

# The Shoulder Pain and Disability Index: The Construct Validity and Responsiveness of a Region-Specific Disability Measure

**Background and Purpose.** The purposes of this study were (1) to assess the construct validity of the Shoulder Pain and Disability Index (SPADI) and (2) to determine whether the SPADI is more responsive than the Sickness Impact Profile (SIP), a generic health status measure. **Subjects.** The sample consisted of 94 patients who were diagnosed with a shoulder problem and referred to six outpatient physical therapy clinics. **Methods.** Clinically meaningful change was determined by use of an ordinal rating scale designed to determine whether the patient's shoulder function was improved, the same, or worse following treatment. Spearman rho correlations were calculated for the initial visit SPADI and SIP scores. The standardized response mean (SRM) was used to measure responsiveness for the patients who were judged to be improved. One-tailed paired *t* tests ( $\alpha=.01$ ) were used to determine whether differences existed among SRM values. **Results.** Correlations between the SPADI and SIP scores ranged from  $r=.01$  to  $r=.57$ . The SRM value was higher for the SPADI total score (SRM=1.38) than for the SIP total score (SRM=0.79). **Conclusion and Discussion.** Most correlations between SPADI and SIP scores provided support for the construct validity of the SPADI. The SPADI does not appear to strongly reflect occupational and recreational disability and is more responsive than the SIP. [Heald SL, Riddle DL, Lamb RL. The Shoulder Pain and Disability Index: the construct validity and responsiveness of a region-specific disability measure. *Phys Ther.* 1997;77:1079–1089.]

**Key Words:** *Disability, Functional status, Health status, Responsiveness, Shoulder, Sickness Impact Profile, Validity.*

*Susan L Heald*

*Daniel L Riddle*

*Robert L Lamb*

Increasingly, third-party payers are requiring documentation of the health-related quality of life (HRQL) of patients who receive physical therapy services.<sup>1,2</sup> *Health-related quality of life* refers to an individual's ability to perform tasks of everyday living and to fulfill various social roles that are satisfying to that individual.<sup>3,4</sup> Many scales have been developed to assess a patient's HRQL.<sup>5-8</sup> Measurement of HRQL can encompass a wide variety of variables, including activities of daily living (ADL), social roles, emotional state, intellectual functioning, and state of perceived well-being.<sup>5-8</sup>

In addition to generic HRQL scales, other scales have been designed for measuring disability and can be classified as condition-specific scales (eg, Arthritis Impact Measurement Scale<sup>6</sup>) or region-specific scales (eg, Oswestry Low Back Pain Questionnaire<sup>9</sup>). Several region-specific scales have been designed for the shoulder.<sup>10-15</sup> Data for reliability and validity have been reported for only one of the region-specific scales designed for the shoulder: the Shoulder Pain and Disability Index (SPADI).<sup>14,16</sup> The SPADI is a self-administered questionnaire that consists of two dimensions, one for pain and the other for functional activities, and requires 5 to 10 minutes for a patient to complete. The pain dimension consists of five questions regarding the severity of an individual's pain. Functional activities are assessed with eight questions designed to measure the degree of difficulty an individual has with various ADL that require upper-extremity use. To answer the questions, patients place a mark on a 10-cm visual analog scale for each question. Verbal anchors for the pain dimension are "no pain at all" and "worst pain imaginable," and those for the functional activities are "no difficulty" and "so difficult it required help." The scores from both dimensions are averaged to derive a total score. Table 1 lists the SPADI items.

Roach et al<sup>14</sup> provided evidence for the test-retest reliability of the total scores and scores for both dimensions of the SPADI based on data from 23 male subjects (intraclass correlation coefficients=.64-.66). Internal

**Table 1.**  
Shoulder Pain and Disability Index<sup>4</sup>: Items Listed by Dimension

<p>Pain dimension: How severe is your pain?</p> <ol style="list-style-type: none"> <li>1. At its worst?</li> <li>2. When lying on the involved side?</li> <li>3. Reaching for something on a high shelf?</li> <li>4. Touching the back of your neck?</li> <li>5. Pushing with the involved arm?</li> </ol> <p>Disability dimension: How much difficulty do you have?</p> <ol style="list-style-type: none"> <li>1. Washing your hair?</li> <li>2. Washing your back?</li> <li>3. Putting on an undershirt or pullover sweater?</li> <li>4. Putting on a shirt that buttons down the front?</li> <li>5. Putting on your pants?</li> <li>6. Placing an object on a high shelf?</li> <li>7. Carrying a heavy object (eg, 10 lb)?</li> <li>8. Removing something from your back pocket?</li> </ol>
--

consistency was good, with Cronbach's alpha values of .86 to .95.<sup>14</sup> To examine the construct validity of the SPADI, Roach et al<sup>14</sup> performed a factor analysis with and without varimax rotation. The factor analysis without rotation resulted in all items from both dimensions loading strongly onto one factor. The SPADI therefore appears to measure one construct. Results of the varimax rotation showed several functional items loading onto both of two factors, indicating that the division between the two dimensions may not be warranted.<sup>17</sup> These results provided some evidence to support the construct validity of the SPADI, but the factor analysis suggested that the scale may not reflect two separate dimensions.

Roach et al<sup>14</sup> examined the criterion-related validity of the SPADI by using measurements of shoulder active range of motion (AROM) as criteria for function. Because there is no research to indicate that shoulder AROM measurements are related to function, we do not believe that the authors provided evidence for the criterion-related validity of the SPADI.

The lack of an adequate criterion measure for disability has created difficulties when evaluating the validity of region-specific questionnaires. Investigators<sup>18-22</sup> have

SL Heald, PT, is Physical Therapist, Rehabilitation Services, Martha Jefferson Hospital, 459 Locust Ave, Charlottesville, VA 22901 (USA). Ms Heald was a candidate for the Master of Science degree in physical therapy at Medical College of Virginia, Virginia Commonwealth University, at the time this research was conducted. Address all correspondence to Ms Heald.

DL Riddle, PhD, PT, is Associate Professor, Department of Physical Therapy, School of Allied Health, Medical College of Virginia, Virginia Commonwealth University, Richmond, Va.

RL Lamb, PhD, PT, is Professor and Chairman, Department of Physical Therapy, Medical College of Virginia, Virginia Commonwealth University.

This study was approved by the Virginia Commonwealth University Committee on the Conduct of Human Research.

*This article was submitted April 5, 1996, and was accepted April 25, 1997.*

compared newly developed region-specific and condition-specific scales with generic HRQL scales that have established reliability and validity. Some authors<sup>23</sup> contend that if moderate to strong correlations are found between the new scales and the established scales, the construct validity of the new scales for measuring some aspect of HRQL is supported. We believe that an established HRQL scale is an acceptable measure for determining whether the SPADI has construct validity for making inferences about the extent of an individual's disability.

A recently published study<sup>16</sup> examined the construct validity of measurements obtained with the SPADI on a sample of 102 patients (98 male, 4 female) with shoulder problems. The patients' ages ranged from 47 to 66 years ( $\bar{X}=60$ ). Williams et al<sup>16</sup> reported correlations between the SPADI and the SF-20 ( $r=-.25$  to  $r=-.50$ ) and the Health Assessment Questionnaire ( $r=.61$ ). These correlations supported the construct validity of the SPADI for inferring the magnitude of disability. Although Williams et al<sup>16</sup> provided evidence for the construct validity of the SPADI based on correlations with established HRQL scales, they did not control for comorbidity due to pathologies other than those of the shoulder. Ideally, when evaluating the validity of SPADI scores, the subjects should not have other health problems influencing their HRQL. The presence of comorbidity has been known to have varying effects on specific and generic questionnaires.<sup>24,25</sup>

The study by Williams et al<sup>16</sup> is an example of how an established HRQL scale can be used to determine the validity of a newer region-specific disability scale. Because of its ubiquitous acceptance, the Sickness Impact Profile (SIP) has frequently been used by researchers for assessing the validity of new disability measures.<sup>19,20,22</sup> The SIP is a generic HRQL scale that has been widely studied and established as a scale with acceptable levels of reliability and validity.<sup>5,23,26-28</sup> Test-retest reliability data have produced Pearson product-moment correlation coefficients ( $r$ ) of .75 to .92 for the overall score.<sup>23</sup> Cronbach's alpha values for internal consistency were high, ranging from .94 to .97.<sup>23</sup>

One purpose of our study was to determine the construct validity of measurements obtained with the SPADI for inferring the extent of disability. To assess the construct validity of the SPADI, we examined the associations between scores on the SPADI and the SIP. The SPADI is used to calculate a total score, a pain score, and a disability score. The SIP is used to calculate a total score, physical dimension and psychosocial dimension scores, and several category scores. The categories of ambulation, mobility, and body care and movement make up the physical dimension. The categories of social

interaction, communication, alertness behavior, and emotional behavior make up the psychosocial dimension. Separate scores are also calculated for the categories of sleep and rest, eating, work, home management, and recreation and pastimes.

We hypothesized that we would find the following associations:

1. Correlations of the SPADI total, pain, and disability scores with scores from those categories of the SIP that relate to shoulder function would be .5 or higher. Many researchers who have compared region-specific or condition-specific scales with generic scales have suggested that correlations of .5 or higher support the construct validity of the region- or condition-specific instrument.<sup>16,29,30</sup> The SIP scores that appear to be related to shoulder function are the total score, the physical dimension score, and the scores for the categories of body care and movement, sleep and rest, work, home management, and recreation and pastimes.
2. Correlations of the SPADI total, pain, and disability scores with scores from those categories of the SIP that do not relate to shoulder function would be .3 or lower. The SIP scores that appear to be unrelated or only weakly related to shoulder function are the psychosocial dimension score and the scores for the categories of ambulation, mobility, social interaction, communication, alertness behavior, emotional behavior, and eating.

Recent literature has also emphasized the importance of examining a score's responsiveness. *Responsiveness* is a measurement's capacity to reflect clinically meaningful changes.<sup>23,25</sup> For the purposes of this study, a *clinically meaningful change* was defined as the agreement between independent therapist and patient judgments that the patient's shoulder function had improved or worsened after treatment.<sup>19,27,32</sup> Measurements obtained from an HRQL or region-specific disability measure may be reliable and concurrently valid yet not reflect clinically meaningful changes that occur over time.<sup>34,35</sup>

Condition-specific and region-specific scales are thought to be more responsive to meaningful changes than generic HRQL scales.<sup>31,32,34</sup> If the SPADI is a more responsive scale than generic scales and is valid for inferring the extent of disability, then physical therapists should use the SPADI to assess the disability of patients with shoulder problems. Whether the SPADI is more responsive than the SIP is not known.

The second purpose of this study was to determine whether the SPADI is more responsive than the SIP to

meaningful changes in individuals with shoulder pathology.

## Method

### Subjects

Subjects recruited for this study were patients seen at six outpatient physical therapy clinics being used as data collection centers. Patients included were consecutive patients referred for outpatient physical therapy for their shoulder pain. The shoulder region was defined as the glenohumeral joint, acromioclavicular joint, sternoclavicular joint, scapulothoracic joint, and all structures crossing or attaching to these joints. The pain had to have been present for at least 1 week.

The following patients were excluded from the study: (1) patients who were cognitively impaired, illiterate, or did not speak English as a primary language, (2) patients referred for physical therapy for any region of the body in addition to the shoulder, (3) patients who had concurrent pain or impairment in both upper extremities but who were referred for treatment of only one upper extremity, (4) patients who had shoulder pain or dysfunction that, based on the physical therapists' judgment, was caused by pathology in an anatomical area other than the shoulder complex, (5) patients who had, in the therapists' judgment, decreased HRQL due to a systemic disease or pathology in addition to their shoulder pathology, and (6) patients who did not complete the data collection forms within 24 hours of their initial or final visit.

Over a 16-month period, 183 patients were deemed eligible and were recruited to participate in the study. Twenty-seven of these patients refused to participate, and 45 patients were eliminated by the therapists based on the exclusionary criteria. Eight patients did not return their completed initial visit forms. The remaining 103 patients completed questionnaires on the day of their initial visit. Three of the 103 patients did not complete the SPADI correctly, so they were eliminated from the sample. Six patients scored "not applicable" ("NA") on at least one dimension of the SPADI, so these patients were eliminated from the sample. Data from the remaining 94 patients were used for analysis of the construct validity of the SPADI. Table 2 presents a description of the subjects included in the evaluation of construct validity. The patients' ages ranged from 19 to 82 years ( $\bar{X}=44.8$ ,  $SD=14.0$ ). Diagnoses varied widely, with "impingement/tendinitis" being the most common diagnosis. The average duration of physical therapy for the patients was 10 weeks.

Five of the 94 patients did not return for a final visit. Fifteen patients were given forms during their final visit,

**Table 2.**

Description of Subjects Included in the Analysis of Construct Validity

Characteristic	No. of Subjects (n=94)
Gender	
Male	59
Female	32
Not reported	3
Therapist's diagnosis	
Impingement/tendinitis/bursitis	28
Instability/dislocation	13
Rotator cuff syndrome	13
After arthroscopic surgery	10
Pain/stiffness	6
Adhesive capsulitis/frozen shoulder	5
After rotator cuff repair	4
After fracture	3
Sternoclavicular or acromioclavicular joint subluxation	1
Contusion	1
Weakness	1
Diagnosis not reported	1

but the forms were not returned. The examiners gave no explanation as to why another 38 patients did not complete the forms during their final visit. A total of 36 patients completed data forms during their final visit.

### Examiners

Examiners were physical therapists employed at six physical therapy clinics. A total of 28 therapists collected data at the six data collection centers. Fifty-seven percent of the therapists were female and 43% were male. The average age of the examiners was 30.8 years ( $SD=3.7$ ). The therapists had an average of 5.8 years ( $SD=4.2$ ) of clinical experience in the area of orthopedic physical therapy.

### Instrumentation

The SPADI is a self-administered questionnaire accompanied by a set of written instructions that takes approximately 5 to 10 minutes to complete. Each question requires the patient to place a mark on a 10-cm visual analog scale that has verbal anchors on each end of the line. Questions on the SPADI are divided into two dimensions (Tab. 3).

The SIP consists of 136 items describing activities that can be affected by health status. Table 3 provides a complete list of categories and dimensions addressed by the SIP. The SIP may be self-administered and takes approximately 20 to 30 minutes to complete. Patients are asked to mark each item that describes their own health-related behavior.

To provide a clinical criterion for judging the meaningfulness of changes in the SIP and the SPADI, we created

**Table 3.**Description of Initial-Visit Shoulder Pain and Disability Index<sup>14</sup> and Sickness Impact Profile<sup>5</sup> Scores<sup>a</sup>

Instrument	$\bar{X}$	SD	Range
Shoulder Pain and Disability Index			
Pain dimension	46.3	27.5	0-100
Disability dimension	33.9	28.1	0-100
Total	40.1	26.1	0-100
Sickness Impact Profile			
Ambulation	0.6	2.3	0-14.4
Mobility	1.4	4.4	0-28.0
Body care and movement	2.3	4.4	0-35.3
Physical dimension			
Social interaction	1.2	2.9	0-19.4
Communication	3.8	9.6	0-73.8
Alertness behavior	1.5	4.6	0-23.7
Emotional behavior	1.7	6.2	0-41.8
Psychosocial dimension			
Sleep and rest	8.7	12.9	0-54.6
Eating	4.3	7.0	0-27.0
Work	8.8	12.9	0-79.2
Home management	0.9	3.2	0-22.3
Recreation and pastimes	11.8	23.5	0-70.1
Total	11.5	14.2	0-66.5
	15.6	14.9	0-72.3
Total	4.5	5.8	0-26.6

<sup>a</sup>All subjects who completed data forms during initial visit (n=94).

an ordinal scale with the following levels of measurement: improved, stayed the same, and worsened. This scale is very similar to the one used by Deyo and Centor<sup>32</sup> to assess the responsiveness of health status measures used for patients with low back pain. Patients and therapists in our study each made independent judgments regarding the patient's shoulder function during the final clinic visit. A meaningful change was considered to have occurred if the patient and the therapist agreed in their independent judgments that the patient's shoulder function had either improved or worsened.<sup>32</sup>

### Procedure

Data were collected on the day of the patients' initial visit and again during their final visit to the physical therapy clinic. Between the patients' initial and final visits, they received physical therapy as the therapist deemed appropriate. No physical therapy evaluation or treatment variables were controlled.

During their initial visit to the clinic, patients first signed a written informed consent form and then were given a SPADI and an SIP to complete within 24 hours. During the final visit, the patients again completed an SPADI and an SIP. Patients and therapists also made independent judgments during the final visit as to whether the patients' shoulder function was improved, worsened, or stayed the same.

### Data Analysis

The SPADI and the SIP were scored according to scoring systems described by the developers of the scales.<sup>11\*</sup> The range of possible SPADI and SIP scores was 0 to 100, with higher scores indicating greater amounts of disability.

To determine the construct validity of the SPADI, the Spearman's rank-order correlation<sup>†</sup> was used to describe the relationship between SPADI and SIP scores obtained during the initial visit. All dimensions of the SPADI and the SIP were examined.

The responsiveness of the SPADI and the SIP to meaningful change was determined by calculating the standardized response mean (SRM).<sup>36</sup> The SRM is a variation of the effect size described by Cohen.<sup>37</sup> The SRM is calculated by subtracting each patient's initial score from the final score to obtain the change score. The mean of the patients' score changes is then divided by the standard deviation of the patients' change scores. Cohen suggests that an absolute value of 0.2 to 0.4 represents a small effect of an intervention, an absolute value of 0.5 to 0.7 shows a moderate effect, and an absolute value of 0.8 or greater is a large effect. To perform a statistical comparison of SRMs for the SIP and the SPADI, the sampling distribution of the SRM values was needed. The jackknife procedure was used to estimate the sampling distributions.<sup>36</sup> The jackknife procedure is designed to estimate the standard error for a sample of data. The estimated variances calculated with the jackknife procedure were used to determine 95% confidence intervals for the SRMs.

Paired *t* tests corrected for multiple comparisons were performed on the estimated sample means of SRM values to determine the statistical significance of the differences in responsiveness between the SIP and SPADI scores.

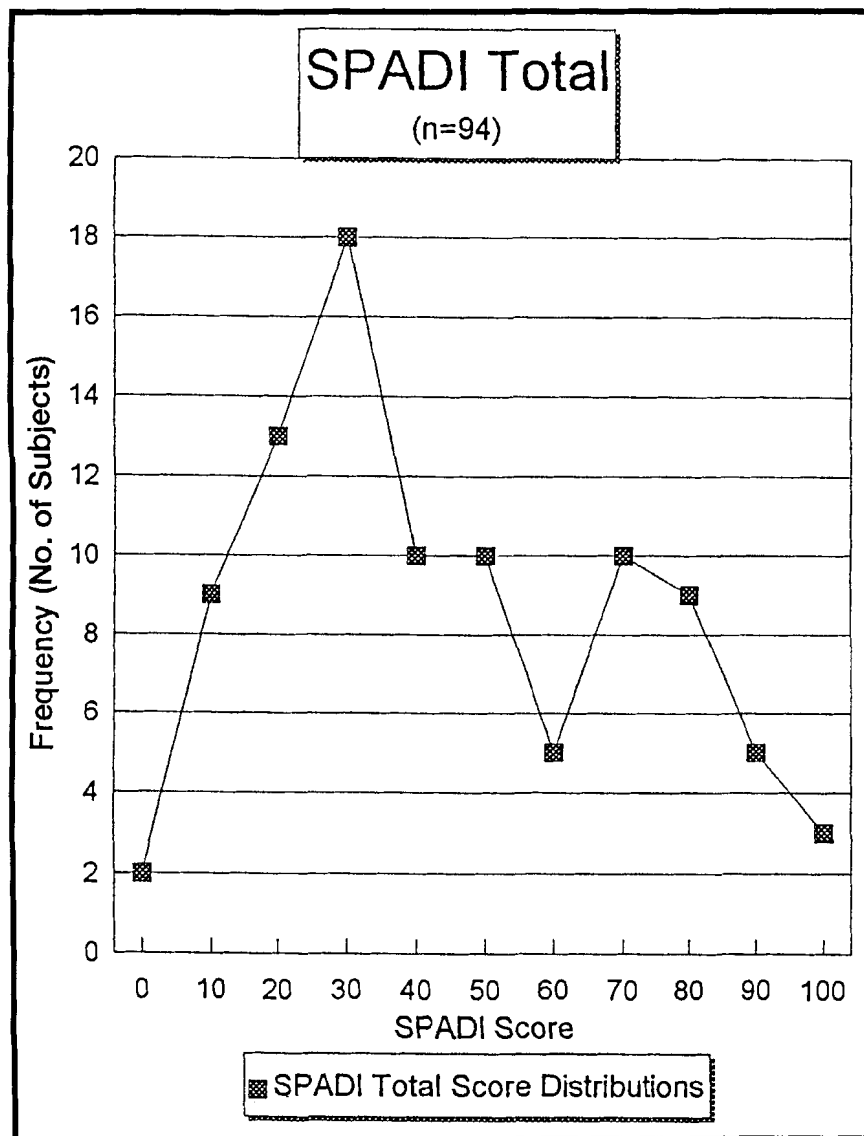
### Results

#### Descriptive Analysis of the Data

Descriptive statistics for the SPADI and SIP scores obtained during the initial visit are summarized in Table 3. Mean scores on the SPADI were higher for the pain dimension ( $\bar{X}$ =46.3, SD=27.5, range=0-100) than for the disability dimension ( $\bar{X}$ =33.9, SD=28.1, range=0-100). Mean scores on the SIP ranged from 0.6 (SD=2.3) for ambulation to 15.6 (SD=14.9) for recreation and pastimes. The SPADI total scores for the initial visit range from 0 to 100 ( $\bar{X}$ =40.1, SD=26.1). The SIP total scores ranged from 0 to 26.6 ( $\bar{X}$ =4.5, SD=5.8). Standard

\* Scoring instructions for the SIP were provided by Johns Hopkins University.

† SYSTAT for Windows, version 5, SPSS Inc, 444 N Michigan Ave, Chicago, IL, 60611.



**Figure 1.** Frequency distribution of Shoulder Pain and Disability Index<sup>14</sup> (SPADI) total scores.

deviations of both the SIP and SPADI scores were relatively high.

The SPADI pain dimension and total scores appeared to resemble a normal distribution. Figure 1 illustrates the distribution of SPADI total scores. The SIP scores did not resemble a normal distribution. The distributions of most of the SIP scores were skewed to the low end of the scale, in a J-shaped curve.<sup>5</sup> Figure 2 provides an example of the J-shaped curve of SIP physical dimension score distributions. For example, 64% of the patients scored 0 on the SIP physical dimension, with no one scoring as high as 30 (Fig. 2). Only the recreation and pastimes and home management categories exhibited less skewed distributions. Forty percent of the patients scored 0 on the home management category, and 31% of the patients scored 0 on the recreation and pastimes category.

### Construct Validity of the SPADI

Correlations between the SPADI total, pain dimension, and disability dimension scores and the SIP scores hypothesized to be related to the SPADI scores ranged from .21 to .57. Correlations between the SPADI scores and the SIP scores hypothesized to be unrelated or only weakly related to the SPADI scores ranged from .02 to .43. Table 4 presents the correlations between the SPADI and SIP scores.

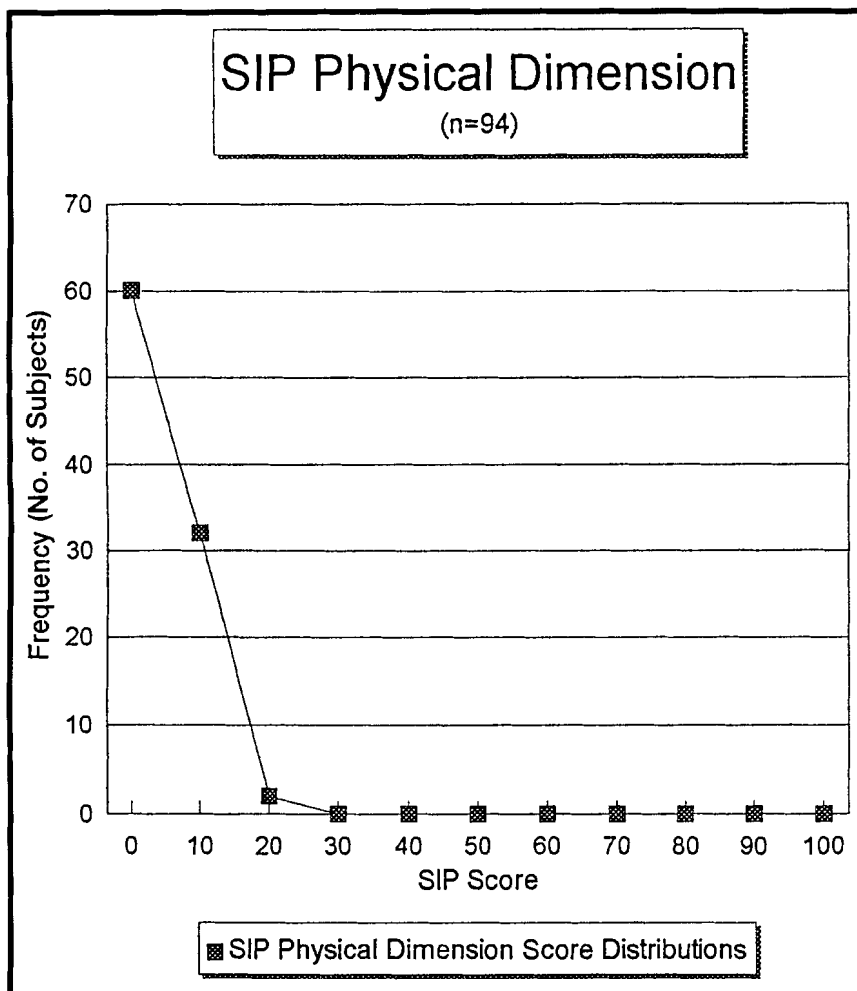
### Responsiveness

Scores from patients who completed forms during both initial and final visits were included in the analysis of responsiveness. There was consensus between therapist and patient judgments on all patients. Only two patients were classified as showing no meaningful change, and no patients were reported as having worsened shoulder function. Only those subjects who showed meaningful improvement in shoulder function were included in the analysis of responsiveness (n=34). Table 5 shows the SPADI and SIP change scores, standard deviations of changes, SRMs, and 95% confidence intervals for the SRMs. The absolute values of the SRMs are presented in Table 5.<sup>36</sup>

The SRMs for the SPADI ranged from 1.04 to 1.54 (Tab. 5). The SRMs for the SIP ranged from 0.00 to 1.10, with the majority of values being below 0.5 (Tab. 5). Paired *t* tests corrected for multiple comparisons were used to determine

the statistical differences in responsiveness between the SIP and SPADI scores. The *t* tests were performed between the SPADI pain dimension, disability dimension, and total scores and those categories of the SIP that are traditionally used to infer the extent of disability (physical dimension, psychosocial dimension, total). The *t* tests were also done on the SRMs for those categories of the SIP with the lowest percentage of patients scoring 0 (sleep and rest, home management, work, and recreation and pastimes). Those SIP categories were considered to be most applicable to patients with shoulder problems.

There generally were differences in responsiveness between the SIP physical dimension, psychosocial dimension, and total scores and the SPADI total scores. The SPADI pain dimension, disability dimension, and



**Figure 2.** Frequency distribution of Sickness Impact Profile<sup>5</sup> (SIP) physical dimension scores.

total scores were not more responsive than the scores for the SIP recreation and pastimes category.

## Discussion

### Construct Validity of the SPADI

The data provide some support for our hypotheses that SPADI scores would be more strongly correlated with certain SIP scores. For example, the correlations between the SIP total, home management, and body care and movement scores and the SPADI scores all approximated the hypothesized value of .5 ( $\rho = .44-.57$ ). Correlations between other SIP scores and the SPADI scores, however, were lower than .5. The SIP physical dimension, recreation and pastimes, and work scores were only weakly correlated with the SPADI scores ( $\rho = .21-.43$ ).

The physical dimension score of the SIP is a summary score that includes the ambulation and mobility dimension scores, two dimension scores with lower correlations than the hypothesized values. Because the physical

dimension score is a summary score, the correlations between SPADI scores and the SIP physical dimension score were not as high as we had hypothesized. The correlations between SPADI scores and the work and recreation and pastimes scores of the SIP were lower than we expected. The SPADI does not appear to adequately measure occupational and recreational disability.

The data also provide some support for our hypothesis that certain SIP scores would be weakly related to SPADI scores. Correlations between SPADI scores and those SIP scores that we hypothesized to be weaker or absent ranged from .01 to .43. Correlations between the SPADI scores and the SIP eating, mobility, and alertness behavior scores were the lowest of all the measures ( $\rho = .01-.20$ ). Correlations between the SPADI scores and some of the scores of the SIP that deal with emotional health, however, were somewhat higher ( $\rho = .32-.43$ ). The SPADI appears to measure some elements of psychosocial disability, although to what extent is unclear from this study.

The literature examining the construct validity of region-specific disability scales as compared with HRQL scales generally describes correlation values greater than .5 as being supportive of the construct validity.<sup>16,22,29,30</sup> Weinberger et al<sup>22</sup> concluded that the Arthritis Impact Measurement Scale is valid for inferring disability based on the correlations with the SIP. Other researchers<sup>19,38</sup> have drawn similar conclusions when attempting to validate scales designed to measure the extent of disability in patients with low back pain. Perhaps correlations with the SIP were higher in these other studies because of the nature of the pathology involved. Because low back pain and arthritis tend to have a more global effect on an individual's HRQL, it is likely that more items on the SIP were applicable to these patients.

The correlations found in our study may also be attributable to the skewed distribution of SIP scores toward the low end of the scale. If more of the items on the SIP were applicable to patients with shoulder pathology, there may have been less of a ceiling effect on the SIP scores, which may have resulted in improved correlations between the SIP and SPADI scores. The percentage of patients scoring 0 on the SIP during the initial visit

**Table 4.**

Spearman Correlations Between Sickness Impact Profile<sup>5</sup> and Shoulder Pain and Disability Index<sup>14</sup> Scores<sup>a</sup> at the Initial Visit

	Shoulder Pain and Disability Index		
	Pain	Disability	Total
Sickness Impact Profile			
Ambulation	.16	.28	.23
Mobility	.17	.20	.20
Body care and movement	.44	.48	.49
Physical dimension	.30	.43	.39
Social interaction	.33	.37	.38
Communication	.32	.32	.32
Alertness behavior	.12	.17	.14
Emotional behavior	.36	.34	.39
Psychosocial dimension	.37	.42	.43
Sleep and rest	.42	.36	.42
Eating	.03	.01	.02
Work	.32	.34	.36
Home management	.51	.49	.56
Recreation and pastimes	.21	.24	.25
Total	.50	.54	.57

<sup>a</sup>All subjects who completed data forms during initial visit (n=94).

ranged from 5% (SIP total) to 94% (ambulation). In contrast, 4% of the patients scored 0 on the SPADI pain dimension, 6% of the patients scored 0 on the disability dimension, and 1% of the patients scored 0 on the SPADI total.

Examination of the individual items on the SIP shows that relatively few of the SIP items were considered to be relevant to patients with shoulder problems. For example, within the ambulation category, all 12 statements refer to the individual's ability to walk. None of these items, theoretically, would apply to someone with disability resulting from shoulder pathology.

Conversely, there were several items on the SIP that were marked frequently by the patients in this study. Table 6 shows the response frequencies of the six SIP items selected most often by the patients with shoulder pathology. More individuals marked items in the home management and recreation and pastimes categories than other categories. The home management and recreation and pastimes categories also had higher mean scores than most other categories. The large number of patient responses to these items can be explained by examining the individual items.

For example, an item checked by 49% of the patients was the SIP home management item "I am doing less of the regular daily work around the house than I would usually do." Another item frequently checked in the recreation and pastimes category was "I am cutting down on some

**Table 5.**

Shoulder Pain and Disability Index<sup>14</sup> and Sickness Impact Profile<sup>5</sup> Change Scores, Standardized Response Means (SRM), and 95% Confidence Intervals for SRM

Instrument	Change Scores		SRM <sup>a</sup>	95% Confidence Interval <sup>b</sup>
	X	SD		
Shoulder Pain and Disability Index				
Pain dimension	-37.4	24.3	1.54	0.87,2.26
Disability dimension	-28.4	27.2	1.04	0.64,1.49
Total	-33.0	23.9	1.38	0.83,1.92
Sickness Impact Profile				
Physical dimension	-0.3	1.8	0.15	-0.46,0.62
Psychosocial dimension	-2.6	5.6	0.47	-0.18,1.01
Sleep and rest	-4.8	20.5	0.24	-0.70,0.87
Eating	0.0	2.0	0.00	-0.40,0.40
Work	-2.9	21.6	0.13	-0.22,0.51
Home management	-9.9	12.5	0.79	0.44,1.15
Recreation and pastimes	-10.5	9.6	1.10	0.62,1.67
Total	-2.6	3.3	0.79	0.32,1.21

<sup>a</sup>SRM=standardized response mean for change scores of subjects who improved clinically (n=34).

<sup>b</sup>Confidence interval of SRM values calculated with jackknife estimates of variance.

of my usual physical recreation or activities." Both of these statements have obvious applicability to individuals who are diagnosed with shoulder pathology.

The correlations in our study are similar to those reported by Williams et al<sup>16</sup> in their investigation of the construct validity of the SPADI. The researchers reported Pearson correlations ranging from  $-.27$  (health perceptions) to  $-.50$  (physical functioning) between the SPADI and the SF-20, a previously validated generic HRQL scale. The similarities between the correlations found in our study and those reported by Williams et al<sup>16</sup> support the meaningfulness of our findings. The correlations for most SIP scores hypothesized in our study to be related to SPADI scores provide reasonably strong evidence to support the construct validity for the SPADI. In addition, most of the SIP scores hypothesized to be either weakly related or unrelated to SPADI scores provide additional evidence to support the construct validity for the SPADI.

#### Responsiveness of the SPADI

We used the SRM to measure the responsiveness of both the SIP and the SPADI. Other investigators<sup>29,36</sup> have used Cohen's benchmarks to qualify the responsiveness of disability and health status measures. According to Cohen's benchmarks, the SRM values for the SIP shown in Table 5 were small or moderate. Only the SIP's recreation and pastimes category had a large SRM (1.10), representing a large degree of responsiveness. In contrast, all SPADI scores were highly responsive accord-



**Table 6.**  
Percentage of Subjects Responding to Frequently Selected Sickness  
Impact Profile<sup>5</sup> (SIP) Items<sup>a</sup>

SIP Dimension and Item	Subjects Responding (%)
Home management <i>I am doing less of the regular daily work around the house than I would usually do.</i>	49
Home management <i>I am not doing heavy work around the house.</i>	49
Emotional behavior <i>I keep rubbing or holding areas of my body that hurt or are uncomfortable.</i>	45
Sleep and rest <i>I sleep less at night, for example, wake up too early, do not fall asleep for a long time, awaken frequently.</i>	42
Recreation and pastimes <i>I do my hobbies and recreation for shorter periods of time.</i>	28
Emotional behavior <i>I often moan and groan in pain or discomfort.</i>	19

<sup>a</sup> All subjects who completed data forms during initial visit (n=94).

ing to Cohen's benchmarks. Standardized response means were generally higher for the SPADI than for the SIP, ranging from 1.04 to 1.54 (Tab. 5). Tests of differences indicated that the SPADI is more responsive than the SIP when used for patients with shoulder pathology. Williams et al<sup>16</sup> also found the SPADI to be a highly responsive scale.

One factor that may have adversely affected our ability to compare the responsiveness of SIP and SPADI scores was the skewness of the SIP data. With the exception of the recreation and pastimes score of the SIP, however, the SPADI scores are clearly more responsive than even those SIP scores that were not as severely skewed.

In our study, we defined meaningful changes as being determined by agreement between the patient's and therapist's judgments as to whether the patient's function had improved, stayed the same, or worsened. Results of the judgments revealed only two patients who did not show meaningful improvements in their condition. The two patients who did not show improvement were reported to have stayed the same. There was consensus between therapist and patient judgments on all patients. Because of the heavily skewed results toward patients who demonstrated meaningful improvement in function, SRM values were determined only for those patients who demonstrated meaningful improvement.

The high percentage of patients who experienced meaningful improvement may have been due to differences in

the extent of disability between those patients who completed the study and those patients who were lost to follow-up. To determine whether a difference in disability existed between patients who completed the study (n=36) and patients who completed forms only during the initial visit (n=58), a *t* test was performed. No differences were found between the SIP or SPADI scores of those patients who completed the study and the scores of those patients who did not complete the study. Therefore, the extent of disability present in patients who completed the study did not appear to be different from that of patients who did not complete the discharge forms. Factors other than disability, however, still may have created a selection bias with the decrease in patient sample size between initial and final visits.

#### *Limitations Associated With Use of the SPADI*

When analyzing the data collected in our study, a few limitations of the SPADI were noted. One of the problems encountered was the way that patients interpreted the instructions for completion of the SPADI. Three patients used words to mark the visual analog line rather than placing a mark on the line. The questionnaires completed in this manner were not scorable according to the developer's instructions. The written instructions provided with the SPADI could therefore be clarified, which might increase the number of forms that are correctly completed. For example, a sample question and response could be added to the instructions to indicate to the respondent how to correctly mark the 10-cm line.

Data for another group of patients (n=6) who scored "NA" on one or both dimensions of the SPADI also were not included in the data analysis. A dimension was scored "NA" if more than two items were left blank or marked "NA" by the patient. We observed that all of the patients who scored "NA" had just undergone rotator cuff surgery or shoulder arthroplasty. The patients apparently scored "NA" because they were instructed by their surgeon not to actively move their shoulder at the time they filled out their questionnaires during the initial visit. The wording of the SPADI limits the applicability of the SPADI to only those patients with shoulder pathology who are instructed to perform active movements with their involved shoulder.

Roach et al<sup>14</sup> and Williams et al<sup>16</sup> presented evidence to support the reliability and validity of the SPADI for measuring HRQL in individuals with shoulder pathology. All subjects, however, in the study by Roach et al<sup>14</sup> were men, and 98% of the subjects studied by Williams et al<sup>16</sup> were men. Examination of the questions comprising the SPADI reveals that at least one of the questions is biased toward male patients. The gender-biased question is "How much difficulty do you have removing some-

thing from your back pocket?" Because many men carry items, such as a wallet, in their back pants pocket, and women generally do not, the question is biased toward men.

In conclusion, the SPADI may be more useful with revised instructions that would clarify for the respondent the correct procedure for completing the questionnaire. The SPADI does not appear to be applicable to patients who are instructed not to move their involved shoulder because of their condition. In addition, because of the apparent bias in one of the SPADI items, the SPADI may be more applicable to male patients than to female patients.

## Conclusions

Evidence for the construct validity of the SPADI is moderately strong, based on the patterns of correlations with the SIP. The correlations between the SPADI scores and the work and recreation and pastimes scores of the SIP, however, suggest that the SPADI may not readily measure occupational and recreational disability. An extremely skewed distribution of SIP scores appears to indicate that relatively few SIP items are applicable to individuals with shoulder problems.

The SRM values for the SIP and SPADI indicate that the SPADI is more responsive to change than the SIP. The superior responsiveness of the SPADI supports the notion that region-specific scales are more responsive than generic health status questionnaires. Because the SIP has limited applicability to patients with shoulder problems and because the SPADI is more responsive to change, the SPADI would be preferred over the SIP for measuring the extent of disability in individuals with shoulder problems. Consideration should be given to the possible effects of the large dropout rate on our conclusions related to responsiveness. Two issues that warrant further study of the SPADI are the patient instructions and the potential influences of gender bias.

## Acknowledgments

Special thanks to the following clinical facilities that participated in the data collection phase of this study: Rehabilitation Services, Martha Jefferson Hospital, Charlottesville, Va; Spectrum Physical Therapy, Charlottesville, Va; Island Sports Physical Therapy, East Meadows and East Northport, NY; Professional Therapies of Roanoke, Vinton, Va; and Chippenham Sports Medicine Center, Richmond, Va.

## References

- 1 Mathews J. *Practice Issues in Physical Therapy: Current Patterns and Future Directions*. Thorofare, NJ: Slack Inc; 1989.
- 2 Thier SO. Forces motivating the use of health status assessment measures in clinical settings and related clinical research. *Med Care*. 1992;30:ms15–ms22.

- 3 Stewart DL, Abeln SH. *Documenting Functional Outcomes in Physical Therapy*. St Louis, Mo: Mosby; 1993:213.
- 4 Jette AM. Using health-related quality of life measures in physical therapy outcomes research. *Phys Ther*. 1993;73:528–537.
- 5 Bergner M, Bobbitt R, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care*. 1981;19:787–805.
- 6 Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis: the Arthritis Impact Measurement Scale. *Arthritis Rheum*. 1980;23:146–152.
- 7 Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30:473–481.
- 8 Jette AM, Davies AR, Cleary PD, et al. The Functional Status Questionnaire: reliability and validity when used in primary care. *J Gen Intern Med*. 1986;1:143–149.
- 9 Fairbank JCT, Couper JC, Davies JB, et al. The Oswestry Low Back Pain Questionnaire. *Physiotherapy*. 1980;66:271–273.
- 10 Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop*. 1981;155:7–20.
- 11 Barrett WP, Franklin JL, Jackins SE, et al. Total shoulder arthroplasty. *J Bone Joint Surg Am*. 1987;69:865–880.
- 12 Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop*. 1987;214:160–164.
- 13 Imatani RJ, Hanlon JJ, Cady GW. Acute complete acromioclavicular separation. *J Bone Joint Surg Am*. 1975;57:328–332.
- 14 Roach KE, Budiman-mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care and Research*. 1991;4:143–149.
- 15 Viikari-Juntura E, Takala EP, Alaranta H. Neck and shoulder pain and disability: evaluation by repetitive gripping test. *Scand J Rehabil Med*. 1988;20:167–173.
- 16 Williams JW, Holleman DR, Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. *J Rheumatol*. 1995;22:727–732.
- 17 Helmstadter GC. *Principles of Psychological Measurement*. New York, NY: Appleton-Century-Crofts; 1964.
- 18 Roland M, Morris R. A study of the natural history of back pain: development of a reliable and sensitive measure of disability in low back pain. *Spine*. 1983;8:141–144.
- 19 Deyo RA. Comparative validity of the Sickness Impact Profile and shorter scales for functional assessment in low-back pain. *Spine*. 1986;11:951–954.
- 20 Laupacis A, Muirhead N, Keown P, Wong C. A disease-specific questionnaire for assessing quality of life in patients on hemodialysis. *Nephron*. 1992;60:302–306.
- 21 Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol*. 1993;46:153–162.
- 22 Weinberger M, Samsa GP, Tierney WM, et al. Generic versus disease-specific health status measures: comparing the Sickness Impact Profile and the Arthritis Impact Measurement Scale. *J Rheumatol*. 1992;19:543–546.
- 23 Bruin AF, Witte LP, Stevens F, Diederiks JP. Sickness Impact Profile: the state of the art of a generic functional status measure. *Soc Sci Med*. 1992;35:1003–1014.

- 24** Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care*. 1989;27:5217-5232.
- 25** Deyo RA, Patrick DL. Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Med Care*. 1989;27:s254-s268.
- 26** Bergner M, Bobbit RA, Pollard WE, et al. The Sickness Impact Profile: validation of a health status measure. *Med Care*. 1976;14:57-67.
- 27** Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res*. 1984;19:275-289.
- 28** Pollard WE, Bobbit RA, Bergner M, et al. The Sickness Impact Profile: reliability of a health status measure. *Med Care*. 1976;14:146-155.
- 29** Katz JN, Larson MG, Phillips CB, et al. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. 1992;30:917-925.
- 30** Beaton DE, Richards RR. Measuring function of the shoulder: a cross-sectional comparison of five questionnaires. *J Bone Joint Surg Am*. 1996;78:882-890.
- 31** Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*. 1986;40:171-178.
- 32** Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis*. 1986;39:897-906.
- 33** Stratford PW, Binkley J, Solomon P, et al. Assessing change over time in patients with low back pain. *Phys Ther*. 1980;74:528-533.
- 34** Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38:27-36.
- 35** MacKenzie CR, Charlson ME, Gigioia D, Kelley K. Can the Sickness Impact Profile measure change? an example of scale assessment. *J Chronic Dis*. 1986;39:429-438.
- 36** Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28:632-642.
- 37** Cohen J. *Statistical Power Analysis and the Behavioral Sciences*. New York, NY: Academic Press Inc; 1977.
- 38** Kopec JA, Esdaile JM, Abrahamawicz M, et al. The Quebec Back Pain Disability Scale: measurement properties. *Spine*. 1995;20:341-352.