

Original article

Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates

Katherine Milette¹, Marie Hudson², Murray Baron²,
Brett D. Thombs¹ and Canadian Scleroderma Research Group*

Abstract

Objective. The reported rates of depressive symptoms in patients with SSc are high. The Center for Epidemiologic Studies Depression Scale (CES-D) is the only measure of depressive symptoms validated for SSc patients. The objective of this study was to assess the internal consistency reliability, convergent validity and strength of association with clinical correlates of the 9-item version of the Patient Health Questionnaire depression scale (PHQ-9) compared with the CES-D in SSc.

Methods. We conducted a cross-sectional, multicentre study of 566 SSc patients who were assessed with the PHQ-9 and CES-D scales, and through clinical histories and medical examinations. Internal consistency reliability was assessed with Cronbach's α , convergent validity with Pearson's correlation and the relationship of socio-demographic and clinical variables with the PHQ-9 and CES-D scores using multiple linear regression.

Results. Scale reliability was good for the PHQ-9 ($\alpha=0.87$) and similar to the CES-D ($\alpha=0.90$). Correlations of the PHQ-9 total score were -0.68 with mental health, -0.43 with physical health, 0.44 with disability, 0.40 with pain and 0.79 with fatigue, which were all in the expected direction and similar to the results with the CES-D. Regression coefficients of clinical correlates did not differ significantly between models using the PHQ-9 and CES-D.

Conclusion. The PHQ-9 is reliable and valid for use as a measure of depressive symptom severity in patients with SSc and performs similarly to the CES-D. However, the PHQ-9 is advantageous because it is half the length of the CES-D, easily administered and scored, and is increasingly used across many patient groups for assessment in research and clinical settings.

Key words: Systemic sclerosis, Scleroderma, Depression, Assessment, Center for Epidemiologic Studies Depression Scale, Patient Health Questionnaire 9-item version.

¹Department of Psychiatry and ²Division of Rheumatology, McGill University and Jewish General Hospital, Montréal, Québec, Canada.

Submitted 11 September 2009; revised version accepted 25 November 2009.

Correspondence to: Brett D. Thombs, Jewish General Hospital, 4333 Cote Ste Catherine Road, Montreal, Quebec H3T 1E4, Canada. E-mail: brett.thombs@mcgill.ca

*See Appendix 1 for a list of members of the Canadian Scleroderma Research Group

Introduction

SSc is a chronic autoimmune CTD characterized by thickening and fibrosis of the skin, involvement of internal organs and significant morbidity and mortality [1]. SSc is associated with high levels of fatigue and pain and substantial reductions in quality of life [2, 3]. A systematic review found that between 36 and 65% of patients with

SSc report depressive symptoms above cut-off levels for potential clinical significance based on validated questionnaires [4].

A number of different measures have been used to assess psychological distress in SSc, including the Beck Depression Inventory (BDI) [5–9], the Zung Self-rating Depression Scale (Zung SDS) [10], the Center for Epidemiologic Studies Depression Scale (CES-D) [11], the Montgomery–Asberg Depression Rating Scale (MADRS) [12], the Hospital Anxiety and Depression Scale (HADS) [13], the Delusions Symptoms State Inventory/states of Anxiety and Depression scale (DSSI/sAD) [14], the General Health Questionnaire (GHQ) [15] and the Psychosocial Adjustment to Illness Scale (PAIS) [9]. The BDI, Zung SDS, CES-D and MADRS assess depressive symptoms. The HADS and DSSI/sAD assess symptoms of both anxiety and depression [13]. The GHQ is a screening instrument used to estimate the likelihood that a patient will be diagnosed with a psychiatric illness and has been widely used in rheumatic diseases [15]. The PAIS assesses adjustment to medical illness and measures seven different domains of functioning, including psychological distress [9].

Only the CES-D [16], however, a 20-item questionnaire of symptoms of depression, has been validated for use among patients with SSc. A previous study reported that the CES-D had good internal consistency reliability and convergent validity among 470 patients with SSc from the Canadian Scleroderma Research Group (CSRG) Registry and that its 4-factor structure was consistent with structures reported in non-medical settings [17].

In primary care, there is little evidence that any particular depression assessment instrument performs better than other instruments [18]. Therefore, other considerations, such as an instrument's brevity, readability and comprehensibility are important to consider. The Patient Health Questionnaire depression scale (PHQ-9) [19] is shorter and more feasibly administered for both research and clinical purposes than the CES-D. The PHQ-9 is a 9-item measure of depression severity with a user-friendly response format, short administration time and easy scoring. The instrument was developed for both screening and severity assessment, and its 9 items map directly onto the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* [20] criteria for a major depressive episode, which is not the case for the CES-D. The PHQ-9 has been translated into at least 25 languages and been shown to be as accurate as longer tools for identifying major depression in a range of settings, countries and patient populations [20, 21].

The objective of this study was to compare the internal consistency reliability, convergent validity and clinical correlates of the PHQ-9 with the internal consistency reliability, convergent validity and clinical correlates of the CES-D, which has already been validated for patients with SSc. Since prior research has found that different depression symptom assessment scales tend to perform similarly [8], we hypothesized that the CES-D and PHQ-9

would not differ substantively in their performance characteristics.

Patients and methods

Patient sample

The study sample consisted of patients who completed CSRG Registry annual visits from November 2007 through July 2009 and completed the PHQ-9 and CES-D. Patients in the Registry were recruited from 15 centres across Canada. In order to be eligible for the Registry, patients must have a diagnosis of SSc made by the referring rheumatologist, be ≥ 18 years of age and be fluent in English or French. Specific diagnostic criteria were not required for entry into the Registry. The ACR criteria [13], published in 1980, have been shown to be outdated with improved understanding of SSc disease processes. Subsequent classification systems have been proposed, but none has gained universal approval [14]. Thus, one of the objectives of the Registry is to improve upon existing diagnostic systems. The Registry patients undergo extensive clinical history, physical evaluation and laboratory investigations and complete a series of self-report questionnaires. Patients from all sites provided informed consent and the McGill University Research Ethics Board approved the study.

Measures

Symptoms of depression were assessed with the CES-D and PHQ-9. Self-report measures of mental health function (SF-36 Mental Component Summary), physical function (SF-36 Physical Component Summary), disability (HAQ–Disability Index), pain (McGill Pain Questionnaire–Short Form) and fatigue (Functional Assessment of Chronic Illness Therapy–Fatigue Scale) were used to establish convergent validity. Analyses of clinical correlates of depression included socio-demographic data, measures of disease severity and duration, and specific SSc symptoms.

CES-D. The 20-item CES-D assesses the frequency of occurrence of symptoms during the past week on a 0–3 Likert-type scale ('rarely or none of the time' to 'most or all of the time'), and total scores range from 0 to 60. Standard cut-offs are ≥ 16 for 'possible depression' and ≥ 23 for 'probable depression' [16]. The CES-D has been shown to have good reliability and convergent validity with related self-report measures in a sample of 470 SSc patients from the CSRG Registry [17].

PHQ-9. The 9-item PHQ-9 rates the frequency of symptoms over the past 2 weeks on a 0–3 Likert-type scale ('not at all' to 'nearly every day'). The total score ranges from 0 to 27 and the standard cut-off threshold for 'moderate' depression severity is a score of ≥ 10 [19–21].

The Short-Form 36 Health Survey Questionnaire. The Short-Form 36 Health Survey Questionnaire (SF-36) [22, 23] is the most widely used and evaluated health outcome measure and has extensive evidence for its validity

and reliability in multiple populations [17]. It consists of eight domains, including physical functioning, social functioning, role limitations related to physical problems, role limitations related to emotional problems, mental health, vitality, bodily pain and general health perceptions. Each domain can be scored separately, with scores ranging from 0 (the worst health state) to 100 (the best health state). Domain scores can also be summarized into a Physical Component Summary score and a Mental Component Summary score. The Physical Component Summary and Mental Component Summary are scored using norm-based scoring based on a general population sample to produce *T*-scores for each patient [mean (s.d.) of 50 (10)]. Version 2 of the SF-36 was used in this study.

HAQ–Disability Index. The HAQ–Disability Index is a 20-item self-administered measure intended to assess functional ability in patients with arthritis [24]. Questions can be divided into eight categories (dressing, standing, eating, walking, toileting, reaching, gripping and instrumental activities) and the use of assistive aids and devices to help with function is recorded. Scores are derived as the average of the score for the most abnormal activity in each of the eight categories, taking into account the use of assistive aids and devices. Scores range from 0 (no disability) to 3 (severe disability) [25].

McGill Pain Questionnaire–Short Form. The McGill Pain Questionnaire–Short Form [26] is a 15-item pain rating index, containing 11 items referring to the sensory dimension and 4 that refer to the affective dimension of pain. Each descriptor is ranked on a 4-point intensity scale (0 = none, 1 = mild, 2 = moderate and 3 = severe) with a total score ranging from 0 to 45. The McGill Pain Questionnaire–Short Form has been used extensively in different patient groups and has excellent psychometric properties [22].

Functional Assessment of Chronic Illness Therapy–Fatigue Scale. The Functional Assessment of Chronic Illness Therapy–Fatigue Scale [27] is a 13-item measure that assesses fatigue severity and impact with items on a 0–4 Likert scale. The total scores range from 0 to 52 and higher scores indicate more severe fatigue. The Functional Assessment of Chronic Illness Therapy–Fatigue Scale has strong psychometric properties, including internal consistency, stability, validity and responsiveness to change in a number of chronic illnesses, including cancer [27], RA [28] and others [29].

Higher scores on the HAQ–Disability Index, McGill Pain Questionnaire–Short Form and Functional Assessment of Chronic Illness Therapy–Fatigue Scale indicate greater disability, pain and fatigue, respectively. These were expected to be positively associated with higher scores on the CES-D and PHQ-9. Higher scores on the SF-36 Mental Component Summary score and SF-36 Physical Component Summary score indicate better function, and were expected to be negatively associated with CES-D scores. The association of the CES-D and PHQ-9 with the SF-36 Mental Component Summary score was

expected to be the most robust since the SF-36 Mental Component Summary score measures the mental health and has a strong depression component.

Demographic information was based on self-report and included age, gender, education, marital status and race/ethnicity. Patients' medical histories and disease characteristics were obtained via clinical histories and examinations by study physicians. Limited skin disease was defined as skin involvement distal to the elbows and knees with or without face involvement. SSc disease duration was determined as the time from onset of non-Raynaud's symptoms based on a clinical history obtained by study physicians. Global disease severity was rated by study physicians on a 0–10 numerical rating scale [30]. Skin involvement was assessed using the modified Rodnan skin score, ranging from 0 to 51 [31]. Tender joint count was recorded by study physicians using a 28-joint count [32]. The number of gastrointestinal symptoms was determined by patient report from a checklist that included weight loss, anorexia, dysphagia, reflux, pyrexia, choking at night, early satiety, bloating, nausea/vomiting, constipation, diarrhoea, malabsorption, fecal incontinence, antibiotics for bacterial overgrowth and hyperalimentation. Shortness of breath was assessed by the patient on a 0–10 numerical rating scale [33].

Data analyses

Internal consistency reliability was evaluated using Cronbach's α . Convergent validity of the PHQ-9 and CES-D with other self-report measures was assessed using Pearson's correlation coefficient. We compared the associations of the CES-D and PHQ-9 with clinical correlates of depressive symptoms using the linear regression model and predictor variables that were published in a previous assessment of clinical correlates of the CES-D [34]. Clinical correlates that were entered into the model included demographic (age, sex), socioeconomic (married or living as married vs single/divorced/widowed, greater than high school education vs high school or less), global disease severity (disease duration and physician-rated global severity) and specific symptoms (total skin score, number of tender joints, number of gastrointestinal symptoms and breathing problems). Variables chosen for the original model for the CES-D [31], which was replicated in this study, were based on theoretical and clinical rationales, rather than statistical significance. For comparison purposes, standardized regression coefficients and CIs were calculated [35] for models with the CES-D and PHQ-9 as the dependent variables. All tolerance values in the regression models were between 0.69 and 0.97, and all bivariate correlations between variables included in the models were ≤ 0.49 , indicating that multicollinearity was not an issue. All analyses were conducted using SPSS version 16.0 (Chicago, IL, USA), and all statistical tests were two-sided with a significant $P < 0.05$.

Results

Sample characteristics

Table 1 shows patient demographic and disease characteristics. Of 566 patients included in the study, 88% ($n=497$) were females and 92% were White ($n=520$). The mean (s.d.) age of the sample was 56.5 (11.8) years, 48% ($n=274$) of patients completed some post-secondary education and 70% ($n=398$) were married or living as married.

The mean duration since the onset of non-Raynaud's symptoms was 12.1 (9.1) years (median 9.9), and the mean duration since the diagnosis of SSc was 9.4 (7.9) years (median 7.5). Approximately 30% ($n=169$) of patients had diffuse SSc, the mean physician-rated global severity score was 2.9 (2.2) (median 2.0) and mean total skin score was 9.8 (8.8) (median 7.0). The mean number of tender joints was 1.5 (4.9) (median 0.0), mean number of gastrointestinal problems was 3.4 (2.8) (median 3.0) and mean breathing problems on a 0–10 scale was 1.9 (2.6) (median 1.0).

The mean (s.d.) CES-D score was 13.7 (10.6) (median 11.0; Fig. 1). Over a third scored at least 16 on the CES-D ($n=194$, 34.3%), a standard cut-off for 'possible depression', and 20.8% ($n=118$) scored ≥ 23 for 'probable depression'. The mean PHQ-9 score was 5.9 (5.4) (median 5.0; Fig. 2). Just over 20% scored ≥ 10 on the PHQ-9 ($n=117$, 20.7%), which is a standard cut-off for moderate severity of depressive symptoms. The correlation of the CES-D with the PHQ-9 was 0.77. Based on Cohen's κ , chance-corrected agreement between the CES-D and PHQ-9 was 0.49 with a CES-D cut-off of 16 and 0.54 with a CES-D cut-off of 23, both of which are considered moderate based on rules of thumb developed by Altman [36]. Table 2 compares patients above cut-off thresholds based on CES-D scores of ≥ 16 and ≥ 23 vs PHQ-9 scores ≥ 10 .

Reliability of the PHQ-9 and CES-D

Overall scale reliability was good and similar for the CES-D ($\alpha=0.90$) and PHQ-9 ($\alpha=0.87$). Corrected item-total correlations for individual CES-D items ranged from 0.28 to 0.72. Corrected item-total correlations for individual PHQ-9 items ranged from 0.47 to 0.78.

Convergent validity

Pearson correlations (with 95% CIs) between the CES-D total score and related self-report measures were: SF-36 Mental Component Summary score -0.76 ($-0.79, -0.72$); SF-36 Physical Component Summary score -0.33 ($-0.40, -0.25$); HAQ-Disability Index 0.42 (0.35, 0.48); McGill Pain Questionnaire-Short Form 0.40

Fig. 1 Distribution of CES-D total scores.

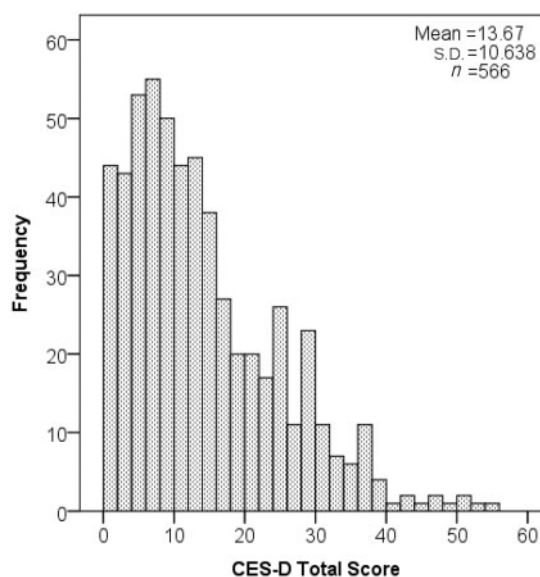


TABLE 1 Patient demographic and disease characteristics

Variables	All patients ($n=566$)	CES-D		PHQ-9	
		r	P -value	r	P -value
Female sex, n (%)	497 (87.8)	0.00	0.940	-0.01	0.852
White, n (%)	520 (91.9)	-0.07	0.095	-0.06	0.171
More than high school education, n (%)	274 (48.4)	-0.07	0.093	-0.07	0.079
Married or living as married, n (%)	398 (70.3)	-0.06	0.177	-0.07	0.103
Diffuse SSc, n (%)	169 (29.9)	0.06	0.126	0.03	0.553
Age, mean (s.d.), years	56.51 (11.80)	-0.03	0.455	-0.03	0.516
Time since onset of non-Raynaud's symptoms, mean (s.d.), years	12.11 (9.12)	0.03	0.500	0.04	0.346
Time since diagnosis of SSc, mean (s.d.), years	9.38 (7.93)	0.00	0.969	-0.01	0.743
Physician-rated global disease severity (range 1–10), mean (s.d.)	2.90 (2.20)	0.17	<0.001	0.20	<0.001
Modified Rodnan total skin score, mean (s.d.)	9.79 (8.80)	0.14	0.001	0.13	0.003
Number of tender joints, mean (s.d.)	1.54 (4.85)	0.11	0.012	0.11	0.007
Number of gastrointestinal symptoms, mean (s.d.)	3.38 (2.84)	0.25	<0.001	0.27	<0.001
Breathing problems (range 1–10), mean (s.d.)	1.93 (2.62)	0.31	<0.001	0.29	<0.001

(0.33, 0.47); and Functional Assessment of Chronic Illness Therapy–Fatigue Scale 0.69 (0.62, 0.71). The correlations between the PHQ-9 total score and the same set of self-report measures were: SF-36 Mental Component Summary score -0.68 ($-0.64, -0.73$); SF-36 Physical Component Summary score -0.43 ($-0.36, -0.49$);

HAQ–Disability Index 0.44 (0.37, 0.50); McGill Pain Questionnaire–Short Form 0.40 (0.32, 0.46); and Functional Assessment of Chronic Illness Therapy–Fatigue Scale 0.79 (0.75, 0.82). All correlations were in the expected direction. The only significant difference between the CES-D and PHQ-9 scores was with the Functional Assessment of Chronic Illness Therapy–Fatigue Scale, for which the correlation was significantly higher for the PHQ-9, although the magnitude of the difference was not large.

Fig. 2 Distribution of PHQ-9 total scores.

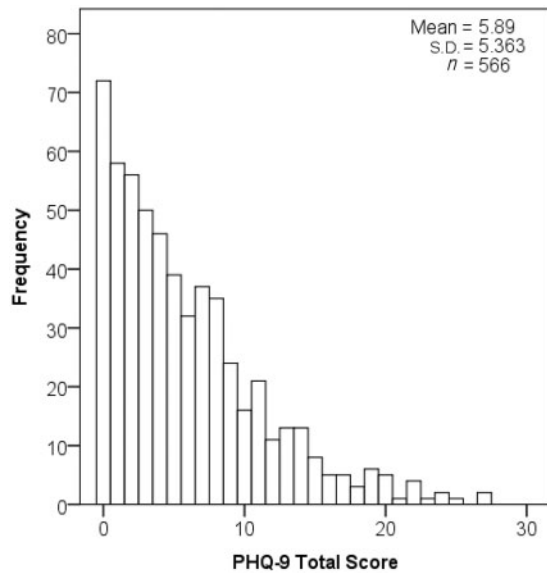


TABLE 2 Patients above and below standard cut-off scores on the CES-D vs the PHQ-9

	CES-D <16	CES-D ≥ 16	CES-D <23	CES-D ≥ 23
PHQ-9 < 10	352	97	406	43
PHQ-9 ≥ 10	20	97	42	75

Predictors of symptoms of depression

Results from the linear regression models are shown in Table 3. Clinical correlates accounted for 14.5% of the variance in CES-D scores vs 19.5% of the variance for the PHQ-9 scores. There were no significant or substantive differences between coefficients produced using the CES-D vs the PHQ-9 for any of the independent variables included in the models, as the CIs overlapped in all cases. Education level, modified Rodnan skin score, tender joint count, breathing problems and a number of gastrointestinal symptoms were independently associated with CES-D scores. For the PHQ-9, education level and modified Rodnan skin score were not statistically significant, although the magnitude of their coefficients departed only minimally from estimates produced by the CES-D.

Discussion

This study is the first study to assess the reliability and validity of the PHQ-9 for measuring depressive symptoms in patients with SSc. Estimates of internal consistency reliability and convergent validity were good and similar to those of the CES-D, which has been previously validated in SSc [17]. Regressions of the PHQ-9 and CES-D scores on clinical correlates, including socio-demographic, disease duration and severity, and specific SSc symptoms, did not differ significantly or

TABLE 3 Clinical correlates of depressive symptoms as measured by the CES-D and PHQ-9 depression scales

Variables	CES-D		PHQ-9	
	Standardized regression coefficient (95% CI)	P-value	Standardized regression coefficient (95% CI)	P-value
Age, years	-0.047 ($-0.13, 0.03$)	0.247	-0.052 ($-0.13, 0.03$)	0.190
Female sex	0.023 ($-0.05, 0.10$)	0.556	0.002 ($-0.07, 0.07$)	0.952
More than high school education	-0.083 ($-0.16, -0.01$)	0.033	-0.027 ($-0.10, 0.05$)	0.469
Married or living as married	-0.041 ($-0.12, 0.34$)	0.282	-0.028 ($-0.10, 0.05$)	0.454
Time since onset of non-Raynaud's symptoms, years	-0.021 ($-0.10, 0.06$)	0.602	-0.020 ($-0.10, 0.06$)	0.619
Physician-rated disease severity (0–10)	0.007 ($-0.08, 0.10$)	0.871	0.012 ($-0.07, 0.10$)	0.786
Modified Rodnan skin score	0.089 ($0.01, 0.17$)	0.036	0.031 ($-0.05, 0.11$)	0.456
Tender joint count	0.097 ($0.02, 0.17$)	0.012	0.095 ($0.02, 0.17$)	0.011
Breathing problems (0–10)	0.278 ($0.19, 0.37$)	<0.001	0.325 ($0.24, 0.41$)	<0.001
Number of gastrointestinal symptoms	0.214 ($0.13, 0.29$)	<0.001	0.251 ($0.17, 0.33$)	<0.001

substantively depending on whether the PHQ-9 or CES-D was used.

The PHQ-9 scale was developed in 1999 [19] and is used increasingly in clinical and research settings due to its short length and the ease with which it can be administered and scored [8–10]. Another advantage is the growing data on its performance as a severity measure and a depression screening tool across many different medical patient groups [9, 10]. This study showed that the PHQ-9 is a reliable and valid measure of depressive symptoms among patients with SSc that can be potentially used for both research and clinical purposes as a measure of severity of depressive symptoms. The similar psychometric performance of the PHQ-9, along with its ease of use and relative brevity, make it attractive compared with the longer CES-D for use in SSc.

A lower proportion of patients crossed the threshold for potentially clinically significant symptoms of depression using the PHQ-9 compared with the CES-D based on standard cut-off scores. This finding in itself is not necessarily surprising. The cut-off scores across depression screening tools do not always produce consistent rates of patients who screen positive for possible depression. For instance, a systematic review of depression screening tools used in post-myocardial infarction patients [36] found that the two most commonly used tools, the BDI and the HADS, produced consistently different rates of possible cases across studies. The BDI identified 31.1% of 10 785 patients across eight studies as having at least 'mild symptoms of depression' based on a cut-off of ≥ 10 . The HADS, on the other hand, identified only 7.3% of 830 patients across four studies as 'probable cases' based on a cut-off of ≥ 11 and 15.5% of 863 patients from four studies using a somewhat more liberal cut-off of ≥ 8 . There is a need for cut-off scores to be set and validated specifically for patients with SSc. No studies of patients with SSc have validated any depression questionnaire against diagnoses of major depression based on a structured clinical interview for the purpose of establishing diagnostic accuracy, and studies are needed to do this. Currently, the CSRG is conducting such a study with the PHQ-9. In addition, future research should examine the relatively high correlation between both the PHQ-9 and CES-D and fatigue. The high correlations reported here suggest the possibility that depression scores may be influenced more than that would be desired by symptoms of the disease, such as fatigue. On the other hand, a previous study that compared 400 patients from the CSRG Registry with a matched sample of 400 community respondents on the CES-D found that high rates of depressive symptoms in SSc were not due to relatively higher rates of endorsement of somatic symptoms among SSc patients compared with non-medically ill respondents [37].

A potential limitation of the study is that it was based on a convenience sample of patients with SSc. The present sample of SSc patients generally had stable disease (median disease duration 12 years since the onset of non-Raynaud's symptoms). Patients who are not being

cared for by a rheumatologist and patients with very severe SSc who were too sick to participate or who died earlier in their disease course, were not included in the present study. Thus, there may have been an over-representation of healthier patients in our SSc sample (survival cohort), and results may therefore not be generalizable to the full spectrum of SSc. Nonetheless, even in this potentially 'healthier sample' the prevalence and severity of depressive symptoms were high. Despite these limitations, the demographic and clinical characteristics of CSRG Registry patients are consistent with other outpatient SSc samples that have been reported in the research literature [1]. In addition, patient data were drawn from 15 centres across Canada, which is a major strength of the study.

In summary, the PHQ-9 is a reliable and valid measure of depressive symptoms in SSc that performs similar to the CES-D, but is more quickly and easily administered and scored. Based on feasibility, the PHQ-9 is recommended for assessment of symptom severity in research and clinical settings with SSc patients. Research is needed, however, to assess the diagnostic accuracy of the PHQ-9 against diagnoses of major depressive disorder based on structured clinical interviews.

Rheumatology key messages

- The PHQ-9 is reliable and valid in SSc.
- The PHQ-9 performs like the CES-D, but is more quickly and easily administered and scored.

Acknowledgements

B.D.T. and M.H. are supported by New Investigator Awards from the Canadian Institutes of Health Research and Établissement de Jeunes Chercheurs awards from the Fonds de la Recherche en Santé Québec. M.B. is the director of the Canadian Scleroderma Research Group, which receives grant funding from the Canadian Institutes of Health Research, the Cure Scleroderma Foundation, the Scleroderma Society of Canada, the Ontario Arthritis Society, Actelion Pharmaceuticals and Pfizer Pharmaceuticals.

Disclosure statement: The authors have declared no conflicts of interest.

References

- 1 Mayes M, Lacey J, Beebe-Dimmer J *et al.* Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum* 2003;48:2246–55.
- 2 Thombs BD, Hudson M, Bassel M, Taillefer SS, Baron M, Canadian Scleroderma Research Group. Sociodemographic, disease, and symptom correlates of fatigue in systemic sclerosis. *Arthritis Rheum* 2009;61:966–73.
- 3 Hudson M, Thombs BD, Steele R, Panopalis P, Newton E, Baron M. Health-related quality of life in systemic

- sclerosis: a systematic review. *Arthritis Rheum* 2009;61: 1112–20.
- 4 Thombs BD, Taillefer SS, Hudson M, Baron M. Depression in patients with systemic sclerosis: a systematic review of the evidence. *Arthritis Rheum* 2007;57:1089–97.
 - 5 Beretta L, Astori S, Ferrario E, Caronni M, Raimondi M, Scorza R. Determinants of depression in 111 Italian patients with systemic sclerosis. *Reumatismo* 2006;58: 219–25.
 - 6 Matsuura E, Ohta A, Kanegae F *et al*. Frequency and analysis of factors closely associated with the development of depressive symptoms in patients with scleroderma. *J Rheumatol* 2003;30:1782–7.
 - 7 Benrud-Larson LM, Heinberg LJ, Boling C *et al*. Body image dissatisfaction among women with scleroderma: Extent and relationship to psychosocial function. *Health Psychol* 2003;22:130–9.
 - 8 Benrud-Larson LM, Haythornthwaite JA, Heinberg LJ *et al*. The impact of pain and symptoms of depression in scleroderma. *Pain* 2002;95:267–75.
 - 9 Roca RP, Wigley FM, White B. Depressive symptoms associated with scleroderma. *Arthritis Rheum* 1996;39: 1035–40.
 - 10 Mozzetta A, Antinone V, Alfani S *et al*. Mental health in patients with systemic sclerosis: a controlled investigation. *J Eur Acad Dermatol Venereol* 2008;22: 336–40.
 - 11 Nietert PJ, Mitchell HC, Bolster MB, Curran MY, Tilley BC, Silver RM. Correlates of depression, including overall and gastrointestinal functional status, among patients with systemic sclerosis. *J Rheumatol* 2005;32:51–7.
 - 12 Legendre C, Allanore Y, Ferrand I, Kahan A. Evaluation of depression and anxiety in patients with systemic sclerosis. *Joint Bone Spine* 2005;72:408–11.
 - 13 Richards HL, Herrick A, Griffin K, Gwilliam PD, Fortune DG. Psychological adjustment to systemic sclerosis: exploring the association of disease factors, functional ability, body related attitudes and fear of negative evaluation. *Psychol Health Med* 2004;9:29–39.
 - 14 Angelopoulos NV, Drosos AA, Moutsopoulos HM. Psychiatric symptoms associated with scleroderma. *Psychother Psychosom* 2001;70:145–50.
 - 15 Hyphantis TN, Tsifetaki N, Pappa C *et al*. Clinical features and personality traits associated with psychological distress in systemic sclerosis patients. *J Psychosom Res* 2007;62:47–56.
 - 16 Radloff LS. The CES-D scale: a self-report depression scale for research in general population. *Appl Psych Meas* 1977;1:385–401.
 - 17 Thombs BD, Hudson M, Schieir O, Taillefer SS, Baron M. Reliability and validity of the Center for Epidemiologic Studies Depression Scale in patients with systemic sclerosis. *Arthritis Rheum* 2008;59:438–43.
 - 18 Williams JW, Pignone M, Ramirez G, Perez Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002;24:225–37.
 - 19 Kroenke KS, Robert L. The PHQ-9: a new depression diagnostic and severity measure. *Psychiat Ann* 2002;32: 509–15.
 - 20 Furukawa TA. Assessment of mood: guides for clinicians. *J Psychosom Res* 2009 (in press).
 - 21 Wittkamp KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007;29:388–95.
 - 22 Ware JJ, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center, 1993.
 - 23 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
 - 24 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23: 137–45.
 - 25 Hudson M, Steele R, Taillefer S, Baron M. Quality of life in systemic sclerosis: psychometric properties of the World Health Organization Disability Assessment Schedule II. *Arthritis Rheum* 2008;59:270–8.
 - 26 Melzack R. The short-form McGill Pain Questionnaire. *Pain* 1987;30:191–7.
 - 27 Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *J Pain Symptom Manage* 1997;13: 63–74.
 - 28 Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811–9.
 - 29 Chandran V, Bhella S, Schentag C, Gladman DD. Functional Assessment of Chronic Illness Therapy-Fatigue Scale is valid in patients with psoriatic arthritis. *Ann Rheum Dis* 2007;66:936–9.
 - 30 Clements PJ, Seibold JR, Furst DE *et al*. High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis trial: lessons learned. *Semin Arthritis Rheum* 2004; 33:249–63.
 - 31 Clements P, Lachenbruch P, Siebold J *et al*. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. *J Rheumatol* 1995;22:1281–5.
 - 32 Prevo ML, van't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;38:44–8.
 - 33 Steen VD, Medsger TA. The value of the Health Assessment Questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. *Arthritis Rheum* 1997; 40:1984–91.
 - 34 Thombs BD, Hudson M, Taillefer SS, Baron M. Prevalence and clinical correlates of symptoms of depression in patients with systemic sclerosis. *Arthritis Rheum* 2008;59: 504–9.
 - 35 Cohen J, Cohen P, West SG, Aiken LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd edition. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2003.
 - 36 Altman DG. Practical statistics for medical research. London, England: Chapman and Hall, 1991.

37 Thoms BD, Fuss S, Hudson M *et al.* High rates of depressive symptoms among patients with systemic sclerosis are not explained by differential reporting of somatic symptoms. *Arthritis Rheum* 2008;59: 431–7.

Appendix 1

CSRG Investigators: J. Pope, London, Ontario; J. Markland, Saskatoon, Saskatchewan; N. Khalidi,

Hamilton, Ontario; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; A. Masetto, Sherbrooke, Québec; E. Sutton, Halifax, Nova Scotia; E. Kaminska, Hamilton, Ontario; P. Docherty, Moncton, New Brunswick; J.-P. Mathieu, Montréal, Québec; M. Abu-Hakima, Calgary, Alberta; D. Smith, Ottawa, Ontario; S. LeClercq, Calgary, Alberta; S. Ligier, Montréal, Québec; and S. Mittoo, Winnipeg, Manitoba.