Patient Health Questionnaire-9 in Spinal Cord Injury: An Examination of Factor Structure as Related to Gender

Claire Z. Kalpakjian, PhD¹; Loren L. Toussaint, PhD²; Kathie J. Albright, PhD¹; Charles H. Bombardier, PhD³; James K. Krause, PhD⁴; Denise G. Tate, PhD¹

¹Model SCI System, Department of Physical Medicine and Rehabilitation, University of Michigan, Ann Arbor, Michigan; ²Luther College; ³Department of Physical Medicine and Rehabilitation, University of Washington School of Medicine, Seattle, Washington; ⁴College of Health Profession, Medical University of South Carolina, Charleston, South Carolina

Received November 15, 2007; accepted July 30, 2008

Abstract

Background/Objective: Despite the attention depression after spinal cord injury (SCI) has received, research and clinical practice have been hampered by inadequate emphasis on reliable and valid measurement. Assessment of symptoms in persons with SCI is challenged by the presence of “trans-diagnostic” symptoms and unexamined effects of gender. The objective of this study was to examine the factor structure of the Patient Health Questionnaire-9 (PHQ-9; the 9-item depression scale of the Patient Health Questionnaire) and determine whether the structure replicates across gender.

Methods: A total of 1,168 women and men were matched on level/completeness of SCI, follow-up year, and age to create 584 pairs. Exploratory factor analysis examined 1- and 2-factor models and congruence in 2 randomly split half samples to establish congruence of the factor solution and replication across gender.

Results: The 1- and 2-factor solutions fit the structure of the items accounting for 41% to 51% of original item variance. Congruence between random samples was uniformly high for the 1-factor solution ($r = 0.791\text{--}0.948$) but variable for the 2-factor solution. Although congruence was high for the combined sample and men ($r = 0.90\text{--}0.97$ and $0.71\text{--}0.94$, respectively), it was variable for women ($r = 0.29\text{--}0.85$).

Conclusions: Although there was support for the 1-factor structure of the PHQ within and between sexes, the low congruence between sexes and within women for the 2-factor structure indicates potentially important differences about how certain symptoms may be experienced or interpreted differently by men and women with SCI. Future research should focus on where sexes diverge in cognitive, affective, and somatic dimensions of depressive symptoms and whether sex-specific or sex-neutral measures are warranted.

J Spinal Cord Med. Apr 2009;32(2):147–156

Key Words: Spinal cord injuries; Affective disorders; Depression; Gender issues; Psychometrics; Factor analysis

Please address correspondence to Claire Z. Kalpakjian, PhD, Model SCI System, University of Michigan, Department of Physical Medicine and Rehabilitation, 300 N. Ingalls, Room NI 2A09, Ann Arbor, MI 48109; p: 734 763 9374; f: 734 936 5492 (e-mail: clairez@umich.edu).

This study was funded by the University of Michigan Model Spinal Cord Injury Care System (H133N060032), Northwest Regional Spinal Cord Injury Systems (H133N060033), and the Georgia Regional Spinal Cord Injury System (HH133N060009) funded by the National Institute of Disability and Rehabilitation Research, Office of Special Education and Rehabilitative Services, US Department of Education.

© 2009 by the American Paraplegia Society

INTRODUCTION

Depression has been studied extensively among people with spinal cord injury (SCI) (1), and its symptoms have been shown to be highly prevalent in this population (2). Depressive symptoms are associated with a myriad of negative outcomes for persons with SCI including lower functional independence (3), more secondary complications (4), poorer community and social integration (5–7), and lower self-appraised health (8). Accurate screening for depression and the proper assessment of severity of depressive symptoms is not only important for clinical settings, but also for research; accurate assessment in outcomes from clinical trials of depression interventions

requires reliable and valid measures that are also sensitive to change.

Despite the attention depression has received, research and clinical practice have been hampered by inadequate emphasis on reliable and valid measurement. In 2 influential reviews, Elliot and Frank (1) and Frank et al (9) observed that studies of depression in people with SCI have been fraught with conceptual imprecision, a diversity of measures, and little attention paid to the broader literature on depression including widely accepted diagnostic criteria. Somatic symptoms have been a key feature of depression in various nosologic systems including the Diagnostic and Statistical Manual of Mental Disorders (DSM) (10) and as far back as the time of Hippocrates (11). Common features of depression include gastrointestinal symptoms, sleep disturbance, headaches, appetite changes, fatigue, and aches and pains of a diffuse nature. Valid measurement of depressive symptoms in persons with SCI is challenged by the presence of “transdiagnostic” symptoms (12) that may be etiologically related to major depression or may be the primary and/or secondary effects of the SCI. As a result, clinicians and researchers may be confused about whether to “count” the somatic symptoms of major depression toward the diagnosis of depression or to ignore them as not etiologically related to major depression. Somatic symptoms also present a challenge to clinicians in primary care settings. The center of that debate focuses on the underrecognition of somatic symptoms as indicative of depression (11). Kroenke (13) argued that the vast majority (70–90%) of patients with depression or anxiety who present to primary care physicians complain of somatic symptoms vs psychological ones, unless they are specifically asked about the latter. This has been supported by a large study conducted by Simon et al (14), who found that, across 1,146 individuals diagnosed with depression from 14 different countries, 45% to 95% presented only with somatic complaints.

Measuring depression is further complicated by gender. In the context of SCI, these issues represent a largely unexplored intersection of 2 areas of investigation that have themselves disregarded differences by sex. First, depression research has focused largely on women, given the higher incidence and prevalence of depression among women (approximately 2:1 M:F ratio) with far less, but growing, interest in understanding unique aspects of depression in men. Second, much of the SCI research in general has focused on men given their overrepresentation in the SCI population (4:1 ratio), with few investigations of sex-related differences in depression prevalence or symptomatology (15).

Although there is wide empiric support for gender disparities in depression rates across populations, the issue is not without debate (16). In recent years, a number of studies have addressed the question of whether there are true sex-related differences in depres-

sion symptomatology or whether artifacts such as different help-seeking behavior, symptom reporting, quality of symptoms, self-ratings, sex-biased diagnosing, gender role socialization, hypo-emotionality, or the inability to describe feelings account for these observed differences (16–18). Of particular relevance to SCI, it has also been proposed that because men often engage in high-risk, antisocial behaviors such as aggression, violence, and alcohol and drug use, that these may represent a different pattern of depression expression (16,18). Research also suggests that significantly more so than men, women can present with symptoms related to weight and appetite change, sleep disturbance, psychomotor retardation, feelings of worthlessness and guilt, gastrointestinal distress, sympathetic arousal, and expressed emotion (19,20). The complex intersection of somatic symptoms and gender and SCI presents a considerable challenge to clinicians and researchers using depression screening and severity measures or diagnostic criteria with this population. One step toward some clarification of this complexity is to examine the underlying structures of depression measures and whether dimensions are identified (eg, cognitive/affective vs somatic) and whether they hold across gender.

Factor analysis of depression scales has been widely used to examine aspects of symptom coverage, the degree to which items represent different symptom clusters or determine the equivalence factor structures across samples (eg, age or sex). It also has been used to conduct comparison studies to determine treatment effectiveness for specific symptoms, the specificity of certain symptoms, or whether subgroups of patients can be distinguished on the basis of specific symptom factors (21). The Patient Health Questionnaire depression scale (PHQ-9) (22) has been used as a depression screening instrument in follow-up interviews conducted by the Model SCI Systems (NIDRR) centers in the past 7 years. The aims of this study were to examine the factor structure of the PHQ-9 and the congruence of dimensions across sex.

METHODS

Study Sample

The study sample was drawn from the Model SCI System's National SCI Statistical Center database (NSCISC), University of Alabama, Birmingham, AL. Participants in this database have sustained traumatic injuries (ie, injuries that occur from external force) and complete follow-up interviews at 1 year after injury and every 5 years thereafter. Data used in these analyses were collected from follow-up interviews conducted from October 2000 through March 2003, with participants having sustained their injuries from 1975 to 2002. Using available data from follow-up years 1, 5, 10, 15, 20, or 25, only cases with complete scores on the PHQ-9 were included for a total pool of $N = 2,741$.

Table 1. Items from the PHQ-9

Over the last 2 weeks, how often have you been bothered by any of the following problems? ^a	
PHQ1	Little interest or pleasure in doing things
PHQ2	Feeling down, depressed, or hopeless
PHQ3	Trouble falling asleep, staying asleep, or sleeping too much
PHQ4	Feeling tired or having little energy
PHQ5	Poor appetite or overeating
PHQ6	Feeling bad about yourself, feeling that you are a failure, or feeling that you have let yourself or your family down
PHQ7	Trouble concentrating on things such as reading the newspaper or watching television
PHQ8	Moving or speaking so slowly that other people could have noticed. Or being so fidgety or restless that you want to hurt yourself in some way
PHQ9	Thinking that you would be better off dead or that you want to hurt yourself in some way

^a Item response options included the following: 0 = not at all; 1 = several days; 2 = more than half the days; 3 = nearly every day.

Variables used to match men and women were selected based on their potential to contribute to depression. Age is a commonly used matching variable, and in this study, we allowed for a 2-year band for matching. We also selected injury characteristics that have been associated with depression in SCI, namely time since injury, which was captured by follow-up year. Level and completeness of injury have been equivocal in their association with depression; however, we chose to be conservative and match on these characteristics as well. From this pool, 1,168 women and men were matched on level (ie, tetraplegia and paraplegia) completeness (ie, complete and incomplete) of SCI, follow-up year, and current age (± 2 years) to create 584 matched pairs. Other demographic characteristics were allowed to vary freely.

Variables and Outcome Measure

Demographic and Injury Characteristics. Basic demographic information such as age, education, race, marital status, employment status, and education are routinely collected during Model SCI System follow-up interviews. Injury level and severity are based on the International Standards for Neurological Classification of SCI, as developed by the American Spinal Injury Association (ASIA) and International Medical Society of Paraplegia (23).

Depression Measure. The PHQ-9 (22) was developed as a diagnostic, self-report tool for identifying depression

in the context of primary care. The items duplicate the criteria for diagnosing depression as adopted by the 4th edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (24). Items from the PHQ-9 are provided in Table 1. In addition to general populations (25), the PHQ-9 also has been used in SCI (2), traumatic brain injury (26), and stroke (27) samples and found to be a valid measure of probable depression in the context of disability. The PHQ-9 also has been shown to detect changes in depression resulting from pharmacologic treatment (28).

Statistical Methods

For this analysis of the PHQ-9 factor structure in an SCI sample, we chose to conduct exploratory (EFA) rather than confirmatory factor analysis (CFA) for several reasons. Most importantly, CFA presumes an established theory. Despite the increasingly widespread use of the PHQ-9, there are very few studies examining its factor structure in the extant literature that provide this theoretic foundation for a factor structure. An exception is Huang et al (29), who found a unidimensional structure of the PHQ-9 across 4 ethnic groups. Second, responses on the items of the PHQ-9 are highly skewed, as would be expected with depression symptoms—most respondents experience very little mood disturbance. Distribution assumptions for CFA are far more stringent than for EFA; although there are several strategies for dealing with nonnormal item distributions, most if not all of these approaches are limited in some way and are not desirable solutions. Therefore, based on clinical insights and some limited literature, we chose to examine a 1-factor model of depression and a 2-factor model that allowed for distinct dimensions of psychologic/emotional vs somatic aspects of depression to emerge.

Principal-axis factor analyses were used in this study. We specified 1 and 2 factors for extraction in these exploratory factor analyses and used oblique (ie, Promax) rotation. Oblique rotation was used because we did not presume that factors would be independent. The Cronbach α was used to assess the internal consistency of the extracted factors. Total sample size was large enough to allow us to assess factor congruence across 2 random samples. Random Sample 1 (RS1) and Random Sample 2 (RS2) each had 292 pairs of men and women each ($N = 584$ each). Congruence was examined between men and women within RS1 and RS2 and between men from RS1 and RS2 and women from RS1 and RS2. The “combined sample” refers to the sample of both men and women within each of the random samples ($N = 584$ each). Figure 1 graphically shows how the total sample was divided and analyses were performed. No significant differences were found within and between groups on demographic and injury matching variables after the random samples were created. The analyses were carried out in the following order:

Figure 1. Graphical representation of random samples and analyses.

- Analysis 1: 1- and 2-factor solutions were examined in RS1 (in the combined sample and then men and women separately).
- Analysis 2: 1- and 2-factor solutions were examined in RS2 (combined sample and then men and women separately).
- Analysis 3: factor solutions were cross-validated between RS1 and RS2 for the combined sample and men and women separately; ie, the similarity or congruence of the factor coefficients was examined.
- Analysis 4: congruence of the factors across men and women within each random sample (ie, gender congruence) was examined in RS1 and RS2.

Congruence of the factor structure across the random samples and between sexes was evaluated using 3 statistical indices; more details about these indices are given in the Appendix. High sample congruence coefficients indicate that the factor loading patterns replicate, and hence, should be considered reliable and generalizable across samples (that is, between RS1 and RS2). High gender congruence coefficients indicate that the factor loading patterns do not differ markedly between men and women (within RS1 and within RS2 and between men and women from RS1 and RS2).

The first test of congruence was a Pearson correlation coefficient. Pearson correlations are conducted on the factor loadings from 2 separate factor analyses and show

the extent to which factors are perfectly congruent ($r = 1.0$) or perfectly incongruent ($r = 0.0$). However, no guidelines exist to determine cut-off values for less than perfect congruence. The second test of congruence was the widely used Tucker congruence coefficient (CC) (30). Much like the Pearson correlation coefficient, this index assesses the similarity of 2 separate factor-loading patterns. In the original work of Tucker (30), he provided very rough guidelines for interpretation, stating that no more than 0.459 indicates lack of congruence whereas at least 0.900 is indicative of congruent factors. More recent work with this statistic (31) has shown that CCs from 0.85 to 0.94 can be taken as evidence of “fair” factor similarity, whereas CCs equal to or exceeding 0.95 suggests that 2 factors can be considered equal. The Pearson correlation coefficient and Tucker’s CC suffer the limitation that both reflect associations between variables but do not take into account mean level differences. Hence, high correlation and/or congruence coefficients could be observed resulting from similar patterns of loadings despite rather large differences in the magnitude of the loadings. To address this issue, the congruence test (32) also was computed, where values close to 0 indicate high levels of congruence and values close to 1 indicate incongruence. Unlike the Pearson correlation and the CC, the congruence test assesses differences between loadings in 2 separate factor matrices. SPSS 15.0 was used to conduct statistical analyses.

RESULTS

Demographic and Injury Characteristics

The sample consisted of a majority of white non-Hispanic individuals. Average age was slightly more than 40 years. The most common cause of injury was motor vehicle crash. Level and severity of injury were generally evenly distributed among the sample. Complete demographic characteristics of the sample by sex are given in Table 2.

Exploratory Factor Analyses

Analysis 1: 1- and 2-Factor Solutions in Random Sample 1 (Combined Sample, Men, and Women Separately). One-Factor Solution. In RS1, the 1-factor solution fit the structure of the PHQ-9 items well in the combined sample and explained approximately 41% of the original item variance (Table 3). This solution fit equally well for both men (43% original item variance) and women (38% original item variance; Table 3).

Two-Factor Solution. When 2 factors were specified for extraction from the 9 PHQ items, the solution was also satisfactory for RS1. This was the case in the combined sample, for men, and for women. The 2-factor solution accounted for approximately 46%, 50%, and 43% of the variance, respectively. Correlations between factors were 0.51 for the combined sample, 0.47 for men, and 0.39 for women (Table 4).

Analysis 2: 1- and 2-Factor Solutions in Random Sample 2 (Combined Sample, Men, and Women Separately). One-Factor Solution. In RS2, the 1-factor solution also fit the structure of the PHQ-9 items well in the combined sample, explaining approximately 44% of the variance in the original PHQ-9 items (Table 3). The 1-factor solution fit equally well for both men (45% original item variance) and women (42% original item variance).

Two-Factor Solution. When 2 factors were specified for extraction from the 9 PHQ items, the solution for RS2 was similar to the solution for RS1. This was the case in the combined sample, for men, and for women. The 2-factor solution accounted for approximately 48%, 51%, and 49% of the variance, respectively. Correlations between factors were 0.45 for the combined sample, 0.47 for men, and 0.44 for women (Table 4).

In the 2-factor solutions in analyses 1 and 2 for the combined samples in RS1 and RS2, factor 1 was generally characterized by cognitive/affective items such as feeling down and depressed, feeling bad about self, and feeling better of dead. Factor 2 was generally characterized by somatic items such as sleep problems, lack of energy, and eating problems. Having little interest or pleasure and moving slowly or being fidgety cross-loaded for both random samples.

Analysis 3: Cross-Validate Factor Solutions Between RS1 and RS2 for the Combined Group, Men, and Women. The congruence of the 1- and 2-factor solutions between RS1 and RS2 was examined using the 3 congruence indices and are summarized in Table 5. Congruence for the 1-factor model between RS1 and RS2 was quite high. The

Table 2. Demographic Characteristics of the Total Sample (N = 1,168)

	N (%)	
	Women	Men
Current age, years (Mean, SD) ^a	41.1 (13.1)	41.6 (13.2)
Age at injury, years (Mean, SD)	31.8 (13.3)	31.8 (13.2)
Marital status		
Single/never married	197 (33.7)	232 (39.7)
Married	212 (36.2)	224 (38.3)
Divorced	115 (19.7)	102 (17.4)
Separated	23 (3.9)	18 (3.1)
Widowed	38 (6.5)	9 (1.5)
Ethnic background		
White non-Hispanic	461 (78.8)	444 (75.9)
African American	81 (13.8)	102 (17.9)
Asian-Pacific Islander	6 (1.0)	2 (0.3)
Native American	9 (1.5)	12 (2.1)
Other/missing	28 (4.7)	22 (2.8)
Education level		
Less than high school	63 (10.8)	101 (17.3)
High school diploma	301 (51.5)	298 (50.9)
Associate's degree	47 (8.0)	48 (8.2)
Bachelor's degree	106 (18.1)	81 (13.8)
Master's degree	44 (7.5)	36 (6.2)
Doctorate	10 (1.7)	9 (1.5)
Other/missing	14 (2.4)	10 (2.0)
Employment status		
Working	156 (26.7)	158 (27.0)
Homemaker	47 (8.0)	2 (0.3)
Retired	41 (7.0)	56 (9.6)
Student	46 (7.9)	49 (8.4)
Unemployed	245 (41.9)	272 (46.5)
Other/missing	50 (8.6)	49 (8.2)
Level/severity of injury ^a		
Paraplegia, incomplete	108 (18.5)	108 (18.5)
Paraplegia, complete	180 (30.8)	180 (30.8)
Tetraplegia, incomplete	183 (31.3)	183 (31.3)
Tetraplegia, complete	114 (19.5)	114 (19.5)
Etiology of injury		
Motor vehicle crash	363 (62.1)	254 (43.4)
Fall	68 (11.6)	100 (17.1)
Gun shot wound	44 (7.5)	75 (12.8)
Medical/surgical complication	26 (4.4)	5 (0.9)
All other	84 (14.3)	151 (25.8)
Year of follow-up ^a		
Year 1	181 (30.9)	181 (30.9)
Year 5	129 (22.1)	129 (22.1)
Year 10	83 (14.2)	83 (14.2)
Year 15	69 (11.8)	69 (11.8)
Year 20	60 (10.3)	60 (10.3)
Year 25	63 (10.8)	63 (10.8)

^a Use of the variable in matching cases.

CC and congruence test were both approaching the limit of perfect congruence for the 1-factor solution for RS1 combined compared to RS2 combined, RS1 men compared with RS2 men, and RS1 women compared with RS2 women. The Pearson correlation was more

Table 3. Factor Loadings for RS1 and RS2 (1-Factor Model, Analyses 1 and 2)

	Combined RS1 Factor 1	Combined RS2 Factor 1	Men RS1 Factor 1	Men RS2 Factor 1	Women RS1 Factor 1	Women RS2 Factor 1
PHQ1—little interest or pleasure	0.768	0.725	0.782	0.706	0.751	0.742
PHQ2—down, depressed, hopeless	0.760	0.772	0.810	0.787	0.713	0.756
PHQ3—sleep problems	0.599	0.661	0.567	0.728	0.634	0.585
PHQ4—tired, little energy	0.648	0.696	0.670	0.687	0.625	0.714
PHQ5—eating problems	0.658	0.635	0.674	0.612	0.643	0.662
PHQ6—feeling bad about self	0.709	0.751	0.743	0.789	0.671	0.708
PHQ7—poor concentration	0.585	0.621	0.558	0.622	0.615	0.624
PHQ8—moving slow, fidgety	0.433	0.497	0.443	0.499	0.427	0.499
PHQ9—better off dead, hurt self	0.504	0.523	0.583	0.557	0.412	0.491
Eigenvalue	3.666	3.917	3.890	4.064	3.458	3.793
Coefficient α	0.854	0.868	0.864	0.876	0.842	0.858
Percent variance explained	41%	44%	38%	45%	45%	42%

variable and ranged from 0.791 to 0.948. Hence, the 1-factor model replicated from RS1 to RS2 almost perfectly in the combined sample, in men, and in women.

Congruence between factors extracted in RS1 and RS2 were more variable for the 2-factor model. For the RS1 and RS2 combined samples, the 2-factor structure was congruent, with coefficients approaching the limit of perfection for the Pearson coefficient, CC, and congruence test. Congruence was generally high between RS1 and RS2 men, although not as consistently as for the combined sample. Finally, congruence of the 2-factor structure between RS1 and RS2 women was more variable for the 3 congruence indices. Although the Pearson coefficients were quite low, the CC and congruence test values were fair. Hence, for women, the 2-factor structure is less generalizable than it is for men and for the combined group.

Analysis 4: Test Factor Congruence Between Sexes Within RS1 and RS2. The congruence of the factor structure (1-factor and 2-factor solutions) for men vs women within RS1 and RS2 was examined; congruence coefficients are given in Table 6. For the 1-factor solution, gender congruence was very high in both RS1 and RS2; the CC and congruence test were both approaching the limit of perfect congruence. The Pearson correlation, often a more conservative test than the congruence coefficient, was greater than or equal to 0.80. Hence, the 1-factor solution was virtually identical for men and women within RS1 and RS2.

The 2-factor solution showed a different pattern of results. In both RS1 and RS2, the CCs were at the low end of the recommendation of Tucker for “fair” similarity. Pearson correlations were much lower, and the congruence test was much higher than for the 1-factor solution. Hence, when 2 factors were extracted, there

Table 4. Factor Loadings for RS1 and RS2 (2-Factor Model, Analyses 1 and 2)

	Combined RS1		Combined RS2		Men RS1		Men RS2		Women RS1		Women RS2	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
PHQ1—little interest or pleasure	0.453	0.424	0.496	0.349	0.472	0.427	0.550	0.307	0.479	0.335	0.573	0.308
PHQ2—down, depressed, hopeless	0.609	0.266	0.620	0.285	0.648	0.282	0.607	0.262	0.651	0.254	0.894	−0.034
PHQ3—sleep problems	0.066	0.641	0.249	0.533	0.068	0.623	−0.070	0.894	0.434	0.410	0.339	0.425
PHQ4—tired, little energy	0.066	0.703	0.035	0.839	0.100	0.723	0.341	0.518	0.063	0.811	0.318	0.438
PHQ5—eating problems	0.194	0.569	0.325	0.417	0.130	0.690	0.360	0.426	0.310	0.405	0.276	0.515
PHQ6—feeling bad about self	0.619	0.201	0.607	0.274	0.786	0.082	0.556	0.260	0.710	0.200	0.492	0.295
PHQ7—poor concentration	0.492	0.180	0.425	0.297	0.451	0.186	0.627	0.085	0.419	0.297	0.365	0.370
PHQ8—moving slow, fidgety	0.234	0.261	0.254	0.329	0.352	0.155	0.506	0.060	0.087	0.520	0.033	0.523
PHQ9—better off dead, hurt self	0.603	−0.013	0.585	0.034	0.678	−0.004	0.532	0.022	0.670	−0.030	0.350	0.128
Eigenvalue	2.917	2.883	3.017	2.923	3.208	2.927	3.172	2.628	3.332	2.934	2.831	2.265
Coefficient α	0.807	0.740	0.817	0.743	0.831	0.773	0.811	0.762	0.821	0.723	0.778	0.704

Table 5. Factor Congruence Coefficients for 1- and 2-Factor Solutions in RS1 and RS2 (Combined, Men, and Women)

Sample Congruence Coefficients	One-Factor Solution			Two-Factor Solution					
	Combined	Men	Women	Combined		Men		Women	
	Factor 1	Factor 1	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Pearson correlation coefficient ^a	0.948 ^c	0.791 ^c	0.925 ^c	0.942 ^c	0.905 ^c	0.768 ^d	0.715 ^d	0.306	0.299
Tucker congruence coefficient ^a	0.998	0.994	0.998	0.983	0.970	0.941	0.908	0.858	0.773
Congruence test ^b	0.043	0.072	0.053	0.081	0.097	0.165	0.184	0.256	0.265

^a For Pearson and Tucker coefficients, values closer to 1 indicate congruent factors.

^b For the congruence test, values closer to 0 indicate congruent factors.

^c $P < 0.01$.

^d $P < 0.05$.

was far less congruence between men and women within each random sample. In this 2-factor model, congruence was slightly better for factor 1 than for factor 2, but neither was equal to the congruence for the 1-factor model. Thus, if 2 factors were extracted, they seemed to be different factors for men and women.

As shown in Table 4, factor loadings for the men in RS1 and RS2 indicated a cognitive/affective factor that included loss of interest/pleasure, feeling down and depressed, feeling bad about self, poor concentration, feeling better off dead, and moving slowly/feeling fidgety. The somatic factor included sleep problems, fatigue, and eating problems. For women, the most marked difference in loadings was for moving slowly/feeling fidgety to clearly load on the somatic factor. Whereas sleep problems loaded distinctly on the somatic factor for men, for women, there was more cross-loading for this item.

DISCUSSION

In considering the results of this study in the context of depression and SCI, it is important to first acknowledge that mathematical models of complex psychologic phenomena, such as depression, are inherently imperfect and at best give us some useful information for prediction, developing, and testing theories and some approximation of the “truth.” They should not be

interpreted as literal representations (33). Furthermore, despite the prevalence of depression in the SCI population (although figures vary) and a major focus on depression in the psychosocial literature in SCI (1), there is still much to understand. Therefore, this study is a small piece of a complex puzzle that will push our inquiry further to understand the nature of depression in men and women with SCI.

With that in mind, an initial cursory review of the factor analyses might suggest that both 1- and 2-factor solutions are acceptable; both solutions account for acceptable amounts of variance in the original items, and most loadings are of an acceptable magnitude (>0.3 – 0.4) (34,35) and in the expected direction. However, on closer examination, these analyses suggest that the 1-factor model is preferable with respect to gender congruence. The 2-factor model has several shortcomings. Both factors do not replicate well across samples, and the 2-factor solution is not gender congruent. Specifically, the 2-factor solution extracted for the sample of women diverges from that extracted for men and possesses much less congruence between women in RS1 and RS2 and hence generalizability. As such, the PHQ-9 is not likely to be reliable when comparing subscores of cognitive vs somatic symptoms between men and women with SCI. Alternatively, support for the 1-factor model congruence across sex

Table 6. Gender Congruence Coefficients for 1- and 2-Factor Models in RS1 and RS2 (Analysis 4)

Gender Congruence Coefficients ^a	One-Factor Solution		Two-Factor Solution			
	RS1	RS2	RS1		RS2	
	Factor 1	Factor 1	Factor 1	Factor 2	Factor 1	Factor 2
Pearson correlation coefficient ^b	0.808	0.800	0.496	0.470	0.282	0.452
Tucker congruence coefficient ^b	0.994	0.996	0.872	0.843	0.858	0.809
Congruence test ^c	0.079	0.064	0.496	0.386	0.259	0.255

^a Male to female comparisons within each random sample.

^b Values closer to 1 indicate congruent factors.

^c Values closer to 0 indicate congruent factors.

indicates that the PHQ-9 total score could be used to compare women and men.

Support for the unidimensionality of the PHQ-9 has been previously found. Bombardier et al (2) showed that all PHQ-9 items contributed toward probable major depression in a large sample of persons with SCI assessed 1 year after injury. Previous work also has shown the unidimensionality of the PHQ-9 in racially and ethnically diverse groups in primary care settings (non-SCI) (29); all items across ethnic groups had factor loadings greater than 0.50. However, a single dimension has not been common in factor analytic studies of depression measures in persons with SCI. Two-dimensional or 3-dimensional structures are often found representing cognitive/affective and somatic clusters of symptoms.

Frank et al (36) examined the factor structure of the Inventory to Diagnose Depression (37) (a depression screening measure based on the DSM-III-R criteria for major depressive disorder) in an SCI sample (along with rheumatoid arthritis, student, and community samples). A 4-factor structure emerged with a primary factor representing affective and cognitive aspects of depression such as negative self-evaluations, suicidal ideation, and depressed affect. For SCI, the second factor reflected common SCI sequelae such as decreased libido, sleep, appetite, energy, and psychomotor retardation.

Tate et al (38) factor analyzed the Zung Self-Rating Depression Scale (39) in a sample of persons with SCI and found a 2-dimensional structure. The affective factor was comprised of items reflected feeling useful, life is full, enjoying things, feeling hopeful about the future, and enjoying sex. The psycho/somatic factor was comprised of items related to being tired for no reason, trouble sleeping, irritability, crying spells, restlessness, and constipation. Krause et al (40) factor analyzed the Older Adult Health and Mood Questionnaire (41) in a large sample of adults with SCI and found 3 factors reflecting evaluative (ie, negative evaluation of life, hopelessness about future), affective (ie, sadness, tearfulness), and reduced activity domains. It is worth noting, however, that these studies examined different depression measures, but none examined factor congruence. As with most statistical techniques, factor analyses can capitalize on chance associations, and hence, these multifactorial solutions should be replicated in different samples using the same instrument. Although multifactorial solutions have appeared commonly in the literature, the important question is whether these solutions are reliable and useful.

The results of this study raise an important question regarding how dimensions of depression may vary between men and women with SCI. Although there was support for the 1-factor structure of the PHQ between sexes, the low congruence between sexes for the 2-factor structure should not be ignored. These results tell us something important about how certain symptoms may be experienced or interpreted differently

by men and women. This should not be surprising considering the debate over how equitably depression measures and diagnostic criteria represent both men and women's experience of depressive symptoms. For example, in this sample, moving slowly or feeling fidgety loaded on different factors for men and women. For men, this item loaded on the cognitive/affective factor, whereas for women, it loaded on the somatic factor. Sleep problems definitively loaded on the somatic factor for men, whereas for women, sleep problems cross-loaded on both factors. One interpretation may be that, to men, psychomotor slowing or feeling fidgety is experienced more cognitively and/or affectively, such as emotional restlessness or feeling weighed down mentally. Women may experience this symptom more somatically, such as being unable to sit still or feeling physically weighed down.

The results of these analyses support the use of the PHQ-9 as a unidimensional representation of depressive symptoms that is generalizable across samples to men and women. However, its use to understand and compare multiple dimensions of depressive symptoms across sexes or samples is not supported. Further work is needed to more fully explore these dimensions within and between sexes and to determine why there seems to be an unstable structure for women in particular.

The unidimensionality of the PHQ-9 in this analysis also points to the issue of how to handle transdiagnostic symptoms when measuring depression in SCI. As we stated, such transdiagnostic symptoms can potentially lead to over- or underdiagnosis of depression in this population. Williams et al (42) proposed 2 primary approaches to handling the effect of physical illness on the diagnosis of depression that provide some framework for considering the results of this study. The inclusive approach counts depressive symptoms toward the diagnosis of depression irrespective of whether the symptom is judged to be from medical and psychologic causes. The etiologic approach reflects the DSM's criteria that count symptoms toward a diagnosis unless the symptom is clearly accounted for by a medical condition.

The data from this study may potentially support the inclusive model of diagnosing probable major depression when using the PHQ-9; that is, counting somatic symptoms including sleep disturbance, appetite changes, anergia, and so forth toward a diagnosis of major depression rather than assuming they are necessarily nonspecific or only related to SCI itself. Support for the 1-factor model in men and women in this sample suggest that both somatic and nonsomatic symptoms are endorsed on the PHQ-9 in a similar fashion. However, this does not indicate whether somatic symptoms are or are not related to depression vs SCI and warrants continued research into effectively distinguishing the origin of such symptoms before an inclusive approach is embraced. Therefore, until there is evidence to the contrary, ignoring or downplaying somatic symptoms

may contribute to underdiagnosis and undertreatment of depression.

Limitations

One limitation of this study is that data were collected 1 or more years after SCI and hence may not be generalizable to people who are examined soon after SCI, especially during inpatient rehabilitation, when this population is often screened for depressive disorders. During this acute period, rates of general psychologic and physical distress, including transdiagnostic symptoms, may be substantially higher elevating total scores. Consequently, the relationship between somatic symptoms of major depression and cognitive or affective symptoms may be different. Another limitation of this study is the absence of repeated measures on the same participants; these data would allow for the consistency of the factor structure to be determined over time among the same participant cohort. Another shortcoming is the absence of data collected by clinicians to verify the consistency of probable depressive disorders with diagnostic interviews or other methodologies.

Future Directions in Research

One of the most important findings of this study is the low gender congruence of a 2-factor structure of the PHQ-9 and the low stability of the 2-factor structure within women. In light of support for the 1-factor model, this lack of congruence for the 2-factor model likely indicates there are potentially important differences between sexes in their experience of depressive symptoms that may be missed. Given our limited knowledge of gender differences in the experience of depression in the context of SCI, this is an area clearly needing further exploration. Specifically, where sexes diverge in cognitive, affective, and somatic dimensions of depressive symptoms and whether sex-specific or sex-neutral measures are warranted should be a focus of future research.

Longitudinal research also is needed to identify changes in depressive symptoms over time and the stability of factor structures across time as it relates to adjustment to SCI. Optimally, this research would start shortly after injury during inpatient rehabilitation. It will also be important to establish the consistency of factor structure across race-ethnicity within the SCI population. The unidimensionality of the PHQ-9 items in this sample does not represent the extant literature on factor structures of other depression measures, both screening and severity, which has more often supported 2- and 3-dimensional models. This may be a function of the sample and/or the PHQ-9 itself and replication in other SCI samples is recommended. Finally, exploration of the dimensionality of the Structured Diagnostic Interview for DSM (43) (SCID) and how similarly the PHQ-9 is to that structure would be worthwhile in further understanding the dimensions of depression in this population and differences between sex.

ACKNOWLEDGMENTS

The authors thank Jane Walters, data coordinator, University of Michigan Model SCI Care System, for assistance in data collection.

APPENDIX Congruence Coefficient

$$CC_{pq} = \frac{\sum_{i=1}^n a_{ip} \cdot a_{iq}}{\sqrt{\sum_{i=1}^n a_{ip}^2 \cdot \sum_{i=1}^n a_{iq}^2}}$$

where ...

CC_{pq}: is the congruence coefficient
n : number of variables in two samples
p : number of factor in first sample
q : number of factor in second sample
a : factorial loading

The congruence coefficient (CC), originally popularized by Tucker (1951), is computed as the sum or the cross-products of the factor loadings taken as a ratio of the square root of the product of the summed loadings squared (see above formula). Unfortunately, the CC can produce spuriously high values when factors containing loadings with the same sign and a large proportion of highly loading variables are compared. The CC has no agreed upon test of statistical significance.

Congruence Test

$$CT_{pq} = \sqrt{\frac{\sum_{i=1}^n (a_{ip} - a_{iq})^2}{n}}$$

(See above for definitions of variables in equation)

The congruence test (CT) is computed as the square root of the average sum of the squared deviations between the factor loadings from two samples. The CT does not suffer from overestimation as CC does in some instances; it has no agreed upon test of statistical significance.

REFERENCES

1. Elliott TR, Frank RG. Depression following spinal cord injury. *Arch Phys Med Rehabil*. 1996;77(8):816–823.
2. Bombardier CH, Richards JS, Krause JS, Tulsy D, Tate DG. Symptoms of major depression in people with spinal cord injury: implications for screening. *Arch Phys Med Rehabil*. 2004;85(11):1749–1756.
3. Malec J, Neimeyer R. Psychologic prediction of duration of inpatient spinal cord injury rehabilitation and performance of self-care. *Arch Phys Med Rehabil*. 1983;64(8):359–363.
4. Herrick S, Elliott T, Crow F. Social support and the prediction of health complications among persons with spinal cord injuries. *Rehabil Psychol*. 1994;39(4):231–250.
5. Elliott T, Shewchuck R. Social support and leisure activities following severe physical disability: testing the mediating effects of depression. *Basic Appl Soc Psych*. 1995;16(4):471–487.

6. Fuhrer JM, Rintala DH, Hart KA. Depressive symptomatology in persons with spinal cord injury who reside in the community. *Arch Phys Med Rehabil*. 1993;74(3):255–260.
7. MacDonald M, Nielson W, Cameron M. Depression and activity patterns of spinal cord injured persons living in the community. *Arch Phys Med Rehabil*. 1987;68(6):339–343.
8. Schulz R, Decker S. Long-term adjustment to physical disability: the role of social support, perceived control, and self-blame. *J Pers Soc Psychol*. 1985;48(5):1162–1172.
9. Frank R, Elliott T, Corcoran J, Wonderlich S. Depression after spinal cord injury: is it necessary? *Clin Psychol Rev*. 1987;7(6):611–630.
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
11. Greden JF. Physical symptoms of depression: unmet needs. *J Clin Psychiatry*. 2003;64(suppl 7):5–11.
12. Rosenthal M, Christensen BK, Ross TP. Depression following traumatic brain injury. *Arch Phys Med Rehabil*. 1998;79(1):90–103.
13. Kroenke K. The interface between physical and psychological symptoms. *Prim Care Companion J Clin Psychiatry*. 2003;5(suppl 7):11–18.
14. Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J. An international study of the relation between somatic symptoms and depression. *N Engl J Med*. 1999;341(18):1329–1335.
15. Kalpakjian C, Albright K. An examination of depression through the lens of spinal cord injury: comparative prevalence rates and severity in women and men. *Womens Health Issues*. 2006;16(6):380–388.
16. Cochran SV, Rabinowitz FE. Gender-sensitive recommendations for assessment and treatment of depression in men. *Prof Psychol Res Pr*. 2003;34(2):132–140.
17. Moller-Leimkuhler AM, Bottlender R, Strauss A, Rutz W. Is there evidence for a male depressive syndrome in inpatients with major depression? *J Affect Disord*. 2004;80(1):87–93.
18. Brownhill S, Wilhelm K, Barclay L, Schmied V. 'Big build': hidden depression in men. *Aust N Z J Psychiatry*. 2005;39(10):921–931.
19. Carter JD, Joyce PR, Mulder RT, Luty SE, McKenzie J. Gender differences in the presentation of depressed outpatients: a comparison of descriptive variables. *J Affect Disord*. 2000;61(1–2):59–67.
20. Marcus SM, Young EA, Kerber KB, et al. Gender differences in depression: findings from the STAR*D study. *J Affect Disord*. 2005;87(2–3):141–150.
21. Shafer A. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton and Zung. *J Clin Psychol*. 2006;61(1):123–146.
22. Spitzer RL, Kroenke K, Williams JBW. Validation and utility of a self-report version of PRIME-MD—the PHQ primary care study. *JAMA*. 1999;282(18):1737–1744.
23. American Spinal Injury Association. *International Standards for Neurological and Functional Classification of Spinal Cord Injury*. Chicago, IL: American Spinal Injury Association; 2002.
24. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9—validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.
25. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2006;28(1):71–77.
26. Fann JR, Bombardier CH, Dikmer S, et al. Validity of the patient health questionnaire-9 in assessing depression following traumatic brain injury. *J Head Trauma Rehabil*. 2005;20(6):501–511.
27. Williams LS, Brizendine EJ, Plue L, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*. 2005;36(3):635–638.
28. Lowe B, Schenkel I, Carney-Doebbeling C, Gobel C. Responsiveness of the PHQ-9 to psychopharmacological depression treatment. *Psychosomatics*. 2006;47(1):62–67.
29. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med*. 2006;21(6):547–552.
30. Tucker LR. *A Method for Synthesis of Factor Analysis Studies*. Washington, DC: Department of the Army; 1951.
31. Lorenzo-Seva U, Berge J. Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*. 2006;2(2):54–67.
32. Herrero FJ, Grossi FJ, y Cuesta M. Factor analysis and factor congruence. Paper presented at: 6th European Meeting of the Psychometric Society; July 17–19, 1989; Leuven, Belgium.
33. MacCallum RC. Working with imperfect models. *Multivariate Behav Res*. 2003;38(1):113–139.
34. Stevens J. *Applied Multivariate Statistics for the Social Sciences*. 4th ed. Mahwah, NJ: Lawrence Erlbaum & Assoc.; 2002.
35. Tabachnick B, Fidell L. *Using Multivariate Statistics*. 4th ed. Needham Heights, MA: Allyn & Bacon; 2001.
36. Frank R, Chaney J, Clay D. Dysphoria: a major symptom factor in persons with disability or chronic illness. *Psychiatry Res*. 1992;43(3):231–241.
37. Zimmerman M, Coryell W, Corenthal C, Wilson S. A self-report scale to diagnose major depressive disorder. *Arch Gen Psychiatry*. 1986;43(11):1076–1081.
38. Tate D, Forchheimer M, Maynard F, Davidoff G, Dijkers M. Comparing two measures of depression in spinal cord injury. *Rehabil Psychol*. 1993;38(1):53–61.
39. Zung W. A self-rating depression scale. *Arch Gen Psychiatry*. 1965;12(1):63–70.
40. Krause J, Kemp BJ, Coker J. Depression after spinal cord injury: relation to gender, ethnicity, aging, and socioeconomic indicators. *Arch Phys Med Rehabil*. 2000;81(8):1099–1109.
41. Kemp BJ, Adams BM. The Older Adult Health and Mood Questionnaire: a measure of geriatric depressive disorder. *J Geriatr Psychiatry Neurol*. 1995;8(3):162–167.
42. Williams JW, Noel PH, Cordes JA, Ramirez G, Pignone M. Is this patient clinically depressed? *JAMA*. 2002;287(9):1160–1170.
43. First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), Clinician Version: User's Guide*. Washington, DC: American Psychiatric Press; 1997.