REAP) grant E2964F, and American Academy of Pain Management grant.**

Address all correspondence and requests for reprints to Michael E. Clark, PhD; Clinical Director Chronic Pain Rehabilitation Program (2CWR), James A. Haley Veterans Affairs Hospital, 13000 B. B. Downs Blvd, Tampa, FL 33612-4798; 813-972-2000, ext. 7484; Michael.Clark2@med.va.gov.

## INTRODUCTION

The development of effective pain treatment strategies requires the availability of precise and practical measures of treatment outcomes [1]. Within the Veterans Health Administration (VHA), the importance of measuring pain treatment effectiveness has been noted in the VHA's National Pain Initiative [2]. Pain treatment specialists have long recognized the need for a uniform method of measuring pain treatment outcomes [3–7]. The development and use of such a system would allow improved evaluation of short- and long-term pain treatment effectiveness, better estimates of the cost effectiveness of interventions, direct comparisons of the outcomes of treatment programs or methods, and improved consumer satisfaction monitoring. Additionally, uniform, multifacility measurement systems promote improved program quality and meet the highest interdisciplinary rehabilitation measurement standards as promulgated by the American Congress of Rehabilitation [4].

Current standards for pain treatment are based on the biopsychosocial conceptualization of chronic pain as a complex, multidimensional phenomenon, with diverse etiologic and sustaining factors [8]. Accordingly, recommendations for comprehensive treatment target multiple domains of patient functioning, including the physical, perceptual, behavioral, and psychosocial status of the individual. Reflecting this multidimensional approach to conceptualization and treatment, current guidelines for pain outcomes assessments mandate the measurement of treatment-related change within each major domain of an individual's chronic pain experience [9]. Pain outcomes measures that provide separate domain or scale scores are preferred over instruments that yield only single summary scores because individuals with chronic pain present with different patterns of dysfunction across domains. The use of a single summary score, as exemplified by the Oswestry Low Back Pain Disability Questionnaire and the Roland-Morris Activity Scale [10,11], may obscure treatment-induced changes in specific outcomes domains.

Presently, no multidomain pain treatment outcomes instruments with separate domain or scale scores meet the criteria just mentioned. Perhaps as a consequence of this lack of pain-specific measures, many Department of Veterans Affairs (VA) and private sector rehabilitation services use general measures of health status, such as the Functional Independence Measure (FIM) or the SF-36 Health Survey [12,13], as indexes of treatment outcomes.

Unfortunately, general health status measures may not adequately assess the multidimensional symptomatology of the pain experience and, as a result, may be incapable of capturing treatment-related change in the major domains of pain-related disability. With respect to the FIM, this deficiency is most apparent in the high baseline functional status scores of pain patients, which fail to discriminate among pain conditions and allow little room for improvement. Similarly, although the SF-36 is widely used and well validated for general health status outcomes [13], the data supporting its stand-alone use as a pain outcomes measure are mixed and suggest that it may be most useful as an adjunct to a "disease-specific" instrument [14].

Without validated instruments designed specifically to assess pain treatment outcomes, pain practitioners instead often use tests originally developed as clinical assessment tools. Two of the most popular multidomain instruments that have been used to assess outcomes in this manner are the West Haven-Yale Multidimensional Pain Inventory (WHYMPI) and the Brief Pain Inventory (BPI) [15,16]. The WHYMPI is a well validated and reliable patient assessment instrument composed of 52 items that span multiple outcomes domains. Nevertheless, while some evidence supports the WHYMPI's sensitivity to treatment-related change [17,18], the bulk of published studies focus on its use as a clinical pain assessment tool. Additionally, although most of the WHYMPI psychometric research has used the original 52-item scale, a subsequent 61-item release, now commonly referred to as the Multidimensional Pain Inventory (MPI), is the version used most often in clinical settings. Furthermore, the WHYMPI does not include measures of some key outcomes dimensions (e.g., medical use and patient satisfaction). Instead, it incorporates items that may not be relevant to direct pain outcomes assessment and contains numerous items that do not include an implicit or explicit time reference, which may complicate efforts to evaluate the effects of discrete treatment episodes. Concerns also exist with respect to some aspects of the methodology employed in the development of the WHYMPI, including an inadequate (120 subjects) sample size for the analyses conducted [19], lack of a single factor analysis of the entire item pool [19], and the apparent use of the same sample for both exploratory and confirmatory factor analyses. Results of subsequent studies of the psychometric characteristics of the WHYMPI raise concerns regarding the stability of the original WHYMPI factor structure

and have identified substantial redundancy between some scales [20,21]. A recent examination of the 61-item MPI found similar weaknesses as well as a factor structure that accounted for less than the 60 percent criterion suggested by Hair, Anderson, and Tatham [19,22].

The BPI is a 32-item inventory designed for use with cancer pain patients to assess pain history, pain intensity, past response to treatment, and pain interference. Primarily a treatment planning tool, the BPI has been translated into many languages and validated for use with cancer and chronic-disease pain patients. However, because the primary purpose of the BPI is to assist providers in planning an effective intervention, many items are not appropriate for outcomes assessments and the scope of those that may be suitable is somewhat limited. For example, the BPI Reactive subscale is a global index of pain interference consisting of seven items, each of which assesses a different domain of functioning. While this may be useful as a gross index of outcomes, it potentially obscures important differences in treatment response among the key domains and prevents empirically based refinement of the intervention approach. Finally, the BPI does not assess pain-related medical use or patient treatment satisfaction.

Another outcomes measurement method used by some practitioners involves assembling a battery of measures by selecting several unidimensional, domain-specific instruments and administering them as an outcomes "package" [1]. Unfortunately, even when practitioners select only reliable and validated instruments for inclusion in the battery, the resultant lack of uniformity in measures precludes comparisons across treatment sites or the development of outcomes benchmarks. Additionally, many of these instruments may be fairly lengthy and, when combined into a battery, may require substantial patient and staff burden.

An alternative approach is to design an instrument specifically to assess pain treatment effectiveness across all outcomes domains deemed important for individuals with chronic pain. Ideally, such an instrument should include measures of the perceptual (i.e., pain), emotional (e.g., depression and anxiety), and functional (e.g., activity levels and pain interference) dimensions of pain [9,23]; be reliable and valid [4]; include a means of measuring both short- and long-term outcomes [9]; incorporate measures of medical use, employment status, and consumer satisfaction [9]; and be as brief as practical [24].

The American Pain Society's Patient Outcomes Questionnaire (APSQO) represents an example of one

such attempt [18]. This instrument was developed as a quality improvement tool and incorporates pain intensity, pain interference, patient satisfaction, and pain medication items. However, the APSQO was designed to assess acute and cancer pain outcomes rather than chronic pain outcomes. In addition, it does not assess any work-related outcomes or changes in medical use, and it includes only one general question pertaining to pain-related emotional changes. Perhaps most importantly, no reliability or validity data were provided and no subsequent empirical studies of its psychometric characteristics exist.

The National Pain Data Bank (NPDB), developed by the American Academy of Pain Management (AAPM), is a more comprehensive pain outcomes instrument designed to be appropriate for the entire spectrum of pain interventions and service sites. Unlike the APSQO, the NPDB includes three separate forms appropriate for intake, posttreatment or interim, and follow-up administration. The original NPDB combined pain history questions with measures of pain intensity, emotional functioning (depression and anxiety), pain-related impairment, interpersonal "closeness," medication use, employment status, consumer satisfaction, and medical resource use. However, a detailed review of the instrument suggested numerous test construction and item content weaknesses that limited its use [25]. Additionally, our qualitative review of NPDB items revealed other problems related to excessive instrument length and complexity, the use of nonuniform rating scales, retrospective ratings of improvement rather than current reports of function, and poor item wording.

Recognizing both the strengths and the weaknesses of the original NPDB, a cooperative VA-AAPM project was implemented to construct a new outcomes instrument that retained the basic structure of the NPDB. The current study reports the results of this conjoint 5-year effort to develop a brief but psychometrically sound pain outcomes instrument, now called the Pain Outcomes Questionnaire-VA (POQ-VA), that assesses all of the key domains and meets relevant professional and accreditation body standards.\*

\*Copies of the POQ-VA can be obtained by completing a request form at <http://www.vachronicpain.org/Pages/POQReq.htm>.

## METHODS

### Subjects

Two samples of subjects were used in this study. All participants were receiving pain treatment services through one of six VHA pain centers, which varied in comprehensiveness from outpatient single provider to inpatient interdisciplinary approaches. Sample 1 consisted of 248 individuals participating in an inpatient interdisciplinary chronic pain rehabilitation program at a single southeastern VA hospital. Sample 2 consisted of 709 individuals who received either inpatient ( $n = 367$ ; 48.2%) or outpatient ( $n = 342$ ; 51.8%) services at one of the six treatment sites. The vast majority of subjects were military veterans, although several veteran dependents also participated. All study procedures were reviewed and approved by local institutional review boards (IRBs)

and VA research and development (R&D) committees prior to data collection. Written informed consent was obtained from the 609 participants involved in the prospective data collection. Informed consent exemption was requested and granted by the local IRB and R&D committees for the retrospective analysis of data for the remaining 348 subjects.

The demographic characteristics of both samples are presented in **Table 1**. To evaluate any potential variation between inpatients and outpatients in Sample 2, we treated these groups as separate samples in the analyses of demographic characteristics. Chi-square analyses revealed that the samples did not differ in gender,  $\chi^2(2, 956) = 0.60$ , nonsignificant (ns), or racial composition  $\chi^2(6, 955) = 10.08$ , ns. One-way analyses of variance (ANOVAs) revealed a significant difference in age,  $F(2, 953) = 14.70$ ,  $p < 0.001$ , but not pain intensity,  $F(2, 951) = 0.25$ , ns. Bon-

**Table 1.**  
Demographic characteristics.

Variables	Sample 1 (n = 248)	Sample 2 (n = 709)	
	Inpatient	Inpatient (n = 367)	Outpatient (n = 342)
Sex %			
Males	87.9	87.5	85.9
Females	12.1	12.5	14.1
Age (yr)*			
M (SD)	50.03 (11.09)	52.14 (10.74)	55.11 (12.41)
Range	24–81	21–82	24–87
Race %			
Caucasian	70.6	74.3	72.4
African American	17.3	13.7	10.3
Hispanic	8.9	9.0	12.6
Native American	2.4	1.9	2.9
Other	0.8	1.1	1.8
Education (yr)			
M (SD)	—	13.41 (2.56)	13.18 (2.39)
Range	—	6–24	6–24
Pain Intensity			
M (SD)	7.11 (1.60)	7.20 (1.62)	7.13 (1.73)
Range	3–10	0–10	2–10
Pain Duration (yr)			
M (SD)	—	14.18 (12.33)	14.23 (13.70)
Range	—	0.50–58.0	0.08–59.0

\*Sample 2 outpatients differed significantly from Sample 1 and Sample 2 inpatients.

M = mean

SD = standard deviation

ferroni corrected post hoc comparisons revealed that the Sample 2 outpatient group was significantly older than both the Sample 2 inpatient group (mean difference = 2.96,  $p < 0.01$ ) and Sample 1 (mean difference = 5.07,  $p < 0.001$ ), which may indicate that older veterans with pain are more likely to be treated as outpatients than as inpatients. Because education and pain duration data were not available for Sample 1, only Sample 2 treatment groups were compared on these variables. No significant differences were found between the Sample 2 treatment groups in education,  $F(1, 704) = 1.46$ , ns, or pain duration,  $F(1, 603) = 0.00$ , ns.

## Measures

### *POQ-VA Development*

As recommended by contemporary test construction experts [26,27], we used a two-stage, rational-empirical approach in the development of the POQ-VA reflecting an iterative process [26], whereby preliminary item pools are refined by sequential item administration and analysis. In the first stage, we identified six broad outcomes domains (pain intensity, pain-related interference in function, emotional functioning, employment status, medical use, and consumer satisfaction) recommended by pain-related professional organizations (the American Pain Society and the International Association for the Study of Pain) or accreditation bodies (Joint Commission for the Accreditation of Healthcare Organizations and The Rehabilitation Accreditation Commission) for pain treatment outcomes assessment. Next, we reviewed pain history and pain outcomes items from the original version of the NPDB Intake (49 items), Discharge (28 items), and Follow-Up (30 items) Questionnaires for relevance, clarity of meaning, psychometric scaling, and utility with respect to the six general domains. We combined original and altered NPDB items with newly developed items and grouped them into categories according to the six domains. Initial item, factor, and concurrent validity evaluations were conducted to empirically derive preliminary scales [28], and we eliminated items with weak psychometric support or poor utility, refined several pain history and outcomes items, and added items to better assess outcomes domains of interest.

The final 19 POQ-VA primary pain outcome items were selected from a total of 35 potential items in the final item pool based on item analyses and principal component analyses. Only data for the items selected for the

POQ-VA are included in this paper. The primary POQ-VA outcomes items use 11-point (0 to 10) rating scales and are imbedded in each of the three questionnaires making up the POQ-VA. The Intake Questionnaire contains the 19 primary items, 23 personal and pain history items tapping demographics (e.g., age, education), pain experience (e.g., pain duration, pain location), employment status, disability status (e.g., VA service connection, type of claims filed), and opioid use (e.g., current use, associated pain relief), and 2 items assessing pain-related medical use (VA and non-VA healthcare visits during the last 3 months). The Discharge or Interim Questionnaire is intended to be administered when treatment is stabilized or terminated and contains the 19 primary outcomes items, a Pain Treatment Satisfaction (PTS) scale that comprises five, 0- to 10-point ratings of satisfaction with different elements of treatment, and 3 medication-use items. The Follow-Up Questionnaire contains the 19 primary items; 12 history items assessing frequency of pain reinjury, employment status, disability status, and opioid use; 2 medical-use questions; and 2 follow-up treatment satisfaction items (overall treatment satisfaction and recommendation to others).

### *Minnesota Multiphasic Personality Inventory-2*

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is a 567-item self-report measure of psychopathology and personality characteristics with well-documented psychometric properties [29,30]. For the present study, the MMPI-2 Anxiety (ANX), Depression (DEP), and Health Concerns (HEA) Content scales were selected a priori as concurrent measures. The ANX and DEP Content scales assess the cluster of cognitive, emotional, and behavioral symptoms associated with diagnoses of anxiety and depression [31]. The HEA scale measures diverse somatic complaints and preoccupation with bodily functioning [31].

### *Sleep Problems Questionnaire*

The Sleep Problems Questionnaire (SPQ) is a four-item measure of the most common symptoms of poor sleep in both healthy and distressed populations. The scale has adequate internal consistency and validity [32].

### *Tampa Scale of Kinesiophobia*

The Tampa Scale of Kinesiophobia (TSK) is a 17-item instrument that was developed to assess kinesiophobia, or the fear of movement and activity due to concerns about injury or reinjury [33]. Psychometric characteristics

of the TSK suggest that it possesses adequate reliability, has a meaningful factor structure [34], and is a good predictor of a range of pain symptoms and behaviors [35]

#### *Sickness Impact Profile*

The Sickness Impact Profile (SIP) is a widely used 136-item measure of perceived impairment [3,36], with high test-retest reliability and internal consistency [37]. The SIP administration instructions were altered by Turner and Clancy to reflect pain-related impairment rather than general physical impairment [38]. The SIP scales have been found to possess good concurrent validity in chronic pain and cancer pain patients [39,40], and they are sensitive to change resulting from multidisciplinary inpatient treatment for chronic pain [11]. Only the SIP Physical (SIP-P) scale was used in the present study.

#### *Pain Visual Analog Scale*

The Pain Visual Analog Scale (PVAS) is a reliable and well-validated measure of pain intensity in acute, cancer, and chronic pain [41–45]. The PVAS was presented as a 10 cm line anchored with the phrases “no pain” and “worst possible pain.”

#### *Physical Capacities Evaluation*

The Physical Capacities Evaluation (PCE) consists of a series of graded behavioral tasks based on those outlined by Woods [46]. Subjects' performance on each task is rated on a 5-point scale. The PCE measures flexibility, strength, and endurance and yields a total score that reflects an individual's demonstrated physical abilities [47].

#### *Pain Questionnaire Feedback Form*

The Pain Questionnaire Feedback Form (PQFF) is a nine-item scale developed to obtain feedback from the participants on their experience completing the items in the POQ-VA pool. Participants responded to the items using a 4-point scale ranging from “strongly agree” to “strongly disagree.” Five items assessed how accurately the participants felt that the POQ-VA described the impact that pain was having on their lives (e.g., “I was able to accurately describe my pain experience”) and the remaining four items measured how clear and easy they believed the form was to complete (e.g., “The form was easy to use”).

## **Procedures**

Subjects in both samples completed the POQ-VA Intake Questionnaires as a component of a brief psychological assessment during their inpatient admission or initial outpatient appointment. Additional measures administered at intake were the MMPI-2, PCE, PVAS, and SIP for Sample 1; the MMPI-2, SPQ, and TSK for the Sample 2 inpatient group; and the PQFF for a subset of the Sample 2 outpatients (n = 240; 70.2%). The POQ-VA Discharge Questionnaire was administered to inpatients who completed the program (Sample 1, n = 224, 90.3%; Sample 2, n = 282, 76.8%) and outpatients (n = 108, 31.6%) who returned for an additional appointment following implementation of their treatment regimen. A subset of both samples (Sample 1, n = 139, 56.0%; Sample 2, n = 208, 29.3%) completed the POQ-VA Follow-Up Questionnaire 3 months after treatment termination. In addition, a subset of Sample 2 outpatients waiting for treatment appointments (test-retest subset, n = 100, 29.2%) completed the Intake Questionnaire twice as part of a test-retest study designed to evaluate the 1- to 2-week temporal stability of the instrument. The average test-retest interval was 8.38 days (standard deviation [SD] = 1.75, range = 7 to 13).

## **RESULTS**

### **Scale Development**

#### *Principal Components Analysis*

For scale development, Sample 2 was randomly divided into validation and cross-validation subsets that were equivalent to demographic characteristics, treatment site, and mean outcomes item scores. A principal components analysis with Varimax rotation, a method that maximizes the independence of the factors [48,49], was used to explore the underlying component structure of 18 of the 19 primary POQ-VA outcomes items in the validation subset of Sample 2 (n = 353). An a priori decision to exclude the average pain intensity item (Pain Numerical Rating Scale [NRS]) from this analysis was made based on the theoretically and empirically supported rational judgment that it should be treated as a separate scale, regardless of its associations with other items [50]. The analysis yielded a five-component solution that accounted for 72 percent of the total variance. The rotated component matrix, eigenvalue, and percentages of variance explained are presented in **Table 2**. As illustrated in the

**Table 2.**  
Principal components analysis of POQ-VA item pool.

Item	Component				
	1	2	3	4	5
Depression	<b>0.83</b>	0.17	0.06	-0.13	0.13
Anxiety	<b>0.79</b>	0.15	0.07	-0.02	0.05
Tense	<b>0.78</b>	0.16	0.12	-0.08	0.19
Concentrate	<b>0.74</b>	0.11	0.10	-0.18	-0.01
Esteem	<b>0.70</b>	0.16	0.17	-0.15	0.14
Dress	0.18	<b>0.87</b>	0.28	-0.08	0.12
Bathe	0.15	<b>0.86</b>	0.28	-0.11	0.12
Bathroom	0.21	<b>0.85</b>	0.20	-0.05	0.05
Groom	0.20	<b>0.81</b>	0.13	-0.09	-0.01
Stairs	0.11	0.19	<b>0.87</b>	-0.09	0.13
Walk	0.08	0.15	<b>0.85</b>	-0.15	0.00
Carry	0.22	0.22	<b>0.74</b>	-0.12	0.20
Cane	0.10	0.34	<b>0.66</b>	-0.08	0.04
Energy	-0.22	-0.05	-0.01	<b>0.84</b>	0.01
Strength	-0.23	-0.13	-0.13	<b>0.78</b>	-0.14
Active	0.00	-0.08	-0.22	<b>0.67</b>	-0.17
Safe	-0.09	-0.05	-0.08	0.30	<b>-0.80</b>
Reinjure	0.30	0.12	0.18	0.00	<b>0.79</b>
Eigenvalue	3.37	3.25	2.85	2.02	1.48
Variance Explained	18.73%	18.05%	15.85%	11.20%	8.20%

Note: Bolded entries denote primary component loadings.

table, primary item loadings were uniformly high, secondary loadings were relatively low, and the total explained variance was well distributed among the five components.

Because these data confirmed our rational conceptualization of the underlying dimensions of pain-related functioning assessed by the 18 POQ-VA items, the component solution was used to develop five linearly scored scales. In order of variance explained, the components and corresponding scales were labeled “Negative Affect” (NA) (five items), “Activities of Daily Living” (ADL) (four items), “Mobility” (four items), “Vitality” (three items), and “Fear of Activity” (Fear) (two items). Correlations between component and scale scores were uniformly high, NA,  $r = 0.95$ ,  $p < 0.001$ ; ADL,  $r = 0.94$ ,  $p < 0.001$ ; mobility,  $r = 0.93$ ,  $p < 0.001$ ; vitality,  $r = 0.95$ ,  $p < 0.001$ ; and fear,  $r = 0.93$ ,  $p < 0.001$ , indicating that the computation of scale scores by means of linear aggregation of item values accurately captured component score variance.

### Confirmatory Factor Analysis

We replicated the findings from the exploratory factor analyses in the cross-validation subset of Sample 2 ( $n = 356$ ) using confirmatory factor analysis (CFA). The CFA was conducted on the five-factor model with Amos 4.0 using covariance matrices and the maximum likelihood estimation (MLE) method [51]. MLE was chosen because it is scale invariant, works well with small to medium sample sizes, and produces an overall chi-square statistic for evaluating model fit [48,49].

In the model, each latent factor represented a core POQ-VA scale (i.e., ADL, NA, mobility, vitality, and fear) with the corresponding items serving as observed indicators. Model fit was evaluated with four different fit indexes: the overall model chi-square ( $\chi^2$ ), the root-mean-square error of approximation (RMSEA), the Bentler-Bonett normed fit index (NFI), and the comparative fit index (CFI). Although there are no precise criteria for evaluating these fit indexes, rule-of-thumb guidelines have suggested that an excellent model fit is indicated when RMSEA values equal 0.06 or less and the NFI and CFI values equal 0.95 or greater. Adequate model fit is suggested when the RMSEA falls in the range of 0.07 to 0.10 and the NFI and CFI are equal to or greater than 0.90 [52].

The results of the CFA on the five-factor model produced an excellent fit:  $\chi^2(125, n = 356) = 304.22$ ,  $p < 0.001$ ; RMSEA = 0.06 (95% CI = 0.05 to 0.07); NFI = 0.98; and CFI = 0.99. All standardized factor loadings were high, ranging from 0.59 to 0.92. The latent factor intercorrelations were medium to large (mean [M] = 0.45, SD = 0.12, range = 0.30 to 0.63), but not one was so high as to suggest a lack of discriminant validity [53]. These findings suggest that the scale composition represents the underlying factor structure of the POQ-VA extremely well.

### Scale Descriptive Statistics, Reliabilities, and Intercorrelations

We computed descriptive statistics, reliabilities, and intercorrelations for the five-core POQ-VA scale and the Pain NRS scores of Sample 2 participants. The results are presented in **Table 3**. Three indexes of reliability were generated for each scale, and two were computed for the Pain NRS. Data provided by the test-retest subset of Sample 2 were used to compute stability and generalizability coefficients. Stability coefficients consisted of the correlation between subjects' test-retest scores. We calculated

**Table 3.**  
POQ-VA core scale descriptive statistics, reliability coefficients, and intercorrelations.

Scale	M	SD	Stability*	Alpha <sup>†</sup>	Generalizability*	Scale Intercorrelations <sup>‡</sup>					
						1	2	3	4	5	6
ADL	11.30	11.15	0.89	0.90	0.93	—	—	—	—	—	—
Fear	12.21	5.24	0.73	0.59	0.85	0.28	—	—	—	—	—
Mobility	25.87	10.35	0.84	0.78	0.91	0.52	0.31	—	—	—	—
NA	27.22	12.28	0.86	0.84	0.92	0.38	0.37	0.31	—	—	—
Vitality	21.05	5.33	0.67	0.78	0.80	0.26	0.39	0.38	0.37	—	—
Pain	7.17	1.68	0.63	—	0.78	0.29	0.19	0.36	0.25	0.18	—

Note: Stability and generalizability coefficients for ADL, fear, mobility, and vitality have been previously reported [54].

M = mean, SD = standard deviation, n = 709, ADL = activities of daily living, NA = negative affect

Available for the test-retest subset (n = 100) of Sample 2 only.

Computed for the cross-validation subset of Sample 2 only (n = 356).

All correlations significant at  $p < 0.01$ .

generalizability coefficients (indexes of reliability that provide an estimate of the proportion of variance in an observed test score that is attributable to the subject's true score [54]) using an ANOVA-based variance partitioning procedure. Finally, because internal consistency coefficients depend on inter-item correlations, the use of the validation subset of Sample 2 would produce artificially inflated estimates of coefficient alphas ( $\alpha$ ). Therefore, only the cross-validation subset of Sample 2 was used to compute coefficient alphas for the five scales.

With the exception of the Vitality scale (0.67) and the Pain NRS (0.63), the observed stability coefficients generally indicate a high degree of temporal stability. Most likely, the lower coefficient obtained for the Vitality scale was an effect of the temporal parameters of one of the three questions making up the scale, which assess the individual's ". . . sense of strength and endurance TODAY." Similarly, pain ratings are known to fluctuate daily, and therefore, the lower values obtained for the Vitality scale and Pain NRS actually may reflect natural variation rather than measurement error. With respect to internal consistency, the ADL, NA, and Mobility scales demonstrated moderate to strong evidence of item homogeneity and, thus, appear to be measuring relatively unidimensional constructs. The lower coefficients obtained for the remaining scales may be a function of length, as the lowest estimate was obtained for the Fear (0.58) scale, which is composed of only two items. In fact, the Spearman-Brown prophecy formula, which is used to estimate changes in reliability resulting from changes in scale length, demonstrates that simply adding two similar items to the existing scale would produce a coefficient of

0.73 for the hypothetical four-item Fear scale. Generalizability coefficients were relatively high across all scales and well above the 0.75 criterion that has been suggested to indicate excellent scale reliability [55]. Collectively, these data indicate that the POQ-VA scales and Pain NRS are functioning reliably across a range of patient populations.

Scale intercorrelations generally demonstrated moderate associations between scales, with none being unacceptably high or low. These data suggest that the scales measure related but sufficiently distinct aspects of the chronic pain experience. The strengths of the associations between the Pain NRS and POQ-VA scale scores were consistent with the results of previously reported relationships between pain severity and pain interference [16].

### Sex Differences

To examine potential gender differences, we compared the five-core POQ-VA scale and the Pain NRS scores of Sample 2 men and women using a multivariate analysis of variance (MANOVA). The results failed to reveal a significant multivariate effect of sex: Wilk's  $\lambda = 0.97$ ;  $F(12, 1328) = 1.28$ , ns; partial  $\eta^2 = 0.01$ .

### Convergent and Discriminant Validity

POQ-VA scale and Pain NRS correlations with external criteria were examined for inpatients in both samples. In the Sample 1 data set, ADL, Mobility, and Vitality scales and the Pain NRS were correlated with SIP-P, PCE, and PVAS scores. Because of item changes, the NA and Fear scales could not be scored from the Sample 1

data set and, therefore, were not examined. In the Sample 2 data set, the POQ-VA scale, and the Pain NRS scores were correlated with SPQ, TSK, and MMPI-2 Anxiety, Depression, and Health Concerns Content scale scores. Criteria for evidence of convergent validity were set at  $r = 0.30$  to  $0.49$  for moderate support and  $r \geq 0.50$  for strong support consistent with standard recommendations for medium and large effect sizes [46]. The resulting correlation coefficients, which are presented in **Table 4**, reveal that the POQ-VA scales generally demonstrate moderate to strong associations with relevant external criteria and weaker relationships with criteria not expected to measure related constructs. This pattern of associations provides evidence of the convergent and discriminant validity of the POQ-VA scales in inpatient samples.

### Sensitivity to Change

A repeated measures MANOVA was used to evaluate the sensitivity of the POQ-VA scales and the Pain NRS to treatment-related change. Intake, discharge, and follow-up scores of Sample 2 subjects completing all three assessments were compared across the POQ-VA scales, and the Pain NRS revealed a significant multivariate within-subjects effect [Wilk's  $\lambda = 0.50$ ,  $F(12, 169) = 14.38$ ,  $p < 0.001$ , partial  $\eta^2 = 0.51$ ]. Univariate analyses,

which are presented with descriptive statistics in **Table 5**, revealed a significant effect of time on all of the measures. Clearly, the POQ-VA scales and Pain NRS scores reflect changes occurring between assessments at intake, discharge, and follow-up (generally 3 months following discharge).

To ensure that the observed changes actually reflect treatment-related improvement and not a measurement artifact, additional analyses were conducted to compare changes over successive administrations of the POQ-VA to treated and untreated groups. A repeated measures MANOVA with a between subjects factor (treatment group) was used to evaluate differences in scale score changes between Sample 2 subjects who completed both intake and discharge assessments and the Sample 2 test-retest subsample who completed the intake form twice. Because POQ-VA scale and NRS items are identical across the two forms, scores produced by the test-retest subsample during the second administration were treated as Time 2 variables in the current analysis. The analysis revealed a significant multivariate effect of time, Wilk's  $\lambda = 0.83$ ,  $F(6, 458) = 15.77$ ,  $p < 0.001$ , partial  $\eta^2 = 0.17$ , and treatment group, Wilk's  $\lambda = 0.95$ ,  $F(6, 458) = 3.85$ ,  $p = 0.001$ , partial  $\eta^2 = 0.05$ , and a significant treatment group by time interaction was found, Wilk's  $\lambda = 0.89$ ,  $F(6, 458) = 9.44$ ,  $p < 0.001$ , partial  $\eta^2 = 0.11$ . Univariate

**Table 4.**  
POQ-VA core scale validity correlation coefficients.

Scales	Sample 1 (n = 248)			Sample 2 Inpatients (n = 348)					
	SIP-P	PCE	VAS	1	ANX	DEP	HEA	TSK	SPQ
ADL	<b>0.42</b>	<b>-0.31</b>	<b>0.33</b>	0.18	0.14	0.14	0.20	0.22	0.12
Fear	—	—	—	0.19	<b>0.30</b>	0.28	0.26	<b>0.59</b>	0.20
Mobility	<b>0.43</b>	<b>-0.33</b>	<b>0.33</b>	0.16	0.12	0.12	0.18	0.26	0.18
NA	—	—	—	<b>0.38</b>	<b>0.63</b>	<b>0.64</b>	<b>0.51</b>	<b>0.36</b>	<b>0.33</b>
Vitality	0.26	<b>-0.32</b>	0.26	0.25	0.26	0.29	0.22	0.16	<b>0.34</b>
Pain	<b>0.33</b>	-0.22	<b>0.67</b>	0.16	0.10	0.08	0.15	0.15	0.17

Note: Bold-faced correlations meet or exceed the 0.30 criterion for convergent validity.

ADL = activities of daily living

ANX = MMPI-2 Anxiety Content scale

DEP = MMPI-2 Depression Content scale

HEA = MMPI-2 Health Concerns Content scale

NA = negative affect

PCE = Physical Capacities Evaluation

SIP-P = Sickness Impact Profile Physical scale

SPQ = Sleep Problems Questionnaire

TSK = Tampa Scale of Kinesiophobia

VAS = Pain Visual Analog scale

1 = MMPI-2 Clinical scale 1

**Table 5.**  
POQ-VA core scale sensitivity to change.

Scale	Intake		Discharge		Follow-Up		<i>F</i> (1,179)
	M	SD	M	SD	M	SD	
ADL	9.42	10.05	7.46	8.55	8.25	9.40	14.32*
Fear	11.78	4.78	8.50	4.70	9.62	4.60	64.05*
Mobility	25.35	10.24	22.07	10.05	22.87	11.11	55.34*
NA	25.03	11.03	21.18	11.27	23.69	12.34	31.71*
Vitality	20.66	4.70	15.98	5.37	18.52	5.19	140.16*
Pain	6.90	1.59	5.66	1.75	6.14	1.99	120.44*

\* $p < 0.001$   
M = mean  
SD = standard deviation  
*F* = ANOVA *F* statistic  
ADL = activities of daily living  
NA = negative affect

analyses revealed significant interaction effects for the Fear,  $F(1, 463) = 19.86, p < 0.001$ ; Mobility,  $F(1, 463) = 3.92, p < 0.05$ ; NA,  $F(1, 463) = 9.69, p < 0.01$ ; Vitality,  $F(1, 463) = 35.36, p < 0.001$ ; and Pain scales,  $F(1, 463) = 25.70, p < 0.001$ ; but not for the ADL scale,  $F(1, 463) = 0.28, ns$ . Among the treated group, all mean scale score changes from intake to discharge were in the expected direction. Follow-up MANOVAs revealed that significant differences were found across the POQ-VA scales and Pain NRS for the treatment groups at Time 2, Wilk's  $\lambda = 0.90, F(6, 467) = 8.76, ns$ , partial  $\eta^2 = 0.10$ , but not Time 1, Wilk's  $\lambda = 0.98, F(6, 464) = 1.33, ns$ , partial  $\eta^2 = 0.02$ . These data suggest treatment-related effects rather than measurement artifact account for the observed changes across administrations.

### Clinically Significant Change

Standard indexes of clinically significant change, such as the Reliable Change Index (RCI) [56], are based on the assumption that treatment will produce sufficient improvement to allow patients to shift from dysfunctional to normal scale score distributions. However, among chronic populations with relatively intractable disorders, one cannot realistically assume that even the best treatments will produce a full return to normal functioning. In such cases, change indexes such as the RCI may not be the most appropriate measure of clinically significant change [56]. Therefore, an effect size approach was selected for use with the POQ-VA scales and Pain NRS. The scale score differences associated

with small, medium, and large effects [53] were calculated for each scale and the Pain NRS. These values are presented with the standard errors of measurement (SEM) in **Table 6**. Our suggested criterion for evidence of clinically significant change is a medium effect, which is greater than the SEM for each scale and represents a shift of at least one-half an SD from one assessment to the next.

### POQ-VA Readability

We conducted readability analyses to determine the minimum reading-level requirements of the POQ-VA. Flesch-Kincaid scores for the Intake, Discharge or Interim, and Follow-Up Questionnaires were 6.7, 7.6,

**Table 6.**  
POQ-VA core scale clinically significant differences.

Scale	Small (0.20)	Medium (0.50)	Large (0.80)	SEM
ADL	2.08	5.20	8.32	2.95
Fear	1.03	2.57	4.11	2.03
Mobility	2.01	5.02	8.04	3.10
NA	2.33	5.84	9.34	3.47
Vitality	1.05	2.63	4.20	2.38
Pain	0.34	0.85	1.35	0.79

Small (0.20), medium (0.50), and large (0.80) are effect sizes recommended by Cohen [53].  
ADL = activities of daily living  
NA = negative affect  
SEM = standard error of measurement

and 7.9, respectively, suggesting that a 7th- to 8th-grade reading level is necessary to adequately comprehend the instrument. Analysis of only the 19 items that make up the core outcomes scales yielded a Flesch-Kincaid score of 5.6, requiring a 6th-grade reading level for satisfactory comprehension.

### POQ-VA Treatment Satisfaction Ratings

The basic psychometric characteristics of the five-item PTS scale were examined in both samples. Calculation of coefficient alphas revealed that the scale has good internal consistency (Sample 1,  $\alpha = 0.90$ ; Sample 2,  $\alpha = 0.83$ ). The PTS demonstrated moderate to strong associations with staff discharge ratings of patient improvement ( $r = 0.60$  and  $0.25$ ) and satisfaction ( $r = 0.41$  and  $0.31$ ), patients' 3-month follow-up ratings of their own improvement ( $r = 0.37$  and  $0.65$ ), and the extent to which they would recommend the intervention to someone with a similar problem ( $r = 0.59$  and  $0.53$ ). In addition, Sample 2 PTS scores were moderately associated with 3-month follow-up ratings of satisfaction ( $r = 0.40$ ). All correlations were significant at the 0.001 level of alpha.

### Medical Use

We evaluated the validity of the POQ-VA medical use items by examining the correspondence between VA medical record visit documentation and participants' reports of the number of VA pain-related healthcare visits in the 3 months before program admission and the 3 months following program discharge using a subset ( $n = 50$ ) of Sample 2 inpatients. Intraclass correlation coefficients revealed good agreement between participants' responses and medical records for the period before intake,  $r = 0.66$ , and adequate agreement for the period following discharge,  $r = 0.44$ .

### POQ-VA Consumer Ratings

To assess patients' impressions of the POQ-VA, we asked a subset of Sample 2 outpatients ( $n = 240$ ) to complete the PQFF following administration of the POQ-VA Intake Questionnaire. The means and SDs on the 4-point scale ranged from a low of 3.02 (0.72) ("I was able to accurately describe my pain experience") to a high of 3.36 (0.61) ("The form was easy to use"). Means for the five pain impact items ( $M = 3.06$ ,  $SD = 0.63$ ) and the four ease-of-use items ( $M = 3.25$ ,  $SD = 0.58$ ) were very similar. All scores fell at the upper end of the 4-point scale

and suggested a moderate to high degree of satisfaction with the instrument.

## DISCUSSION

Designed specifically to measure pain treatment outcomes, the POQ-VA is the first self-report instrument that assesses all the key domains identified by major accreditation bodies and professional societies [9,23]. The POQ-VA was developed with the use of the rational-empirical test development approach [26], iterative item and scale development procedures [26,27], and relatively large samples of veterans with chronic noncancer pain. The POQ-VA includes measures of pain history (descriptive information, pain experience, employment, disability status, and opioid use), average pain intensity, pain interference, emotional distress, pain-related fear, satisfaction with treatment, and medical use. The instrument is designed to be administered at multiple time points with the use of three forms (Intake, Discharge or Interim, and Follow-Up). A 19-item abbreviated form that includes only the five core outcomes scales and the Pain NRS and does not incorporate any veteran-specific content also is available, along with a generic five-item PTS scale.

The results of this study provide support for the reliability and validity of the POQ-VA scales and the POQ-VA Pain NRS as measures of chronic pain treatment outcomes among veterans. Principal components analysis of the items making up the five core outcomes scales (ADL, NA, Mobility, Vitality, and Fear) yielded a well-defined component structure characterized by consistently high primary component loadings and relatively low secondary component loadings, which should enhance core scale stability. Hair et al. suggested that the total variance accounted for by the core items (72%) was well over the 60 percent criterion that indicates an acceptable factor solution in social sciences research [22]. A CFA of this five-factor model in a cross-validation sample produced an excellent fit, providing further support for the validity and stability of these scales. POQ-VA scale correlations with concurrent measures of pain-related function revealed a pattern of strong associations with relevant external criteria (i.e., convergent validity) and weaker relationships with less relevant extra-test measures (i.e., discriminant validity). The Pain NRS score, which has been validated separately in several studies [42,57,58], demonstrated a pattern of convergent-discriminant relationships

similar to that of the core POQ-VA outcomes scales. Similarly, the obtained correlations between VA health records and participants' reports of VA healthcare visits support the validity of the medical use items. Reliability indexes (i.e., stability, consistency, and generalizability) for both the POQ-VA core scales and the Pain NRS indicated that the measures demonstrated adequate to excellent reliability, particularly when the limited number of items on the Fear and Vitality scales is considered. Similar conclusions apply to the POQ-VA PTS scale. The five-item scale had high internal consistency and was associated significantly with concurrent measures of patient satisfaction.

POQ-VA core scales and the Pain NRS score also were found to be sensitive to change in a heterogeneous group of veterans undergoing a variety of pain treatment interventions at six VA pain treatment sites. Significant improvements in all measures from pretreatment to post-treatment were obtained, while scores for 100 outpatients awaiting treatment who completed two administrations of the POQ-VA 7 to 13 days apart did not change. Consistent with most pain outcomes studies [59,60–62], follow-up assessment revealed some decline in functioning but levels remained above those reported at pretreatment. We also estimated the magnitude of POQ-VA scale scores necessary to define clinically significant change. These values can be used to help determine which patients benefit from the provided interventions.

Results of this study also support the general clinical use of the POQ-VA. Readability analyses indicated that a 7th- to 8th-grade reading level was necessary to comprehend the POQ-VA content, and that a 6th-grade reading level was sufficient to understand the items on the core outcomes scales. Mean consumer satisfaction ratings evaluating item content, questionnaire breadth, and ease of completion from a subset of veterans who completed the POQ-VA Intake Questionnaire were uniformly positive. Clinician users anecdotally reported satisfaction with the ease of administration and the use of the POQ-VA scales.

Despite these positive findings, several limitations associated with the validation of the final POQ-VA items must be noted. First, our studies almost exclusively used veterans with chronic pain. Although evidence suggests that veterans with chronic pain are in fact similar to non-veterans with chronic pain [47], additional research is needed to verify the POQ-VA's validity with nonveterans. Second, the psychometric properties of the current employment status question were not evaluated because

we lacked reliable collateral information that could be used for comparisons. Additional research will be needed to determine the reliability and validity of this item before it is used to assess treatment-related changes in employment status. Third, consistent with contemporary test construction recommendations and common practice (e.g., MMPI-2, WHYMPI) [15,26,27,29], validity and reliability data were based on final core POQ-VA items that were extracted from a larger pool of potential items. Possibly, the magnitude of some item reliabilities or correlations between POQ-VA scales and concurrent measures might have differed if only the final POQ-VA items had been administered to subjects, although the magnitude of these changes likely would be small. Nevertheless, replication of these results following administration of only the final POQ-VA items would further prove the instrument's reliability and validity. Fourth, we did not conduct evaluations of potential cultural differences in participant's POQ-VA responses in this study. Such appraisals are necessary and currently underway. In the interim, because a majority of subjects who participated in this study were Caucasian, one should exercise caution when using the POQ-VA to evaluate the treatment response of individual patients from minority groups. Last, the POQ-VA was developed primarily to measure pain treatment outcomes for those experiencing chronic noncancer pain. While the instrument may prove useful for assessing pain treatment effectiveness for patients with acute or cancer pain as well, its appropriateness for use with these populations first needs to be determined empirically.

## CONCLUSIONS

The POQ-VA was designed to serve as a primary pain outcomes tool in settings where comprehensive, multidimensional outcomes data are needed to evaluate the effectiveness of pain interventions. Data reported in this study support the reliability, validity, and clinical use of the POQ-VA for evaluating the effectiveness of treatment for veterans experiencing chronic noncancer pain. As an instrument designed specifically to assess pain treatment outcomes, it has numerous strengths. First, it includes multiple forms that parallel common phases of treatment (i.e., treatment initiation [Intake], treatment stabilization or termination [Interim or Discharge], and treatment reassessment [Follow-Up]), allowing for direct

assessment of change. Second, it is comprehensive, providing measurement of all the major pain outcomes domains, yet it is relatively brief (44 questions in the Intake or longest form). Third, it includes a detailed pain history section that is useful for pain research applications. Fourth, it requires minimal reading skills for completion and is viewed favorably by pain consumers. Fifth, it is sensitive to changes associated with pain treatment. Last, a separate POQ-VA Short Form (19 items) and a five-item PTS questionnaire are available for settings where brevity is a primary concern. Overall, the POQ-VA is a useful alternative for clinicians and researchers interested in the evaluation of the pain intervention effectiveness.

## ACKNOWLEDGMENTS

We would like to thank the staff of the AAPM for their assistance during this project and Karen M. Milo, PhD, for her helpful comments on an earlier version of this manuscript. We also want to express our appreciation to Caridad Bravo-Fernandez, MD; Judith Chapman, PhD; Ramon Cuevas, MD; Manuel Garcia, PhD; and Timothy Lawler, PhD, for serving as local principal investigators during field testing of the instrument.

## REFERENCES

- Clark ME, Gironde RJ. Practical utility of outcome measurement. In: Weiner RS, editor. *Pain management: A practical guide for clinicians*. 6th ed. Boca Raton (FL): CRC Press; 2002.
- Veterans' Health Administration. VHA national pain management strategy. *Pain Assessment: The 5th vital sign*. Washington (DC): Veterans' Health Administration; 2000. p. 27–29.
- Brown D. Quality assessment and improvement activities should be incorporated into our pain practices. *Pain Forum* 1995;4:48–56.
- Johnston MV, Keith RA, Hinderer SR. Measurement standards for interdisciplinary medical rehabilitation. *Arch Phys Med Rehabil* 1992;73(12-S):S3–23.
- Pasero CL, Hubbard L. Development of an acute pain service monitoring and evaluation system. *Qual Rev Bull* 1991; 17(12):396–401.
- Stanton-Hicks M. Treatment of sympathetically maintained pain. *Reg Anesth* 1995;20(1):1,2.
- Turk DC, Rudy TE, Sorkin BA. Neglected topics in chronic pain treatment outcome studies: Determination of success. *Pain* 1993;53(1):3–16.
- Turk DC, Flor H. Chronic pain: A biobehavioral perspective. In: Gatchel RJ, Turk DC, editors. *Psychosocial factors in pain: Critical perspectives*. New York: The Guilford Press; 1999. p. 18–34.
- Rehabilitation Accreditation Commission. Ed. *Standards manual: Medical rehabilitation*. Rev. ed. Tucson (AZ): Rehabilitation Accreditation Commission; 2002.
- Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66(8):271–73.
- Jensen MP, Strom SE, Turner JA, Romano JM. Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. *Pain* 1992;50(2):157–62.
- Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: A new tool for rehabilitation. *Adv Clin Rehabil* 1987;1:6–18.
- Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30(6):473–83.
- Suarez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D. Use of health status measures in patients with low back pain in clinical settings: Comparison of specific, generic, and preference-based instruments. *Rheumatology* 2000;39(7):783–90.
- Kerns RD, Turk DC, Rudy TE. The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 1985; 23(4):345–56.
- Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994; 23(2):129–38.
- Altmaier EM, Lehmann TR, Russell DW, Weinstein JN, Kao CF. The effectiveness of psychological interventions for the rehabilitation of low back pain: A randomized controlled trial evaluation. *Pain* 1992;49(3):329–35.
- Kerns RD, Turk DC, Holzman AD, Rudy TE. Comparison of cognitive-behavioral and behavioral approaches to the outpatient treatment of chronic pain. *Clin J Pain* 1986;1: 195–203.
- Deisinger JA, Cassisi JE, Lofland KR, Cole P, Bruehl S. An examination of the psychometric structure of the Multidimensional Pain Inventory. *J Clin Psychol* 2001;57(6): 765–83.
- Riley JL 3rd, Zawacki TM, Robinson ME, Geisser ME. Empirical test of the factor structure of the West Haven-Yale Multidimensional Pain Inventory. *Clin J Pain* 1999; 15(1):24–30.
- Bernstein IH, Jaremko ME, Hinkley BS. On the utility of the West Haven-Yale Multidimensional Pain Inventory. *Spine* 1995;20(8):956–63.

22. Hair JF Jr, Anderson RE, Tatham RL. *Multivariate data analysis with readings*. 2nd ed. New York: Macmillan; 1987.
23. International Association for the Study of Pain. *Core curriculum for professional education in pain*. 2nd ed. Seattle (WA): International Association for the Study of Pain; 1991.
24. American Pain Society Quality of Care Committee. *Quality improvement guidelines for the treatment of acute pain and cancer pain*. *JAMA* 1995;274:1874–80.
25. UCSD Health Outcomes Assessment Program. *Report of American Academy of Pain Management "health-related quality of life" questions (as part of National Pain Data Bank)*. Sonora (CA): American Academy of Pain Management; 1998.
26. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychol Assess* 1995;7(3):309–19.
27. Green BF. A primer of testing. In: Kazdin AE, editor. *Methodological issues and strategies in clinical research*. 1st ed. Washington (DC): American Psychological Association; 1992.
28. Clark ME, Gironde RJ. Concurrent validity of the National Pain Data Bank: Preliminary results. *Am J Pain Manage* 2000;10:25–33.
29. Butcher JN, Dahlstrom WG, Graham JR, Tellegen A, Kaemmer B. *Minnesota multiphasic personality inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis (MN): University of Minnesota Press; 1989.
30. Butcher JN, Graham JR, Ben-Porath YS, Tellegen A, Dahlstrom WG, Kaemmer B. *MMPI-2 (Minnesota multiphasic personality inventory-2): Manual for administration, scoring, and interpretation*. Rev ed. Minneapolis (MN): University of Minnesota Press; 2001.
31. Graham JR. *MMPI-2: Assessing personality and psychopathology*. 3rd ed. New York: Oxford University Press; 2000.
32. Jenkins CD, Stanton BA, Niemcryk SJ, Rose RM. A scale for the estimation of sleep problems in clinical research. *J Clin Epidemiol* 1988;41:313–21.
33. Kori SH, Miller RP, Todd DD. Kinesiophobia: A new view of chronic pain behavior. *Pain Manag* 1990;3:35–43.
34. Clark ME, Kori SH, Broeckel J. Kinesiophobia and chronic pain: Psychometric characteristics and factor analysis of the Tampa Scale. 15th Annual Scientific Meeting of the American Pain Society. Washington, DC; 1996.
35. Crombez G, Vlaeyen JW, Heuts PH, Lysens R. Pain-related fear is more disabling than pain itself: Evidence on the role of pain-related fear in chronic back pain disability. *Pain* 1999;80(1–2):329–39.
36. Williams RC. Toward a set of reliable and valid measures for chronic pain assessment and outcome research. *Pain* 1988;35(3):239–51.
37. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: Development and final revision of a health status measure. *Med Care* 1981;19(8):787–805.
38. Turner JA, Clancy S. Comparison of operant behavioral and cognitive-behavioral group treatment for chronic low back pain. *J Consult Clin Psychol* 1988;56(2):261–66.
39. Beckham JC, Burker EJ, Lytle BL, Feldman ME, Costakis MJ. Self-efficacy and adjustment in cancer patients: A preliminary report. *Behav Med* 1997;23(3):138–42.
40. Watson JH, Graydon JE. Sickness Impact Profile: A measure of dysfunction with chronic pain patients. *J Pain Symptom Manage* 1989;4:152–56.
41. Breivik EK, Bjornsson GA, Skovlund E. A comparison of pain rating scales by sampling from clinical trial data. *Clin J Pain* 2000;16:22–28.
42. De Conno F, Caraceni A, Gamba A, Mariani L, Abbattista A, Brunelli C, La Mura A, Ventafridda V. Pain measurement in cancer patients: A comparison of six methods. *Pain* 1994;57(2):161–66.
43. Ogon M, Krismer M, Sollner W, Kantner-Rumplmair W, Lampe A. Chronic low back pain measurement with visual analogue scales in different settings. *Pain* 1996;64(3):425–28.
44. Krames ES. Intrathecal infusional therapies for intractable pain: patient management guidelines. *J Pain Symptom Manage* 1993;8(1):36–46.
45. Awonuga A, Waterstone J, Oyesanya O, Curson R, Nargund G, Parsons J. A prospective randomized study comparing needles of different diameters for transvaginal ultrasound-directed follicle aspiration. *Fertil Steril* 1996;65(1):109–13.
46. Woods DA. Rehabilitation aquatics for low back injury: Functional gains or pain reduction? *Clin Kinesiol* 1989;43:96–102.
47. Clark ME. MMPI-2 Negative Treatment Indicators Content and Content Component scales: Clinical correlates and outcome prediction for men with chronic pain. *Psychol Assess* 1996;8:32–38.
48. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR, editors. *Reading and understanding multivariate statistics*. Washington (DC): American Psychological Association; 1995. p. 99–136.
49. Tabachnick BG, Fidell LS. Principal components and factor analysis. In: Tabachnick BG, Fidell LS, editors. *Using multivariate statistics*. 3rd ed. New York: Harper Collins; 1996. p. 635–708.
50. Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, editors. *Handbook of pain assessment*. 2nd ed. New York: The Guilford Press; 2001. p. 15–34.
51. AMOS [computer program]. Arbuckle J, Version 4.0. Chicago (IL): SmallWaters Corp; 1999.

52. Hu Li-tze, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equation Model* 1999;6(1):1–55.
53. Cohen J. Set correlation and contingency tables. *Appl Psychol Meas* 1988;12(4):425–34.
54. Girona RJ, Azzarello L, Clark ME. Test-retest reliability of the National Pain Data Bank. Version 2.0. *Am J Pain Manage* 2002;12:24–30.
55. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *Am J Ment Defic* 1981;86:127–37.
56. Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. In: Kazdin AE, editor. *Methodological issues and strategies in clinical research*. Washington (DC): American Psychological Association 1992; p. 631–48.
57. Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. *Pain* 1999;83(2):157–62.
58. Paice JA, Cohen FL. Validity of a verbally administered numeric rating scale to measure cancer pain intensity. *Cancer Nurs* 1997;20(2):88–93.
59. Berman BM, Singh BB, Lao L, Langenberg P, Li H, Hadhazy V, Bareta J, Hochberg M. A randomized trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. *Rheumatology* 1999;38(4):346–54.
60. Hawley DJ. Psycho-educational interventions in the treatment of arthritis. *Baillieres Clin Rheumatol* 1995;9(4):803–23.
61. Mannon AF, Muntener M, Taimela S, Dvorak J. Comparison of three active therapies for chronic low back pain: Results of a randomized clinical trial with one-year follow-up. *Rheumatology* 2001;40(7):772–78.
62. van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow-up. *Ann Rheum Dis* 2001;60(12):1123–30.

Submitted for publication May 1, 2003. Accepted in revised form August 1, 2003.