# A Critical Review of Questionnaires for Assessing Pain-Related Disability

Richard W. Millard[1,2]

*Standardized questionnaires offer a practical and cost-effective means for measuring moderate disability. Fourteen questionnaires that can be used for assessing pain-related disability are reviewed in this article. Their comparative attributes are summarized. Most of the questionnaires obtain information about specific activities of daily living. They vary in terms of structure, content, and intended applications. Psychometric theory provides the optimal method for evaluating these questionnaires, highlighting the importance of reliability and validity. Being mindful of psychometric qualities will help the evaluator to select an appropriate questionnaire. The present array of existing instruments points toward a need for comparison studies that may eventually result in more uniform methods for evaluating pain-related disability.*

**KEY WORDS:** pain; disability; psychological assessment.

## INTRODUCTION

Accurate disability evaluations are difficult to perform when symptoms of pain are present. Pain is an inherently subjective experience. As a result, objective findings from the physical examination may show poor correspondence to level of functioning at activities of daily living, work, or social integration. The need for improved methods of evaluating pain-related disability has been recognized by the Social Security Administration, which processes a large proportion of claims citing pain as a major reason for disability (1). Self-report questionnaires are a potential way to obtain meaningful estimates of such disability or functional loss.

In using any questionnaire, there is the threat that findings may be biased by self-report. As an alternative, biomechanical approaches are available to collect seemingly more objective data that is observable and quantifiable. This may be col-

[1]Departments of Psychiatry and Anesthesiology, University of Rochester Pain Treatment Center, Rochester, New York.
[2]Correspondence should be directed to Richard W. Millard, PhD, Departments of Psychiatry and Anesthesiology, University of Rochester Pain Treatment Center, 300 Crittenden Blvd., Rochester, New York 14642.

lected using methods that range from standardized tests of lumbar function (2) to direct observation of pain behaviors (3,4). However, the performance demands of a clinical setting may also interfere with this kind of measurement. Administration of a questionnaire may require fewer resources. It may also yield more practical information about conventional activities in the natural environment (5).

The appeal of disability questionnaires is apparent as many self-report measures have proliferated within the last decade. The evaluator is confronted with a bewildering choice of contents, formats, and intended applications. Some instruments only assess disability or functional loss. Others inquire about pain intensity, cognitive influences, physical symptoms, or other topics as well. Psychometric theory provides the best way to evaluate and select a questionnaire. It is founded upon a rich tradition of measuring physical and mental function, pioneered in the last century by Galton and Cattell (6).

The purpose of this review is to acquaint the reader with comparative advantages or limitations of 14 readily available questionnaires that have been employed among groups reporting pain. The emphasis is on practical measurements of moderate disability. Scales restricted to vocational demands, pain behavior, or biomechanical measurements (e.g., trunk strength, aerobic capacity) were not considered. These are organized in three groups according to the intended population: (1) back pain, (2) pain without reference to site, or (3) illness, without reference to pain. Each questionnaire is discussed in terms of background, psychometric findings, and implementation. Psychometric findings are discussed in terms of reliability and validity. Reliability, the consistency or dependability of results, has been almost universally reported in research describing these questionnaires. Validity has been less completely evaluated. In some cases, it has been explored but not labeled as such. Where appropriate, this review will make reference to evidence concerning types of validity (content, construct, criterion). Table I presents a summary of the domain, format, and intended population of each questionnaire.

## DISABILITY OF BACK PAIN

Some questionnaires are restricted to assessing disability associated with back pain. This reflects the relative prevalence of back pain as well as the fact that it tends to interfere with multiple activities and be associated with global disability. Certain actions (e.g., bending, lifting) are especially affected by back pain, and for this reason these questionnaires tend to possess somewhat distinctive content. Additionally, three of the five questionnaires that are reviewed in this section are parts of broader schemes to integrate information about both disability and impairment within a comprehensive evaluation framework.

### Chronic Disability Index (CDI)

*Derivation.* The CDI is a short (9-item) yes/no checklist that comprises one part of a combined method for assessing disability and impairment. It contains in-

Table I. Summary of Questionnaires

| Questionnaire | Content | Format |
|---|---|---|
| **Back pain** | | |
| Chronic Disability Index | Basic activities of daily living | Clinician or patient checklist |
| Disability Questionnaire | Disrupted activities (from SIP) | Patient checklist |
| Low-Back Rating Scale | Physical measurment, patient and clinician perceptions | Clinician and patient ratings |
| Million Scale | Subjective reactions to back pain | Analogue rating by patient |
| Oswestry Scale | Activity limitations | Percentage rating by patient |
| **Pain (non-specific)** | | |
| Chronic Illness Problem Inventory | Physical, psychosocial, illness behavior, marital | Patient rating |
| Functional Rating Scale | Behavioral changes | Clinician or patient rating |
| Multidimensional Paint Inventory, Part IIII | "Uptime" or activity | Patient rating |
| Pain Disability Index | Areas of activity | Analogue rating by patient |
| **Illness (pain or other)** | | |
| Activity Pattern Indicators | Frequency of activity | Patient report |
| Functional Assessment Screening Questionnaire | Difficulty of activity | Patient rating |
| Health Assessment Questionnaire, Disability Index | Difficulty of activity | Patient rating |
| Self-Care Assessment Schedule | Frequency of activities | Patient report |
| Sickness Impact Profile | Disrupted activities, illness behaviors | Patient checklist |

quiries about nine general activities (e.g., walking, sleeping, putting on footwear, etc.) that had been identified as common areas of difficulty due to back pain (7). Some items are borrowed from a previous questionnaire and others are rationally-derived. All questions refer to direct behavioral signs of disability. The CDI assesses disability within an integrative model that links impairment, pain intensity, and psychological distress (7).

*Psychometrics.* Adequate inter-rater reliability was established in interviews of 30 patients $(0.90 > r > 0.73)$ (7). Satisfactory inter-correlation among the constituent items also suggested that the CDI assesses disability as a single construct. Up to 480 patients were included in a comparison of results from the CDI and its associated impairment index. These findings indicated that information about physical impairment was able to explain 46% of variance in scores on the CDI (7). This further supports its construct validity in being able to provide information that is distinguishable from findings that would emerge from conventional physical examinations. CDI findings have also overlapped, but been distinguishable from reports of psychological distress (7).

*Implementation.* The CDI is intended for use among patients with back pain only. It is a brief and practical scale that can be easily repeated. It is relatively unique in being embedded within a larger model for explaining the nature of low-back disorders. Its companion impairment scale for assessing physical impairment

is equally practical and provides a coherent way to include such information within a comprehensive evaluation.

## Disability Questionnaire (DQ)

*Background.* The DQ contains 24 yes/no items from the Sickness Impact Profile (SIP) (8), selected for relevance to back pain (9). Psychosocial content has been mostly eliminated; representative statements include "I get dressed more slowly" and "I find it difficult to get out of a chair" (10).

*Psychometrics.* Over a 3-week interval, DQ scores showed comparable consistency to full-scale SIP scores ($r = 0.83$ among patients with stable functional status) (10). Scores were highly correlated ($r = .85$) to the SIP, suggesting that brevity resulted in the loss of little important information (10).

DQ findings were found to be independent of distress, age, or sex (10). Moderate relationships ($r = 0.42$) were reported to pain intensity and measurement of spine flexion. In comparison to the SIP, scores were equally sensitive to clinical changes among patients with acute pain (10).

*Implementation.* This questionnaire has been studied with acute (10) and chronic (9) patients. Millard and Jones (11) have reported its application for chronic low-back pain, showing moderate agreement between DQ scores and other questionnaire measures. If the full SIP provides a "shotgun" approach for assessing many areas of disability, the DQ is more precise. It avoids unnecessary items, is quickly administered, and maintains good psychometric qualities.

## Low Back Rating Scale (LBRS)

*Background.* The LBRS (12) is a multiple-method approach including measures of functional capacity (trunk strength, range of motion), patient report, and physician rating. Patient perception includes estimated pain intensity as well as reported difficulty at 15 activities of daily living. Physician perception is based upon expected pain or disability and utilization of analgesic medication. Overall, this format was thought to approximate comparable scales for assessing the impact of hip dysfunction (12).

*Psychometrics.* Reliability of the total scale is not reported; separate estimates might be necessary for each method of measurement. When administered to 29 surgical patients and 48 rehabilitation inpatients, scores were found to be sensitive to treatment improvements (12). On the basis of initial findings, information about functional capacity and patient perception is weighted when calculating the total score (12). Certain Minnesota Multiphasic Personality Inventory Scale scores (Hypochondriasis and Hysteria) were predictive of poor treatment response, as measured by the LBRS (12).

*Implementation.* The LBRS has not been widely applied. It includes some information that is less directly related to disability, such as pain intensity and medication usage. However, its combined assessment of patient self-report, functional

capacity, and clinician rating is potentially appealing. Few scales incorporate multiple methods, a format that may potentially enhance validity.

## Million Scale (MS)

*Background.* Million and colleagues (13) selected questions about disability and pain intensity to develop a 15-item checklist for measuring progress among patients with back pain. It is specifically intended to elicit subjective perceptions. A visual analogue rating method is used to record responses. As an adjunct, ten "objective" criteria from physical examination (e.g., straight-leg raising, lumbar extension/flexion) are evaluated in a companion scale.

*Psychometrics.* Very high test–retest reliability was reported for the subjective questionnaire items ($r = 0.97$), yet this was based upon a small sample with only a few hours between administrations (13). High inter-rater reliability is also reported ($r = 0.92$, $n = 36$). Same-day assessments of the ten objective items produced extremely high correlations (averaging $r = 0.99$, $n = 19$) (13).

The sensitivity of the two combined scales was evaluated by comparing 19 patients who received a lumbar corset, either with or without a plastic support. Subjective items were more sensitive to treatment response in these two groups. The objective biomechanical variables did not change. These findings were consistent across three 1-month intervals (13).

*Implementation.* Like the LBRS, the MS yields a combination of patient report and directly-observed biomechanical information. As noted, this multi-method approach is potentially useful. Unfortunately, construct validity of the MS has not been evaluated in comparison to other measures of disability. It may provide more information about pain beliefs than about areas of disabled functioning. Its content is specific to back pain and this limits its generalizability to other pain conditions or chronic illnesses.

## Oswestry Questionnaire (OQ)

*Background.* This brief scale provides a percentage score to reflect level of functioning among individuals receiving physical therapy (14). Items refer to activities of daily living that might be disrupted by low back pain (e.g., lifting, standing, sexual activity, sleeping), rated for difficulty on a 6-point scale. Pain intensity is also measured as one of the ten items.

*Psychometrics.* High test–retest reliability ($r = 0.99$) has been found when comparing responses by 22 patients who completed the scale twice over a 24-hour interval (14). Items were found to be closely related to one another, suggesting good internal consistency (14).

The validity of this measure was initially supported in two ways. First, pain intensity was directly related to reports of disability (14) among 22 patients. Second, recovery from back pain was monitored over a 3-week period for 25 patients. It was found that the percentage of disability declined by an average of about 28%

during the recovery period. This suggested that the scale is sensitive to changes in functioning.

*Implementation.* The OQ was initially developed for acute low back pain, yet it has been applied to evaluate chronic back pain as well. It has been rather highly correlated ($r = 0.70$) to the CDI (7). Other research has shown close correspondence between OQ scores and the degree of relaxation in back muscles during flexion (15). Hazzard and colleagues (16) used the scale to document clinical outcomes in an occupational rehabilitation program and found that it distinguished employed from unemployed patients at 1 year following treatment. This is one of the few instances where the utility of a disability questionnaire has been independently replicated and also related to treatment outcomes.

# DISABILITY OF PAIN (WITHOUT REFERENCE TO SITE)

Pain can occur in areas other than the back, and some questionnaires are appropriate for evaluating disability that accompanies any type of pain complaint. These typically emphasize measurement of disruptions in performing general activities of daily living (ADLs). Although developed for populations with pain, the content of some of these questionnaires may make little or no reference to pain or physical symptoms. In such cases, minimal modification would be necessary to extend their usage to other populations.

## Chronic Illness Problem Inventory (CIPI)

*Derivation.* This questionnaire was developed to provide a problem-oriented record for chronic illnesses (17). The format is modeled after a comparable record for use among patients with cancer. Although intended for various chronic illness groups, the original sample consisted of 115 respondents at a pain treatment facility. Items are grouped into 18 categories reflecting diverse areas of functioning (e.g., finances, appearance, etc.). Responses are given on a 5-point scale.

*Psychometrics.* High internal consistency (mean alpha for the 18 categories = 0.85) and test–retest reliability over a 1-week interval ($r = 0.87$) were reported (17). CIPI findings were generally in agreement with clinical ratings by a psychologist (17). Comparison to other chronic illness groups (38 obese patients and 15 chronic obstructive pulmonary disease patients) indicated the poorest functioning among patients with pain complaints. The ability to distinguish between these groups was construed as evidence of the measure's validity.

*Implementation.* The CIPI appears to have been implemented in few studies. Like the Sickness Impact Profile or Health Assessment Questionnaire, it provides a broad-based estimate of disability: physical limitations, psychosocial functioning, health care behavior, and marital adjustment are all sampled. Evaluators who are primarily interested in self-reported physical limitations may be deterred by the inclusion of these additional topics.

## Functional Rating Scale (FRS)

*Background.* The Functional Rating Scale (18) obtains interval estimates of functioning in terms of work, activities of daily living, time out of bed, and the usage of medications, transcutaneous electrical nerve stimulation (TENS), or other apparatus. The brief format makes it easy for the clinician to complete this scale during an initial interview.

*Psychometrics.* A cross-section of 127 healthy and disabled respondents completed the FRS on two occasions, with about a 3-week interval between administrations. This indicated high test–retest reliability ($r = 0.96$ for the composite group) (18).

Content validity of the FRS was supported by showing that scores were significantly higher among healthy ($n = 98$), as opposed to disabled individuals ($n = 157$) (18). Patients with chronic pain and arthritis patients were both evaluated, although the scale did not differentiate between these groups. Significant changes in FRS scores have been documented among 58 patients with chronic pain who were assessed before and after an individualized pain management program that included physical therapy and behavior modification. Among this group, FRS scores changed an average 40% over a 7-month period (18).

*Implementation.* This scale has not been widely used. It has few items, and some may be infrequently endorsed (i.e., use of TENS). It may be useful when a clinician-rated scale is being sought. When applied in this manner, scores from the FRS have been moderately correlated ($0.44 < r < 0.54$) with patient-report questionnaires for evaluating low-back disability (11).

## Multidimensional Pain Inventory (MPI, Part III)

*Background.* The 18 items within this questionnaire are intended to assess disruptions in activities of daily living (19). They are divided into four content areas: household chores, outdoor work, activities away from home, and social activities. Each item is answered using a 6-point scale.

*Psychometrics.* Good internal consistency ($0.70 < r < 0.86$) and test–retest ($0.83 < r < 0.91$) reliability coefficients have been reported (19) within the content areas. However, there is greater variability between the types of activities ($0.12 < r < 0.49$) (19).

Within the derivation sample of 120 patients in two Veterans Administration hospitals, MPI Part III responses were moderately correlated to perceptions of control. Activity level was also distinguished from general distress, solicitousness by others, and pain severity. Each of these were identified as additional factors within the full MPI (19).

*Implementation.* As part of a larger inventory, these questions provide the opportunity to evaluate disability in relation to both overall distress and pain variables. Although Part III is not designed to be used separately from the full MPI, such an application is conceivable. Findings can be useful in selecting treatment, as items

consist of activities that might be set as goals within behavioral therapy. Like the CDI, the MPI is linked to a parallel system for measuring impairment (20).

### Pain Disability Index (PDI)

*Background*. The Pain Disability Index (PDI) (21) was designed to be a brief and global analogue rating of function across seven content areas: family/home responsibilities, recreation, social activity, occupation, sexual behavior, self-care, and life-support activity.

*Psychometrics*. An alpha coefficient of 0.86 has been reported among a sample of 401 respondents (21), reflecting high internal consistency. The seven items of the PDI have been found to comprise two factors that reflect either obligatory activities (i.e., life support and self-care) or discretionary activities (e.g., occupation, sexual behavior). Good internal consistency has been demonstrated within these factors (0.70 for the obligatory activities and 0.85 for the discretionary activities (21). Test–retest reliability has been reported for 46 patients who completed the PDI at a 2-month interval while awaiting inpatient pain treatment. Given that changes in disability were not anticipated, this revealed relatively poor consistency ($r = 0.44$) (21).

The validity of the PDI has been supported in various ways, including comparisons of high- vs. low-disabled groups, patient vs. nurse ratings, and age/sex differences. Regression analysis has shown that variance in PDI scores can be accounted for by time spent in bed, symptom reporting, and other logically associated variables. These findings are also offered as evidence of the scale's validity (21).

*Implementation*. With only seven items, the PDI is easily understood and can be completed within a few minutes. It is among the briefest measures in this review. Further interviewing would be necessary to obtain information about specific activities. Its 11-point analogue scaling is comparable to common methods for assessing pain intensity, and this may produce a greater overlap between reports of pain and disability (11). As the name implies, this scale is worded to evaluate individuals who are disabled by pain, although subsequent research might demonstrate that it can be easily modified for use among other moderate disabilities.

### DISABILITY OF ILLNESS (WITHOUT REFERENCE TO PAIN)

The remaining category of questionnaires is the most general. These questionnaires have been used with multiple populations, typically at settings for general rehabilitation. Their content remains descriptive, although there is no major emphasis upon pain. They offer the advantage of being able to compare and communicate results across differing populations, potentially equating moderate levels of disability from such diverse conditions as back pain, head injury, or multiple sclerosis.

## Activity Pattern Indicators (API)

*Background.* As an outgrowth of the collaborative Rehabilitation Indicators project during the 1970's, the API was developed to signify functional loss associated with various disorders, including pain (22). The 64 "indicators" or items are grouped into 11 categories that represent areas of functioning, such as Homemaking and Socializing. Patients indicate how often these activities are performed.

*Psychometrics.* Reliability was examined by evaluating agreement at how different raters grouped the items into categories. This yielded a Kappa coefficient of 0.94 (22), reflecting high agreement. Test–retest reliability coefficients (6-week interval) have been reported for activity categories and subscales, ranging between $0.91 > r > 0.48$ (22).

The internal structure of the API has been found to consist of obligatory vs. discretionary types of activities (22). This structure has been subsequently identified in other measures of pain-related disability (21, 23). Responses differed between sexes, but this was not indicated to be a source of bias (22). The relationship between API findings and concurrent signs of disability (such as employment or time spent in bed) is not reported.

*Implementation.* The API was initially developed for research applications. It evaluates pattern of disability instead of providing a global score, and it may respond less well to the practical demands of clinical settings. Further investigation of this measure could extend its appeal, because the indicators represent a wide and potentially useful number of activities that are likely to be monitored during a rehabilitation program.

## Functional Assessment Screening Questionnaire (FASQ)

*Background.* The FASQ was initially developed to assess functioning among moderately-disabled medical populations although its value in evaluating pain-related disability has been demonstrated in a sample of 158 patients with chronic pain (23). It obtains an ordinal rating of how much difficulty is encountered when performing 15 signal activities (e.g., "doing grocery shopping," "cutting your toenails"). These activities are rated for level of difficulty by using a 5-point scale for each item.

*Psychometrics.* Split-half estimates ($r = 0.84$, $n = 158$) and inter-rater responses ($r = 0.71$, between 29 patients and their spouses) have indicated adequate internal consistency (23).

FASQ scores have differentiated individuals on the basis of employment status and type of pain complaint (23). Findings have been minimally related to psychological variables as indicated by Minnesota Multiphasic Personality Inventory scales, except for those containing numerous somatic items (23, 11). As with the API and PDI, it seems to possess a general factor structure that distinguishes obligatory from discretionary activities. In comparison to similar disability questionnaires, FASQ results seem to be minimally associated to negative affectivity or coping style (11).

*Implementation*. The FASQ is a practical alternative to longer measures and provides the ability to distinguish between levels of disability. It also offers the potential advantage of permitting communication across different rehabilitation populations, because the content is not limited to pain. Further investigation is needed to evaluate whether it can be sensitive to treatment-related changes.

## Health Assessment Questionnaire—Disability Index (HAQ-DI)

*Background*. The HAQ-DI was developed as a comprehensive outcome measure within a multi-center rheumatologic database. Its format and content are appropriate for frequent, repeated administrations. In spite of this rather specialized background, it is not a disease-specific measure, and it has been modified for use with other populations (24). Eight types of activities of daily living are represented. Scores can be expressed as "disability units" and used to compare relative levels of functioning over time or across groups.

*Psychometrics*. This scale has benefitted from much investigation. The initial database included 24 settings and over 7000 respondents. As a result, estimates of good internal consistency and inter-rater reliability have been reported (24). Logical relationships have been demonstrated to chronicity, type of illness, distress, and disease variables (e.g., morning stiffness, number of affected joints) (24, 25). High intercorrelation ($r = 0.91$) has also been shown to physical disability (as measured by the Arthritis Impact Measurement Scale) among a sample of 48 patients with rheumatoid arthritis (24).

*Implementation*. Information about use of the HAQ-DI is mostly restricted to rheumatologic samples although it has proven useful in distinguishing disability from pain and distress in the diffuse symptom picture of fibromyalgia ($n = 83$) (25). It has been the subject of considerable investigation and yet it is a relatively brief measure that can be completed in about 5 minutes. This makes it potentially appropriate when repeated administrations are desired. Its "disability units" are an appealing concept that offers the opportunity to make meaningful comparisons of functioning across time and differing pain conditions.

## Self-Care Assessment Schedule (SCAS)

*Derivation*. The SCAS is a short (10 items) checklist that seeks to identify purely behavioral disruptions characterizing disability (26). Respondents indicate the frequency with which key activities (e.g., arise from bed prior to 10 a.m., shop, prepare a meal) occurred over an immediately prior 14-day time frame. It was designed to exclude potentially confounding symptoms of illness or estimates of functional capacity.

*Psychometrics*. Initial studies were conducted upon a mixed population of 583 subjects, comprising groups of psychiatric in- and outpatients, patients with chronic pain, medical/surgical patients, and inhabitants of the community. Cumulative results have generally provided ample psychometric support (26, 27) for the integrity of the SCAS. Group differences emerged in anticipated directions, with the inpa-

tient psychiatric and pain samples exhibiting the greatest behavioral disruption (27). Test–retest reliability estimates for psychiatric patients have been reported over a 3-day interval ($r = 0.82$, $n = 27$) (27). Cronbach alpha coefficients for individual items ranged from 0.70 to 0.75 across multiple clinical groups ($126 > n > 14$), with good agreement across raters ($0.84 > r > 0.78$, patients and friends/relatives). In addition, the activities being assessed occur naturally and with enough frequency to assure an adequate range of scores (27).

Construct validity was specifically demonstrated by showing logical relationships to measures of social adjustment and psychiatric symptoms, and by indicating a logical factor structure among psychiatric outpatients ($n = 151$) (27). SCAS scores have also been closely associated to independent clinical assessments of overall deterioration or improvement ($r = 0.82$, $n = 24$ patients with chronic pain) (27).

*Implementation.* The SCAS is intentionally focused upon a limited number of carefully selected, specific behaviors. In this way, it differs greatly from global scales such as the MS or PDI. The emphasis upon signal behaviors is comparable to the FASQ, although the SCAS inquires less about vocational or recreational demands. The combination of its brief format and large body of psychometric support would make it a potentially suitable scale for either clinical or research purposes. A further advantage is that its validity has been evaluated with various clinical and nonclinical populations (27).

## Sickness Impact Profile

*Background.* This checklist was developed to be a behaviorally-based measure of health status across a range of populations (8). The 136 yes/no items reflect 12 different categories (e.g., Home Management, Mobility, Alertness Behavior). These are grouped into three dimensions (Physical, Psychosocial, and Other) that correspond to broad kinds of disability. Individual items carry differing weights in calculating the total score, which can be expressed as a percentage for the category or a total for the dimension. Combining scores from the three dimensions provides a total "impairment" score that is intended to portray the degree of behavioral impact associated with sickness.

*Psychometrics.* High reliability estimates (up to $r = 0.94$) have been reported in large samples across many medical populations. This has been reported in terms of internal consistency, test–retest and inter-rater agreements, and other methods (8).

There are extensive findings supporting validity of the SIP. Large ($N = 1108$) field studies were initially conducted with health maintenance organization enrollers in addition to smaller groups of medical patients (with hyperthyroidism, rheumatoid arthritis, and hip replacements). SIP findings were logically associated with multiple measurements of health status (8). The SIP has been used to evaluate pain-related disability in a number of studies; in one study of 107 patients with chronic low back pain, scores on the physical dimension were significantly correlated with activity level while psychosocial scores were significantly correlated with measures of distress (28). In another study, the total SIP score was rather highly associated with

depression ($r = 0.60$, Beck Depression Inventory), suggesting the potential influence of negative affectivity upon SIP results (29). There are conflicting reports as to whether the SIP is sensitive to the kinds of clinical changes that occur during rehabilitation of moderately disabled clients. Follick, Smith, and Ahern (28) reported that SIP scores were significantly changed after an outpatient pain rehabilitation program ($n = 14$). In a larger study, however, Turner and colleagues (30) found that the SIP was relatively less sensitive to post-treatment changes.

*Implementation.* The SIP has been subjected to a much higher level of empirical scrutiny than most available questionnaire methods and is not limited to use with pain-related disability. Its application of the term "impairment" suggests that more than functional loss is being evaluated. The length of the SIP may be impractical as it contains many items that refer to considerably more disabling conditions and would be infrequently endorsed. The DQ was devised, in part, to address this potential drawback.

## SUMMARY

With this many questionnaires, the evaluator ought not have difficulty in finding one that is appropriate. Practical considerations may direct the choice of a questionnaire, with attention to cost, time, or the resources that are expended for completion and scoring. This is a specialized form of assessment that does not supplant the need for accompanying medical and psychological data. One may wish to employ alternate assessment methods (such as direct behavioral observation or mechanical assessment of physical function) as well as self-report and consider how findings can be interpreted in relation to other components of an overall evaluation. There is little reason to develop new self-report questionnaires because the existing instruments sample a wide universe of content and can be adapted to virtually all conceivable settings with little or no modification. Rather than deriving more questionnaires, it would be wise for subsequent investigations to consider their application. The following recommendations may guide this work.

First, there are compelling reasons to use psychometric procedures when evaluating methods for assessing functional loss or disability. This offers a common language that can be used to contrast the assorted questionnaires and communicate findings. The goal is to obtain sound and useful measurements. Psychometric theory evolved so that there would be a way to understand and describe how to obtain sound, useful measurements. It is far better to use the terms of reliability and validity than to coin new expressions that may not be as widely employed.

While it is relatively easy to come by evidence of reliability, systematic investigations of validity are more scarce. The domain of information that is sampled by these questionnaires is surprisingly varied and content validity has not been well explored. Criterion validity has sometimes been inferred by the presence of treatment-related changes in questionnaire scores. However, such findings have rarely been replicated and may not reflect clinically important information. Construct validity can be jeopardized by negative affectivity, bias of self-report, and demand characteristics of the test setting. Such variables may collude to render distorted

measurements. Psychometric theory provides a framework for investigating these important questions about the quality of disability questionnaires.

Second, there is a need to compare different questionnaires and different ways of measuring disrupted activity. This too can be phrased in psychometric language, as a question of concurrent validity. There have been few instances where alternate methods have been directly compared. Without such studies, however, it is unclear how much a common phenomenon is being assessed by these various approaches. It is difficult to state whether one method may be superior to another. There are also problems at communicating across settings, because it is hard to equate levels of disability when different questionnaires or assessment methods are used.

A third recommendation is to improve methods for communicating results. The ideal goal would be a uniform assessment of disability. This might yield a standard index, such as disability units from the HAQ (24). It would require more comparative investigations, seeking to equate findings from different approaches. In the end, it would make it possible to meaningfully equate participants in rehabilitation programs across multiple sites. Such a step would go far in helping to unify the present disarray of questionnaires.

No definitive recommendation about the choice of an individual questionnaire is offered in this review. The advent of specialized occupational rehabilitation approaches has shown the importance of measuring disability with pain. The proliferation of questionnaires (and other measurement approaches) is not a new concern. It stems from the venerable question of how behavior and bodily health are related. Questionnaires simply offer one easy and economical way to obtain this information. The evaluator is advised to be cautious about how and what they measure, because findings and interpretations will affect decisions about individual management of pain-related disability.

## REFERENCES

1. Osterweis M, Kleinman A, Mechanic D. *Pain and disability: Clinical, behavioral, and public policy perspectives.* Washington: National Academy Press, 1987, p. 268.
2. Mayer TG, Tencer AF, Kristoferson MS, Mooney V. Use of noninvasive techniques for quantification of spinal range-of-motion in normal subjects and chronic low-back dysfunction subjects. *Spine* 1984; 5: 88-95.
3. Keefe FJ, Wilkins RH, Cook WA. Direct observation of pain behavior in low back pain patients during physical examination. *Pain* 1984; 20: 59-68.
4. Follick, MJ, Ahern DK, Aberger EW. Development of an audiovisual taxonomy of pain behavior: Reliability and discriminant validity. *Health Psychol* 1985; 4: 555-568.
5. Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehab* 1988; 69: 1044-1053.
6. Anastasi, A. *Psychological testing* (3rd Ed.). New York: Macmillan, 1968, p. 7.
8. Waddell G, Main CJ. Assessment of severity in low-back disorders. Spine 1984; 9: 204-208.
8. 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: 787-805.
9. Roland M, Morris R. A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983; 8: 141-144.
10. Deyo RA. Comparative validity of Sickness Impact Profile and shorter scales for functional assessment in low-back pain. *Spine* 1986; 11: 951-954.
11. Millard RW, Jones RE. Construct validity of practical questionnaires for assessing disability of low-back pain. *Spine* 1991; 16: 835-838.

12. Lehmann TR, Brand RA, Gorman TW. A low-back rating scale. *Spine* 1983; 8: 308-315.
13. Million R, Hall W, Nilsen KH, Baker RD, Jayson MI. Assessment of the progress of the back pain patient. *Spine* 1982; 7: 204-212.
14. Fairbank JC, Couper, J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980; 66: 271-273.
15. Triano JJ, Schultz AB. Correlation of objective measure trunk motion and muscle function with low-back disability ratings. *Spine* 1987; 12: 561-565.
16. Hazzard RG, Fenwick, JW, Kalisch SM, Redmond J, Reeves V, Reid S, Frymoyer JW. Functional restoration with behavioral support. A one-year prospective study of patients with chronic low-back pain. *Spine* 1989; 14: 157-161.
17. Kames LD, Naliboff BD, Heinrich RL, Schag CC. The Chronic Illness Problem Inventory: Problem-oriented psychosocial assessment of patients with chronic illness. *Int J Psychiat Med* 1984; 14: 65-75.
18. Evans JH, Kagan A. Development of a functional rating scale to measure treatment outcome of chronic spinal patients. *Spine* 1986; 11: 277-281.
19. Kerns RD, Turk DC, Rudy TE. The West Haven–Yale multidimensional pain inventory. *Pain* 1985; 23: 345-356.
20. Turk DC, Rudy TE, Stieg RL. The disability determination dilemma: Toward a multiaxial solution. *Pain* 1988; 34: 217-230.
21. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: Psychometric properties. *Pain* 1990; 40: 171-182.
22. Rock DL, Fordyce WE, Brockway JA, Bergman JJ, Spengler DM. Measuring functional impairment associated with pain: Psychometric analysis of an exploratory scoring protocol for activity pattern indicators. *Arch Phys Med Rehab* 1984; 65: 295-300.
23. Millard RW. The Functional Assessment Screening Questionnaire: Application for evaluating pain-related disability. *Arch Phys Med Rehab* 1989; 70: 303-307.
24. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: The Health Assessment Questionnaire, Disability, and Pain Scales. *J Rheumatol* 1982; 9: 789-793.
25. Hawley DJ, Wolfe F, Cathey MA. Pain, functional disability, and psychological status: A 12-Month study of severity in fibromyalgia. *J Rheumatol* 1988; 15: 1551-1556.
26. Barnes, D. Benjamin S. The Self-Care Assessment Schedule (SCAS)-I. The purpose and construction of a new assessment of self care behaviours. *J Psychosomat Res* 1987; 31: 191-202.
27. Benjamin S, Barnes D. The Self-Care Assessment Schedule (SCAS)-II. Reliability and validity. *J Psychosomat Res* 1987; 31: 203-214.
28. Follick MJ, Smith TW, Ahern DK. The Sickness Impact Profile: A global measure of disability in chronic low back pain. *Pain* 1985; 21: 67-76.
29. Watt-Watson JH, Graydon JE. Sickness Impact Profile: A measure of dysfunction with chronic pain patients. *J Pain Sympt Mnmt* 1989: 4: 152-156.
30. Turner JA, Clancy S, McQuade KL, Cardenas DD. Effectiveness of behavioral therapy for chronic low back pain: A component analysis. *J Consult Clin Psychol* 1990; 58: 573-579.