SPINE Volume 28, Number 10, pp 1061–1067

Comparison of the Roland–Morris Disability Questionnaire and Generic Health Status Measures

A Population-Based Study of Workers' Compensation Back Injury Claimants

Judith A. Turner, PhD,* Deborah Fulton-Kehoe, MPH,† Gary Franklin, MD, MPH,†‡ Thomas M. Wickizer, PhD,§ and Rae Wu, MD, MPH†

Study Design. Workers' compensation back injury claimants completed baseline and follow-up telephone interviews in a prospective population-based cohort study.

Objective. To compare the Roland–Morris Disability Questionnaire (RDQ) to widely used generic health status measures in a sample of workers with recent work-related back injuries in terms of validity, reliability, responsiveness to change, and floor and ceiling effects.

Summary of Background Data. Little research has directly compared the validity and responsiveness of the RDQ to that of the Short-Form 12 or Short-Form 36 health status measures among individuals with back pain. Furthermore, there is little information concerning the validity, reliability, and responsiveness of the RDQ as a measure of functional outcomes for workers with back injuries.

Methods. Approximately 8 weeks (median) after filing low back injury claims, 309 workers completed the RDQ, Short-Form 12, and Short-Form 36 scales and gave information about their work status in computer-assisted telephone interviews. An average of 5 months later, 284 workers (91.9%) completed the measures again.

Results. The RDQ demonstrated excellent internal consistency and validity through correlations with other measures of physical functioning, ability to discriminate between those working and those not working, and much more responsiveness to change than the Short-Form 12 and Short-Form 36 scales. However, 15% of the sample did not answer one or more RDQ items.

Conclusions. The RDQ is a valid measure of physical disability among workers with back injuries. Its greater responsiveness to change suggests its superiority to the Short-Form 12 and Short-Form 36 as an outcome measure in this population. [Key words: low back questionnaires, Roland–Morris Disability Questionnaire, SF-12, SF-36, workers compensation] **Spine 2003;28:1061–1067**

Multiple instruments are available for assessing health outcomes for patients with low back pain. These include both back pain–specific measures and generic questionnaires that assess functional and psychosocial variables influenced by overall health. Generic instruments can assess effects of multiple health conditions and enable comparisons of patients with different diseases as well as comparisons of patients with healthy individuals. However, condition-specific measures have the potential to be more responsive than generic measures to clinically important changes in the condition under study. Many experts have argued for including both generic and condition-specific measures in clinical research.[1,2] However, in some situations, only a limited number of questions may be asked of study participants, and a choice must be made.

The Roland–Morris Disability Questionnaire (RDQ)[3] is widely used to assess physical disability associated with low back pain. The RDQ has been demonstrated as valid, reliable, and responsive to change.[3–8] The Short Form-12 (SF-12)[9] and the Short Form-36 (SF-36)[10] are widely used, valid, reliable generic health status measures. However, little research has directly compared the RDQ to the SF-12 or SF-36. In a postal survey study of a community sample with low back pain, a modified RDQ was found to be associated moderately ($r = -0.61$) with the SF-36 Bodily Pain scale and highly ($r = -0.78$) with the SF-36 Physical Functioning scale, but less highly with the other SF-36 scales.[8] In a prospective cohort study of patients, one third of whom were receiving workers' compensation, who sought care from orthopedic and neurologic surgeons for sciatica, a modified RDQ and the SF-36 Bodily Pain and Physical Functioning scales showed good responsiveness to change. The modified RDQ was somewhat more responsive than the SF-36 Physical Functioning scale.[2]

From the *Department of Psychiatry and Behavioral Sciences and the Department of Rehabilitation Medicine, University of Washington School of Medicine, Seattle, the †Occupational Epidemiology and Health Outcomes Program, Department of Environmental Health, University of Washington School of Public Health and Community Medicine, Seattle, the ‡Washington State Department of Labor and Industries, Olympia, and the §Department of Health Services, University of Washington School of Public Health and Community Medicine, Seattle, Washington.

Supported by the Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health grant no. 1 R01 OH04069, and the Medical Aid Fund of the Washington State Department of Labor and Industries.

Acknowledgment date: July 23, 2002. First revision date: October 4, 2002.

Acceptance date: October 7, 2002.

Device status/drug statement: The submitted manuscript does not contain information about medical devices or drugs.

Conflict of interest: Federal funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this article.

Address correspondence and reprint requests to Judith A. Turner, PhD, Department of Psychiatry and Behavioral Sciences, Box 356560, University of Washington, Seattle, WA 98195; E-mail: jturner@u.washington.edu

Despite the considerable volume of research on the RDQ, there is a paucity of information from samples of injured workers. The RDQ may be useful for monitoring outcomes of workers with back injuries (*e.g.,* to target workers who do not improve for more intensive treatment) or for research involving this population, especially because it is relatively brief and can be administered in either paper-and-pencil or telephone interview formats. However, it is important to validate a low back pain measure with the population in which it will be used,[8] and the RDQ's psychometric properties have not been established among workers with back injuries. The purpose of this study was to compare the RDQ with widely used generic health status measures in a population-based sample of workers with recent back injury claims in terms of reliability, validity, responsiveness to change, and floor and ceiling effects.

■ Materials and Methods

Participants and Procedure. The Washington State Department of Labor and Industries (DLI) claims database was used to identify workers 18 years of age or older with a work-related back injury claim newly accepted between July 3 and September 22, 2000. Back injuries were identified by the American National Standards Institute (ANSI) codes for the body part injured (*e.g.,* back and injury nature = sprain/strain). In Washington state, approximately 70% of nonfederal workers are insured by the DLI state workers' compensation fund, and 30% by the 382 large self-insured firms. A random sampling strategy was used to obtain a sample with approximately equal numbers of state fund nontime loss (<4 days of injury-related lost work time and thus eligibility for medical, but not time loss, compensation), state fund time loss (≥4 days of lost work time and thus eligibility for time loss compensation), and self-insured time loss back injury claims (information on type of injury available only for workers with time loss claims in self-insured companies).

The sampling strategy identified 710 back injury claims, and attempts were made to contact the claimants by telephone between August 29, 2000 and January 3, 2001. The contacted workers were screened by telephone for study eligibility. The exclusion criteria specified inability to participate because of language, hearing, or other barriers and worker report of no claim for a work-related back injury. Eligible workers were invited to participate in the study, which involved a baseline and a follow-up telephone interview. The workers were not compensated for the baseline interview, but were paid $10 for the follow-up interview. Among these 710 workers, 217 (30.6%) could not be contacted, 45 (6.3%) were ineligible, and 139 (19.6%) declined participation.

Among the 309 workers (46.5% response rate, excluding the ineligible workers) who enrolled and completed a baseline telephone interview, the mean age was 40.5 ± 11.0 years (range, 18–74 years), 194 (62.8%) were male, and 115 (37.2%) were female. In terms of ethnicity and race, 238 (77%) were white, 14 (4.5%) African American, 13 (4.2%) Native American, 13 (4.2%) bi- or multiracial, 12 (3.9%) Hispanic or Mexican American, 11 (3.5%) Asian or Pacific Islander, and 5 (1.6%) other or unknown. Three (1%) declined to give racial or ethnic information. In terms of education, 22 (7.1%) had not graduated from high school; 114 (36.9%) had a high school or general education development diploma; 18 (5.8%) had completed vocational, technical, or trade school; 120 (38.8%) had attended some college; 23 (7.4%) had completed college; and 12 (3.9%) had attended professional or graduate school. The median time from DLI claim receipt to the baseline interview was 8 weeks (range, 3–24 weeks). The study participants did not differ significantly from nonparticipants in terms of male proportion, time loss *versus* nontime loss claims, or self-insured *versus* state fund claims, as analyzed by χ^2 tests. However, the participants were significantly older than the nonparticipants (mean age, 40.5 ± 11.0 *vs* 37.2 ± 11.1 years; $t = 3.95$; $P = 0.000$).

A total of 284 (91.9%) study participants completed a follow-up interview a mean of 5 months after the baseline interview (range, 2.3–8 months). Five claimants (1.6%) declined follow-up participation, and 20 (6.5%) could not be contacted. Follow-up respondents did not differ significantly from nonrespondents in baseline scores on any measure (according to *t* tests), gender, education, or time loss *versus* nontime loss claim status (according to χ^2 tests). However, follow-up nonrespondents were significantly younger (mean age, 36.2 ± 12.7 *vs* 40.9 ± 10.8 years; $t = 2.06$; $P < 0.05$) and more likely to be nonwhite (40% *vs* 20%; $\chi^2 = 5.17$; $P < 0.05$).

Measures. The participants answered sociodemographic and work status questions and completed health status measures in a computer-assisted telephone interview (CATI) conducted by trained interviewers. The SF-36 Physical Functioning, Role–Physical, Bodily Pain, and Mental Health scales were administered because they capture important dimensions of back injury–related quality of life and have been shown to have good psychometric properties.[2,10] The participants also completed the additional items needed to score the SF-12,[9] the 12-item version of the SF-36. The SF-12 yields Physical (PCS) and Mental Component (MCS) summary scores. On the SF-12 and SF-36, higher scores indicate better health status. The SF-36 scores range from 0 to 100. The SF-12 PCS and MCS are scored using norm-based methods and transformed to have means of 50 and standard deviations of 10 in the general U.S. population. Finally, the participants completed the RDQ,[3] which is scored from 0 to 24, with higher scores indicating more severe disability. All of these scales and the work status questions were administered in both the baseline and follow-up interviews. At the follow-up interview, the workers also were asked about change in their injury-related pain.

Statistical Analyses. Internal consistency was assessed by Cronbach's alpha.[11] Coefficients exceeding 0.70 generally are regarded as acceptable, those exceeding 0.80 as good, and those exceeding 0.90 as excellent.[12] The authors also calculated the proportion of questionnaires that could not be scored because of unanswered questions, the correlations of the measures at baseline, and the proportion of the sample with the lowest and highest possible scores for each measure. A scale is said to have ceiling effects when individuals who score as perfectly healthy on the scale are identified as still having health problems. Because these people are at the top of the scale, further improvements in health cannot be detected. Similarly, if a patient scores at the floor of a scale, it will not be possible to detect further deterioration.

To examine the validity of each measure, the scores of study participants who were working were compared with those of participants who were not working, with the assumption that

Table 1. Baseline Health Status Assessment of Back Injury Claimants: Missing Data, Floor and Ceiling Effects, and Cronbach's Alpha

Measure	Missing (%)	Floor (%)	Ceiling (%)	Cronbach's Alpha*
RDQ	14.6	0	0	0.95
SF-12				
PCS	9.7	0	0	—
MCS	9.7	0	0	—
SF-36				
Physical functioning	0	1.3	17.5	0.95
Role–physical	0	38.5	28.2	0.88
Bodily pain	0	2.3	12.3	0.87
Mental Health	0	0	3.9	0.85

* Cronbach's alpha can not be calculated for the SF-12 summary scales.
RDQ = Roland–Morris disability questionnaire; SF = short form; PCS = physical component summary scale; MCS = mental component summary scale; Floor = worst possible score; ceiling = best possible score.

those not working should show worse health status. The study used *t* tests to determine the statistical significance of the differences between groups and effect sizes (the difference between the group means divided by the standard deviation) for the purpose of assessing the extent of the differences.

Several analyses were performed to assess the responsiveness of each measure to change from baseline to follow-up assessment. First, paired *t* tests were used to determine the statistical significance of the change. Next, both the effect size (the mean change in scores divided by the standard deviation of individuals' baseline scores) and the standardized response mean (SRM) (the mean change in scores divided by the standard deviation of individuals' changes in scores) were calculated. For both statistics, absolute values 0.20 or more but less than 0.50 generally are considered small, values 0.50 or more but less than 0.80 are considered moderate, and values 0.80 or more are considered large.[12] Finally, the effect size for subgroups categorized according to workers' report of change in their condition at follow-up assessment was calculated. Responsiveness to change was demonstrated by greater effect sizes for groups reporting the most change.

■ Results

Neither floor nor ceiling effects were seen at baseline for the RDQ, SF-12, or SF-36 Mental Health scales (Table 1). However, the best possible score was obtained by 18% of the sample on the SF-36 Physical Functioning scale and by 12% on the Bodily Pain scale. Even more problematic was the SF-36 Role-Physical scale, on which 39% achieved the worst possible score and 28% the best possible score. Because some workers declined to answer some questions or gave "don't know" responses, scores could not be calculated for all the study participants on questionnaires that required all or almost all of the items to be answered for them to be scored. This was particularly problematic for the RDQ and the SF-12. The RDQ score is the number of items a respondent indicates as describing him or her that day. With interview-administered RDQs, there is no standardized method for scoring if a respondent does not answer one or more questions. Therefore, RDQs with any unanswered questions were excluded from further analyses. This resulted in a high rate (14.6%) of missing data (Table 1). The missing data rate was also high (9.7%) for the SF-12, which cannot be scored if any item is missing.[13] In contrast, there were no missing data for the SF-36 scales, which can be scored if at least half of the items are answered.[10] Internal consistency was good for all the scales, and excellent for the RDQ and SF-36 Physical Functioning scale (Table 1).

The validity of the RDQ was supported by its high correlations with the other physical functioning measures (SF-12 PCS, SF-36 Physical Functioning and Role-Physical) and the SF-36 Bodily Pain scale (absolute values of r, 0.70–0.85; Table 2). The validity of all the measures also was demonstrated by significantly (P = 0.000) worse scores in the subgroup that had not worked the month before the baseline interview (Table 3). The effect size reflecting the degree of difference between the working and nonworking groups was large for all the scales except the SF-12 MCS, and largest for the RDQ and the SF-36 Physical Functioning scale.

Table 2. Correlations of Health Status Scales, Baseline Assessment of Back Injury Claimants

	RDQ	SF-12 PCS	SF-12 MCS	SF-36 PF	SF-36 RP	SF-36 BP
SF-12 PCS	−0.80					
SF-12 MCS	−0.52	0.33				
SF-36 PF	−0.85	0.86	0.43			
SF-36 RP	−0.70	0.85	0.47	0.69		
SF-36 BP	−0.74	0.85	0.43	0.76	0.72	
SF-36 MH	−0.60	0.48	0.87	0.55	0.54	0.53

RDQ = Roland–Morris disability questionnaire; SF = short form; PCS = physical component summary scale; MCS = mental component summary scale; PF = physical functioning; RP = role–physical; BP = bodily pain; MH = mental health.

Table 3. Back Injury Claimants Who Did Versus Did Not Work in the Prior Month: Baseline Health Status Measure Scores and Effect Size of the Difference

	Worked, Past Month*				
	Yes (n = 234)		No (n = 75)		
Measure	M	(SD)	M	(SD)	ES
RDQ	8.09	(5.90)	15.09	(6.34)	−1.17
SF-12					
PCS	43.21	(10.68)	32.52	(9.13)	1.03
MCS	50.81	(9.42)	43.07	(11.73)	0.77
SF-36					
Physical functioning	73.46	(26.74)	41.46	(30.57)	1.15
Role–physical	53.49	(41.05)	12.89	(29.52)	1.05
Bodily pain	57.59	(25.72)	34.84	(23.47)	0.90
Mental health	74.67	(19.25)	58.31	(23.32)	0.81

* Those who did versus those who did not work in the past month differed significantly ($P = 0.000$) on each measure.
RDQ = Roland–Morris disability questionnaire; SF = short form; PCS = physical component summary scale; MCS = mental component summary scale; M = mean; SD = standard deviation; ES = effect size.

Table 4 shows the baseline and follow-up scores for each measure and three indicators of the measures' responsiveness to change in workers' health status: statistical significance (paired *t*-tests), effect size, and SRM. Statistically significant changes were found for all the scales except the SF-12 MCS. However, these changes were modest in size. The effect sizes were small for all the measures, and the SRM was in the moderate range only for the RDQ. This suggests that improvement, on the average, was modest for this sample of injured workers.

To examine the measures' responsiveness to change in the subgroup of workers who improved from baseline to follow-up assessment, the SRMs were calculated separately for the group reporting at follow-up assessment that they had improved, and for the group not working at baseline but working at follow-up assessment (Table 5). In both of these groups, a comparison of the SRM absolute values (higher scores on the RDQ and lower scores on the SF-12 and SF-36 indicate greater disability) shows the RDQ to be much more responsive than the other measures. Table 5 also shows the SRMs for three other subgroups based on the report of change at follow-up assessment as no change, worse, and variable. Among the small group of workers who said they had worsened, the SRM was weak for all the measures.

■ Discussion

Little is known about the relative merits evidenced by the most widely used health status measures (RDQ, SF-36, and SF-12) in various back pain populations. The current study is the first to compare these measures directly in a population-based sample of workers with back injury claims. The findings support the RDQ's convergent validity (as demonstrated by high correlations with other measures of disability), internal consistency, and responsiveness to change in this population. Support for the validity of all the scales studied comes from the finding that all scores were significantly worse among those who had not worked in the preceding month. The RDQ and the SF-36 Physical Functioning scale showed superiority to the other measures in terms of internal consistency and ability to discriminate between those working and those not working. The RDQ was much more responsive to change than the generic measures.

In the small group reporting at follow-up that they had worsened, no measure showed good sensitivity in

Table 4. Baseline to Follow-up Change in Health Status Among Back Injury Claimants

	Baseline (n = 309)		Follow-up (n = 284)					
Measure	M	(SD)	M	(SD)	*t*	*P*	ES	SRM
RDQ	9.19	(6.66)	6.68	(6.74)	8.74	.000	−0.37	−0.58
SF-12								
PCS	40.74	(11.37)	42.30	(11.64)	−2.81	.005	0.14	0.18
MCS	49.14	(10.47)	50.11	(11.02)	−1.49	.14	0.09	0.09
SF-36								
Physical Functioning	65.63	(31.01)	68.69	(29.20)	−2.19	.03	0.10	0.13
Role–physical	44.13	(42.44)	52.73	(43.21)	−3.87	.000	0.20	0.23
Bodily pain	52.51	(26.87)	56.59	(25.24)	−2.97	.003	0.15	0.18
Mental health	71.18	(21.22)	73.48	(21.24)	−2.03	.04	0.11	0.12

RDQ = Roland–Morris disability questionnaire; SF = short form; PCS = physical component summary scale; MCS = mental component summary scale; M = mean; SD = standard deviation; ES = effect size; SRM = standardized response mean.

Table 5. Back Injury Claimants Grouped by Change at Follow-up in Pain* and Work Status† SRMs

Measure	Change in Pain				Follow-up Work Status	
	Improved (n = 150)	Variable (n = 97)	No Change (n = 19)	Worse (n = 18)	Working (n = 36)	Not Working (n = 31)
RDQ	−0.78	−0.48	−0.65	0.10	−0.84	−0.22
SF-12						
PCS	0.32	0.09	−0.32	−0.20	0.33	0.13
MCS	0.12	0.11	0.05	−0.27	0.54	−0.02
SF-36						
Physical functioning	0.23	0.10	−0.38	−0.16	0.50	0.01
Role–physical	0.36	0.16	−0.12	−0.19	0.56	0.01
Bodily pain	0.24	0.25	−0.44	−0.20	0.57	0.14
Mental health	0.16	0.16	−0.04	−0.26	0.63	−0.10

* Total sample.
† Among those not working at baseline.
RDQ = Roland–Morris disability questionnaire; SF = short form; PCS = physical component summary scale; MCS = mental component summary scale; SRM = standardized response mean.

detecting this decline, and the RDQ showed no sensitivity. Although the authors are reluctant to draw conclusions given the small number (n = 18) of workers who said they had worsened and that their worsening possibly was modest, it is possible that the RDQ is better at detecting improvement than deterioration. A literature search yielded no published articles that directly addressed this issue. However, one other study reported a much smaller increase (1–2 points) on the RDQ among patients with sciatica who reported that their leg pain was a little or much worse than the decrease (2.2–9.6 points) on the RDQ among patients who reported that their leg pain was a little or much better.[2] Another study found that an extensively modified RDQ was much better at detecting improvement than at detecting deterioration in work status among primary care patients with back pain.[15] Further research is needed to determine whether the RDQ in fact is relatively insensitive to worsening in back pain–related disability.

It has been suggested that the RDQ may show floor effects for severely disabled patients (*i.e.,* it may not discriminate among patients with varying levels of severe disability), and that it may be most appropriate in settings wherein patients have mild to moderate disability.[7] This sample of workers with back sprain injuries was not severely disabled, on the average, and no floor effects were seen.

At baseline, neither floor nor ceiling effects were seen for the RDQ, SF-12, or SF-36 Mental Health scales. However, the SF-36 Physical Functioning and Bodily Pain scales showed possible ceiling effects (*i.e.,* many workers achieved the best possible score). Even more problematic was the SF-36 Role-Physical scale, which showed substantial floor and ceiling effects. Version 2 of the SF-36 is now available, and revisions to the Role-Physical scale appear to result in decreased floor and ceiling effects and greater responsiveness.[16,17]

Approximately 15% of the RDQs were not scored because workers gave a "don't know" or "prefer not to answer" response to one or more items. Prior survey studies have reported even higher rates of incomplete data for this measure.[8] In the paper-and-pencil format of the RDQ,[7] patients are asked to check only the items that describe them that day. Therefore, it is unknown whether items not checked truly do not apply or whether the patient inadvertently skipped the item or did not want to answer it. The current authors recommend that consideration be given to modifying the RDQ with instructions to patients that they must answer "yes" or "no" to each item, as is typical for self-report measures, and to developing a standardized method for scoring the RDQ when some questions are unanswered. This would ensure greater comparability of results across studies and administration modes: telephone, mail, in-person interview, and in-person questionnaire.

The average physical health status among this sample of workers with back injury claims, as assessed by the SF-12 PCS and the SF-36 Physical Functioning and Bodily Pain scales, was lower than the 25th percentile of the U.S. general population norms at both baseline and follow-up assessment.[10,13] For the sample, improvement from baseline to follow-up assessment, on the average, was only modest. It should be kept in mind that in this sample, time loss compensation claims were overrepresented, which would be associated with greater physical disability. Nonetheless, the poor health status is notable, especially given the fact that baseline interviews were conducted approximately 2 months after claims were filed and follow-up interviews were conducted an average of 5 months after baseline, when recovery from most back sprain injuries would be expected. The extent to which the poor physical health status of these workers preceded or followed the injury cannot be determined from this study.

The study findings should be considered in light of several issues. First, only 46.5% of potentially eligible claimants enrolled in the study. However, support for the generalizability of the results to the population of workers with back injury claims comes from the similar-

ity between participants and nonparticipants in gender distribution and proportion with time loss claims. Although the participants were significantly older, the difference in mean age was small (40.5 *vs* 37.2 years).

Second, follow-up respondents were significantly older and had a greater representation of whites than nonrespondents, raising the possibility that the measures' responsiveness to change might have been different if all study participants had completed the follow-up assessment. However, this possibility appears unlikely given the very high (92%) follow-up response rate and the small number of nonrespondents (n = 25). Furthermore, follow-up respondents and nonrespondents did not differ on baseline scores on any measure, gender, education, or time loss *versus* nontime loss claim status.

Third, the measures were administered in telephone interviews, and it is possible that administration in a paper-and-pencil format might yield different results. The authors chose telephone administration because in-person assessment was not possible (study participants lived throughout the state of Washington) and telephone administration offers several advantages over mailed questionnaires. These advantages include certainty about who is actually completing the measures, lower missing data rates,[18–20] probably higher response rates,[20] and ease of data entry, with responses entered directly into a database during computer-assisted telephone interviews. There have been conflicting findings concerning the comparability of data obtained through telephone interviews and data from paper-and-pencil questionnaires, but studies that have compared measures completed in both formats by the same patients have found high rates of agreement.[21,22]

In summary, the authors recommend the RDQ for use in research involving workers with back injuries. Although the RDQ, SF-12, and SF-36 all are easy to administer in both telephone and paper-and-pencil formats, the RDQ can be scored more easily and quickly. Furthermore, the RDQ proved much more responsive to change than the SF-12 and SF-36 scales in this sample, consistent with other studies that have found condition-specific instruments to be more responsive than generic measures.[14,23] Differences in responsiveness have important implications for clinical research and practice. Information concerning a measure's responsiveness is important in selecting a measure for a particular study, allowing estimation of sample sizes needed to ensure adequate statistical power (the larger the effect size or SRM, the smaller the sample size required).[14]

The RDQ does not assess psychosocial dysfunction or pain intensity. The RDQs authors recommend using other scales to assess these variables.[7] This recommendation is supported by the current findings of only moderate associations between the RDQ and the generic mental health and pain scales. In addition, investigators who wish to compare their back pain samples with samples representing other health conditions may wish to supplement the RDQ with a general health status measure.

■ Key Points

- The Roland–Morris Disability Questionnaire (RDQ) is a valid, reliable, and responsive measure of physical disability in individuals with work-related back injury claims.
- The RDQ is more responsive to change than the SF-12 and SF-36 in this population, but use of the SF-12 and SF-36 as supplementary measures enables the collection of information concerning health status indexes other than physical disability (*e.g.,* mental health), as well as comparisons with other health conditions.
- Workers with back injury claims show substantially below average health status and quality of life relative to the general U.S. population.

Acknowledgments

The authors thank Jerry Gluck and Terri Smith-Weller for their assistance with this study.

References

1. Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;25:3100–3.
2. Patrick DL, Deyo RA, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899–909.
3. Roland M, Morris R. A study of the natural history of back pain: Part 1. Development of a reliable and sensitive measure of disability in low back pain. *Spine* 1983;8:141–4.
4. Beurskens AJHM, de Vet HCW, Koke AJA. Responsiveness of functional status in low back pain: a comparison of different instruments. *Pain* 1996; 65:71–6.
5. Deyo RA. Comparative validity of the Sickness Impact Profile and shorter scales for functional assessment in low back pain. *Spine* 1986;11:951–4.
6. Jensen MP, Strom SE, Turner JA, et al. Validity of the Sickness Impact Profile Roland Scale as a measure of dysfunction in chronic pain patients. *Pain* 1992;50:157–62.
7. Roland M, Fairbank J. The Roland–Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine* 2000;25:3115–24.
8. Underwood MR, Barnett AG, Vickers MR. Evaluation of two time-specific back pain outcome measures. *Spine* 1999;24:1104–12.
9. Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
10. Ware JE, Snow KK, Kosinski M. SF-36 *Health Survey: Manual and Interpretation Guide.* Lincoln, RI: QualityMetric, 2000.
11. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297–334.
12. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis, and Interpretation.* Chichester, England: John Wiley, 2000.
13. Ware JE, Kosinski M, Keller SD. SF-12: *How to Score the SF-12 Physical and Mental Health Summary Scales.* 3rd ed. Lincoln, RI: QualityMetric, 1998.
14. Stucki G, Liang MH, Fossel AH, et al. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995;48:1369–78.
15. Dionne CE, Von Korff M, Koepsell TD, et al. A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. *Spine* 1999;24:2339–45.
16. Jenkinson C, Stewart-Brown S, Peterson S, et al. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health* 1999;53:46–50.
17. Ware JE, Kosinski M, Dewey JE. *How to Score Version Two of the SF-36 Health Survey.* Lincoln, RI: QualityMetric, 2000.
18. McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 Health survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994;32:551–67.
19. O'Toole B, Battistutta D, Long A, et al. Comparisons of costs and data

quality of three health survey methods: mail, telephone, and personal home interview. *Am J Epidemiol* 1986;124:317–28.
20. Perkins JJ, Sanson-Fisher RW. An examination of self- and telephone-administered modes of administration for the Australian SF-36. *J Clin Epidemiol* 1998;51:969–73.
21. Morishita L, Boult C, Ebbitt B, et al. Concurrent validity of administering the Geriatric Depression Scale and the Physical Functioning Dimension of the SIP by telephone. *J Am Geriatr Soc* 1995;43:680–3.
22. Simon GE, Revicki D, Von Korff M. Telephone assessment of depression severity. *J Psychiatr Res* 1993;27:247–52.
23. Garratt AM, Moffett JK, Farrin AJ. Responsiveness of generic and specific measures of health outcome in low back pain. *Spine* 2001;26:71–7.

Point of View

Jon D. Lurie, MD

Currently, a wide variety of health status questionnaires are available to clinicians and researchers. Head-to-head comparisons of differences in health status measures, such as this study by Turner and colleagues, are critical for clinicians and researchers trying to understand which measures to use in different settings. This work adds to the evidence showing the reasonable consistency and validity of the Roland–Morris Disability Questionnaire (RDQ), the Short-Form 36 (SF-36), and the Short-Form 12 (SF-12) in patients with low back problems. The relative responsiveness of these scales is the more interesting and the more complicated question. Because no "gold standard" for better functional health exists, our evaluation of the responsiveness of health status measures will continue to be plagued with the question, "responsive to what?"

The data in Table 5 of this study show that although the RDQ showed more improvement in the group overall at follow-up assessment (responsiveness to "expected" change), and more improvement in the group who rated themselves as improved at follow-up assessment (so-called responsiveness to estimated change),[1] it also showed almost as much improvement in the group who rated themselves as experiencing "no change." In sharp contrast, the physical scales of both the SF-12 and the SF-36 demonstrated lower functional health status scores in the patients whose rating was "no change." The number of patients in the "no change" group was small, and we do not have sufficient information to know what the "right" answer is. However, these data certainly raise the question as to whether part of the RDQ's responsiveness in this study represents spurious improvement or improvement in aspects of functional health that are not important to patients when they consider whether they are better or not.

Another key aspect of this study for readers to keep in mind is that the population studied had very low levels of disability. Whereas the average baseline score on the physical function scale of the SF-36 in this population was 65, the average score in the general population is 85,[2] the average score for patients seen in the National Spine Network was 42,[2] and the average score for workers' compensation patients in the Maine Lumbar Spine Study was 29.[3] This very low level of disability is likely responsible for the large ceiling effects of the SF-36 seen in this study. Whether the findings of this study would be similar for the more severely impaired patients usually seen by spine practitioners requires further study.

References

1. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine* 2000;25:3192–9.
2. Fanuele JC, Birkmeyer NJO, Abdu WA, et al. The impact of spinal problems on the health status of patients: have we underestimated the effect? *Spine* 2000;25:1509–14.
3. Patrick DL, Deyo RD, Atlas SJ, et al. Assessing health-related quality of life in patients with sciatica. *Spine* 1995;20:1899–909.