Discordance between Patient and Physician Assessments of Disease Severity in Systemic Sclerosis

MARIE HUDSON, ANN IMPENS, MURRAY BARON, JAMES R. SEIBOLD, BRETT D. THOMBS, JENNIFER G. WALKER, the Canadian Scleroderma Research Group, and RUSSELL STEELE

ABSTRACT*. *Objective. To describe the magnitude and correlates of discordance between patient and physician assessments of disease severity in patients with systemic sclerosis (SSc).

Methods. Subjects were patients enrolled in the Canadian Scleroderma Research Group Registry. The outcomes of interest were patient and physician global assessments of disease severity (scales ranging from 0-10). Predictors of disease severity represented the spectrum of disease in SSc (skin involvement, severity of Raynaud's phenomenon, shortness of breath, gastrointestinal symptoms and pain, number of fingertip ulcers, tender and swollen joints, creatinine, and fatigue). The results of the analysis were validated in an independent sample of patients with SSc from the United States.

Results. Patients perceived greater disease severity than physicians (mean difference 0.78 ± 2.65). The agreement between patient and physician assessments of disease severity was, at best, modest (intraclass correlation 0.3774; weighted κ 0.3771). Although both patients and physicians were influenced by skin scores, breathlessness, and pain, the relative importance of these predictors differed. Patients were also influenced by other subjective symptoms, while physicians were also influenced by disease duration and creatinine. The predictors explained 56% of the deviance in the patient global assessments and 29% in the physician assessments. These findings were confirmed in the US dataset.

Conclusion. Patients and physicians rate SSc disease severity differently in magnitude and are influenced by different factors. Patient-assessed and physician-assessed measures of severity should be considered as complementary and used together in future studies of SSc. (First Release Sept 15 2010; J Rheumatol 2010;37:2307–12; doi:10.3899/jrheum.100354)

Key Indexing Terms:
OUTCOME ASSESSMENT COHORT STUDIES SYSTEMIC SCLEROSIS

Discordance of assessments between patients and physicians occurs when patients and physicians assign different values to a health trait[1]. Discordance between patient and physician assessments of disease activity has been described in several rheumatic diseases, including rheumatoid arthritis (RA)[2,3], systemic lupus erythematosus[1,4,5], and ankylosing spondylitis[6]. In those studies, when rating disease activity, patient assessments were more strongly associated with subjective symptoms, such as pain, psychological well-being, and function, while physician assessments were more strongly associated with objective findings, including laboratory tests. Discordance has the potential to impede patient care; patients may fail to comply with medical instructions if they are poorly informed of their condition or if physicians fail to appreciate the full effect of disease on their patients.

Little is known about the presence and magnitude of possible discordance in the assessment of disease activity in systemic sclerosis (SSc) in part probably because measuring disease activity in SSc is particularly difficult. Unlike systemic lupus erythematosus and RA, SSc is not characterized by episodes of acute inflammation (manifested by synovitis, pleuritis, dermatitis, and nephritis) that can be easily differentiated from quiescent phases. Instead, the clinical features of SSc are attributable to vascular and connective tissue fibrosis that is more difficult to appreciate and quantify than inflammation and, when it becomes measurable, has often progressed to permanent damage. Many patients, especially those with limited skin involvement, have an indolent course without clear signs of inflammation. Further, elevated acute-phase proteins are inconsistently associated with early SSc, leading some to argue that patients with SSc may have an impaired acute-phase response[7,8].

From the Jewish General Hospital and McGill University, Montreal, Quebec, Canada; the University of Michigan Scleroderma Program, Ann Arbor, Michigan, USA; and Flinders Medical Centre, Bedford Park, South Australia, Australia.

Supported in part by the Canadian Institutes of Health Research, the Scleroderma Society of Canada, the Cure Scleroderma Foundation, and educational grants from Actelion Pharmaceuticals and Pfizer Inc. Dr. Hudson is funded by a New Investigator Award from the Canadian Institutes of Health Research. Additional funding was provided by the Jonathan and Lisa Rye Scleroderma Research Fund and the Marvin and Betty Danto Research Fund at the University of Michigan.

M. Hudson, MD, MPH; M. Baron, MD; B.D. Thombs, PhD; R. Steele, PhD, Jewish General Hospital and McGill University; A. Impens, PhD; J.R. Seibold, MD, the University of Michigan Scleroderma Program; J.G. Walker, MD, Flinders Medical Centre.

Address correspondence to Dr. M. Hudson, Jewish General Hospital, Room A-725, 3755 Cote Ste Catherine Road, Montreal, Quebec H3T 1E2, Canada. E-mail: marie.hudson@mcgill.ca

Accepted for publication July 16, 2010.

Given the difficulty of measuring disease activity in SSc and of separating it from disease damage, disease severity has been proposed as an appropriate measure of disease status in SSc. Indeed, Medsger defines disease severity in SSc as the total effect of disease on organ function at a given point in time, including both reversible (activity) and irreversible (damage) components[9] and, given the difficulties in defining disease activity, this is likely to be a better measure of disease status and possible discordance in SSc.

Thus, we undertook this study to (1) identify the extent to which patient and physician assessments of disease severity differed, and (2) identify and compare the predictors of patient and physician assessments of disease severity in patients with SSc.

MATERIALS AND METHODS

Design. We performed a cross-sectional study of a Canadian sample of patients with SSc and confirmed the results using a sample of patients with SSc from the United States.

Study subjects. The Canadian subjects were patients enrolled in the Canadian Scleroderma Research Group (CSRG) Registry. Patients in this registry are recruited from the practices of rheumatologists across Canada. They must have a diagnosis of SSc made by the referring rheumatologist, be > 18 years of age, be fluent in English or French, and likely to be compliant with study procedures and visits. The patients included were those whose baseline visit was between September 2004 and 2008. The US patients were recruited from the University of Michigan Scleroderma Program between December 2005 and April 2006. A total of 105 sequential ambulatory patients with SSc were recruited and consented to participate in a study on hand functioning. Four subjects did not complete the study.

Outcome measures. The patient and physician global assessments of disease severity in the Canadian patients were done using numerical rating scales (NRS) ranging from 0-10. The NRS scale is simple to complete and score and has been shown to be as reliable and responsive as visual analog scales (VAS) to measure disease activity and function in ankylosing spondylitis[10] and more reliable to assess pain in patients with RA[11]. Physicians were asked to "rate the patient's overall health for the past week" and the NRS was anchored by the descriptors "no disease" and "very severe disease." Patients were asked to "rate your disease in the past week" and the NRS was anchored by "no disease" and "very severe limitation." The patient and physician global assessments of disease severity in the US patients were assessed using a VAS ranging from 0-100 mm, anchored by the descriptors "no severity" and "extremely severe." The scores on the VAS of 0-100 were divided by 10 to be comparable to the NRS ratings ranging from 0-10. Although the wording of the anchors on the global assessments differed slightly, the scores ranging from 0-10 were assumed to be equivalent.

Predictor variables. Potential predictors of disease severity were chosen to represent the spectrum of disease in SSc, and included severity of Raynaud's phenomenon (RP), skin involvement, fingertip ulcers, shortness of breath, joint symptoms, gastrointestinal (GI) symptoms, kidney involvement, pain, and fatigue. In both samples, the methods for data collection were similar. The extent of skin involvement was recorded using the modified Rodnan skin score. Similarly, the number of fingertip ulcers and a simplified 28 swollen and tender joint count[12] were recorded by physical examination by a well-trained health professional using standardized definitions. Creatinine was documented by laboratory testing.

Data on RP, GI symptoms, shortness of breath, and pain were assessed using a self-report measure, the Scleroderma-Health Assessment Questionnaire (S-HAQ)[13]. The S-HAQ consists of the Disability Index of the HAQ (HAQ-DI) and items to measure symptoms specific for SSc using VAS scales. The HAQ-DI is a self-administered measure intended to assess functional ability in arthritis[14]. The disease-specific questions in the S-HAQ relate to the severity of various symptoms, including RP, GI symptoms, shortness of breath, and pain in the past week. Each item is anchored by the adjectives "does not interfere" and "very severe limitation" and scored separately. The Canadian patients answered the disease-specific questions on the S-HAQ using an 11-point NRS, while a 0-100 mm VAS was used by the US patients.

Finally, fatigue was measured using the Vitality subscale of the Medical Outcomes Study Short Form-36 (SF-36) questionnaire[15,16]. The SF-36 Vitality subscale includes 4 Likert items with 5 response options each (all of the time to none of the time) that assess patients' level of fatigue during the previous 4 weeks. Scores are normalized with a mean of 50 and SD of 10. Scores below 50 represent worse fatigue and above 50, less. The SF-36 Vitality subscale has been used to measure fatigue in general population samples and in patients with medical illness and injury. A recent systematic review concluded that the SF-36 Vitality subscale has good evidence for validity, reliability, sensitivity to change, and feasibility in RA[17].

Statistical analysis. The initial analyses were done using the Canadian data. The standard measure of agreement for quantitative measures is the intraclass correlation coefficient (ICC) and for ordered categorical variables, the weighted κ statistic. Using the disease severity scores ranging from 0-10 in turn as continuous or ordinal variables, we calculated the ICC and the weighted κ statistic. We also fit a linear mixed model that isolated heterogeneity due to the physicians from overall disagreement to determine whether physician heterogeneity was responsible for disagreement between patient and physician assessments.

We undertook subsequent analyses to identify the predictors of patient and physician global assessments of disease severity, using generalized linear models (in particular, normal, Poisson, and negative binomial regression models). We fit 3 separate sets of models for each of the patient and physician global assessments of disease severity. The first set included all the selected covariates of severity. The second set included only the physician-recorded correlates. The third set included only the patient-recorded correlates of severity. In all regression models, we adjusted for demographic variables (age, gender, ethnicity, education) as well as disease duration. In multivariate analyses using generalized linear models, we found that a negative binomial regression model fit the data well for 4 of the 6 regression models. We observed underdispersion rather than overdispersion in the 2 other models (both of which used the patient-reported variables), so the results between the negative binomial and Poisson models yielded very similar results. Model fit was assessed using percentage deviance explained, which is analogous to R^2 in standard linear regression models. Finally, because we identified differences in predictors of patient and physician global assessments of disease severity, we undertook a regression analysis to identify the predictors of the differences. We used normal linear regression to predict the difference between patient and physician severity scores, as there was no reason (using either model selection criteria or diagnostics) that suggested a normal assumption was inappropriate.

Lastly, we sought to confirm our findings by running the results of our models in the US data. We used the estimated regression coefficients from the Canadian data to calculate predicted physician and patient severity assessments for the US data and estimated the association between the predicted assessments and the observed assessments using simple linear regression.

At the time of analysis, the CSRG had 936 patients entered in its registry, of which 742 had complete data for the variables of interest in this study. The US sample had 101 subjects, of whom 61 had complete data. Data between patients included and excluded from the analyses were compared and there were no systematic differences. Therefore, only patients with complete data were included in the analyses. All statistical analyses were performed with SPSS v. 13 and the R statistical package[18].

Ethical considerations. Each patient provided informed written consent to participate in the data collection process and ethics committee approval for our study was obtained at each site.

RESULTS

There were 803 patients included in this study, of which 742 were Canadian and 61 from the United States (Table 1). In the Canadian sample, 87% were women, mean age was 55.5 (± 12.4) years, and mean disease duration since the onset of the first non-RP disease manifestation was 10.7 (± 9.0) years. In the US sample, 86% were women, mean age was 51.4 (± 11.4) years, and mean disease duration since the onset of the first non-RP disease manifestation was 7.5 (± 8.4) years. On a scale ranging from 0 to 10, with 0 being the lowest and 10 being the greatest disease severity, the mean patient and physician global assessments of disease severity were 3.63 (± 2.54) and 2.85 (± 2.27), respectively, in the Canadian sample and 4.25 (± 2.59) and 2.04 (± 1.78), respectively, in the US sample. The mean difference between patient and physician assessment was 0.78 (± 2.65) in the Canadian sample and 2.21 (± 2.65) in the US sample. The positive values suggest that, on average, patients perceived greater disease severity than physicians. Of note, the difference in patient and physician ratings of disease severity in diffuse patients was 0.53 (95% CI 0.23, 0.84), and in limited patients, 0.92 (95% CI 0.66, 1.17). This was not statistically significant.

Table 1. Baseline characteristics of study subjects. Fatigue was measured using the Medical Outcomes Study Short Form-36 questionnaire vitality subscale. Scores are normalized with a mean of 50 and standard deviation of 10. Scores below 50 represent worse fatigue and above 50, more vitality. Values are mean (SD) unless otherwise indicated.

	Canadian Subjects, n = 742	US Subjects, n = 61
Age, yrs	55.48 (12.39)	51.44 (11.37)
Disease duration, yrs	10.70 (8.98)	7.48 (8.35)
Skin score (0–51)	10.69 (9.62)	8.20 (7.27)
Raynaud's (0–10)	2.83 (2.87)	4.11 (3.00)
Shortness of breath (0–10)	2.04 (2.60)	2.32 (2.77)
Gastrointestinal symptoms (0–10)	1.80 (2.60)	1.87 (2.52)
Pain (0–10)	3.65 (2.74)	4.23 (2.48)
Number of fingertip ulcers	1.28 (2.46)	0.55 (1.47)
Swollen joint count (0–28)	0.65 (2.32)	0.30 (1.34)
Tender joint count (0–28)	1.45 (3.78)	0.98 (2.74)
Fatigue	48.85 (21.70)	42.34 (22.83)
Creatinine, umol/l	83.77 (53.74)	79.42 (35.45)
Patient assessment of severity (0–10)	3.63 (2.54)	4.25 (2.59)
Physician assessment of severity (0–10)	2.85 (2.27)	2.04 (1.78)
Females, no. (%)	642 (86.50)	55 (85.94)
Diffuse disease, no. (%)	299 (40.30)	31 (48.44)
White, no. (%)	601 (81.00)	58 (90.62)
Education > high school, no. (%)	355 (47.84)	48 (75.00)

Agreement between patient and physician global assessments of disease severity in the Canadian data. Using the disease severity scores ranging from 0-10 either as continuous or ordinal variables, we observed very similar ICC and weighted κ statistics (0.3774 and 0.3771, respectively). The values for these statistics indicate at best only fair agreement between patient and physician assessments of disease severity. We observed a slight difference in the extent of agreement in the 2 disease subsets [ICC of 0.29 (95% CI 0.19, 0.39) in the limited subset and ICC of 0.41 (95% CI 0.31, 0.50) in the diffuse subset], although this was not statistically significant.

A linear mixed model was used to assess the extent to which interphysician variability was responsible for the lack of agreement between the patient and physician severity scores. We did observe statistically significant variability between physicians in their assessments [Bayesian Information Criterion (BIC) of 3202 for a model that accounted for physician heterogeneity vs 3215 for a model that did not]. A difference of 6-10 in the value of the BIC indicates strong evidence against the null hypothesis and a difference of more than 10 indicates very strong evidence[19]. Thus, a difference of 13 suggests very strong evidence against the model, assuming no between-physician heterogeneity in assessments of disease severity. Nevertheless, only about 5% of the overall variability in patient severity scores could be explained by the differences among the physicians themselves.

Thus, based on these analyses, we concluded that agreement between patient and physician assessments of disease severity was, at best, modest. Interphysician variability in assessments accounted for only a small part of the differences in assessments.

Predictors of patient and physician global assessments of disease severity in the Canadian sample. We identified similarities and differences in the predictors of patient and physician global assessments of disease severity (Table 2). The OR reported in Table 2 represent the relative increase in the response (i.e., the patient or physician assessments of severity) for a 1-unit increase in the covariate of interest (e.g., skin score, shortness of breath, etc.). Thus, although skin scores, shortness of breath, and pain were significant predictors of both patient and physician global assessments of disease severity when all covariates were included in the models, their relative effects on physician and patient assessments differed. Thus, an increase of 1 unit in skin score was associated with about a 3% increase in the physician assessment of severity, controlling for all other variables (i.e., about a 15% increase for a 5-unit increase in skin score). In contrast, we estimated only a corresponding 0.9% increase in patient severity assessment for a 1-unit increase in skin score (again controlling for all other variables) or a 4.5% increase in mean patient assessment for a 5-unit increase in skin score. The OR estimates for shortness of breath were fairly similar in the models predicting patient (1.062) and physician (1.094) assessments of severity separately. However, pain had a larger effect in the model predicting patient-assessed severity (1.121), compared to its effect in the model predicting physician-assessed severity (1.032).

In addition, significant predictors of patient assessments included severity of RP, GI symptoms, and fatigue. The coefficient < 1 for fatigue reflects the fact that for the measurement of fatigue, lower scores represent worse fatigue, while for the global assessment, lower scores represent less-severe disease. In turn, other significant predictors of physician assessments included disease duration, with early disease being considered worse, and creatinine. The regression models using all patient-reported and clinical covariates explained 56% of the deviance in the patient global assessments and 29% in the physician assessments, respectively. As expected, the patient-reported variables by themselves explained much more deviance in the patient assessment than the physician assessment (54% vs 14%) and the clinical variables by themselves explained more deviance in the physician assessment than the patient assessment (18% vs 5%). We also noted (but do not show) a significant interaction between disease duration and skin score in the models for the physician assessments ($p < 0.001$) that indicated that the amount by which the physician score would increase for high skin scores would be smaller for patients with longer disease duration.

Finally, given that we found differences in the predictors of patient-assessed and physician-assessed severity, we regressed the difference between the patient and physician assessments to determine what was most associated with the discordance between them (Table 3). Pain, GI symptoms, RP, and fatigue were associated with significantly higher values for the difference (i.e., contributed more to the patient assessment than the physician assessment). Increased skin score and creatinine were associated with significantly lower values for the difference (i.e., contributed more to the physician assessment than the patient assessment). Further, we again found a significant interaction between skin score and duration in this model that suggested that the longer the disease duration, the less an increased skin score would be associated with the difference (data not shown).

Confirmation of the models in the US sample. To confirm our findings, we used the regression coefficients obtained from the Canadian data to predict physician assessments of severity, patient assessments of severity, and the difference between patient and physician assessments in the US patients. In these analyses, we allowed for the US and Canadian data to have different overall means, so as to examine the relationship of severity with the covariates, rather than the overall population mean. We found that the regression coefficients derived from the Canadian data explained 15.7% of the variability in the physician global assessments in the US data. This can be compared to an estimated prediction R^2 of 25.1% in the Canadian data. Similarly, regression coefficients from the Canadian data explained 43.4% of the variability in the patient assessment scores in the US data, compared to a prediction R^2 of 54.8% on the Canadian patient assessments. Finally, the Canadian model for the differences in assessments explained 22.3% of the variability in the difference in assessments in the US data, compared to a prediction R^2 of 33.3% for the Canadian data. Thus, prediction in the US data using the Canadian models was reasonably good.

We also investigated whether individual variables had a different relationship with disease severity in the Canadian and US samples. We found no strong evidence that the relationship between any of the covariates and the patient or

Table 2. Negative binomial regression results to identify predictors of the physician (MD) and patient (Pt) global assessments of disease severity in the Canadian data. This table contains the estimated OR with 95% CI for the 6 different models. Values in bold type indicate CI that do not overlap with 0. Note that creatinine was transformed by taking the square root in order to improve model assumptions and decrease the influence of outlying points. Results are given as square root of creatinine. The coefficient < 1 for fatigue reflects the fact that for the measurement of fatigue, lower scores represent worse fatigue.

	All Covariates		Clinical Covariates Only		Patient Reported Covariates Only	
	MD Assessment	Pt Assessment	MD Assessment	Pt Assessment	MD Assessment	Pt Assessment
Age	1.000 (0.997, 1.006)	1.006 (0.997, 1.004)	1.004 (0.992, 1.009)	0.999 (0.994, 1.004)	1.000 (0.995, 1.004)	1.000 (0.996, 1.003)
Female	1.027 (0.893, 1.184)	1.042 (0.931, 1.168)	0.974 (0.838, 1.134)	0.964 (0.822, 1.130)	0.903 (0.777, 1.051)	0.996 (0.893, 1.112)
White	0.962 (0.849, 1.092)	1.046 (0.947, 1.158)	0.971 (0.851, 1.109)	1.017 (0.887, 1.165)	0.995 (0.866, 1.145)	1.061 (0.961, 1.175)
Disease duration	**0.993 (0.987, 0.999)**	0.997 (0.992, 1.002)	0.995 (0.989, 1.001)	1.001 (0.994, 1.007)	**0.986 (0.980, 0.992)**	0.996 (0.991, 1.000)
> High school	1.025 (0.928, 1.132)	1.024 (0.946, 1.108)	0.995 (0.896, 1.106)	0.943 (0.846, 1.051)	0.992 (0.889, 1.107)	1.013 (0.937, 1.096)
Shortness of breath	**1.094 (1.073, 1.116)**	**1.062 (1.046, 1.078)**	—	—	**1.082 (1.058, 1.106)**	**1.060 (1.044, 1.075)**
Pain	**1.032 (1.010, 1.055)**	**1.121 (1.101, 1.140)**	—	—	**1.045 (1.021, 1.070)**	**1.123 (1.103, 1.142)**
GI symptoms	0.987 (0.966, 1.008)	**1.018 (1.002, 1.034)**	—	—	0.990 (0.967, 1.014)	**1.018 (1.003, 1.033)**
Fatigue	0.999 (0.997, 1.001)	**0.994 (0.992, 0.996)**	—	—	0.998 (0.995, 1.001)	**0.994 (0.992, 0.996)**
Raynaud's	0.983 (0.964, 1.002)	**1.028 (1.013, 1.042)**	—	—	0.984 (0.963, 1.005)	**1.028 (1.014, 1.043)**
Skin score	**1.030 (1.025, 1.035)**	**1.009 (1.005, 1.013)**	**1.030 (1.025, 1.035)**	**1.012 (1.006, 1.018)**	—	—
Fingertip ulcers	1.009 (0.989, 1.028)	1.008 (0.991, 1.024)	1.009 (0.988, 1.030)	1.019 (0.996, 1.042)	—	—
Swollen joints	1.004 (0.980, 1.028)	1.010 (0.992, 1.028)	0.999 (0.973, 1.024)	0.998 (0.973, 1.024)	—	—
Tender joints	0.996 (0.982, 1.009)	0.993 (0.983, 1.003)	1.005 (0.990, 1.020)	**1.023 (1.008, 1.039)**	—	—
Creatinine	**1.036 (1.011, 1.062)**	1.012 (0.991, 1.032)	**1.035 (1.008, 1.063)**	1.008 (0.978, 1.039)	—	—
Deviance explained, %	28.9	55.9	17.5	4.6	13.8	54.3

Table 3. Linear regression results to identify the predictors of the difference between patient and physician global assessments of severity in the Canadian data. Values in bold indicate CI that do not include 0. Note that creatinine was transformed by taking the square root in order to improve model assumptions and decrease the influence of outlying points on the results. Results are given in terms of the square root of creatinine.

	Estimated Coefficient (95% CI)	p
Age	–0.003 (–0.017, 0.011)	0.63
Female	0.015 (0–0.483, 0.453)	0.95
White	0.253 (–0.145, 0.652)	0.21
Duration	0.010 (–0.008, 0.029)	0.28
Education beyond high school	–0.043 (–0.360, 0.274)	0.79
Shortness of breath	0.007 (–0.062, 0.076)	0.85
Pain	**0.327 (0.256, 0.399)**	**< 0.0001**
Gastrointestinal symptoms	**0.136 (0.067, 0.206)**	**< 0.0005**
Fatigue	**–0.015 (–0.023, –0.006)**	**< 0.001**
Raynaud's	**0.164 (0.100, 0.227)**	**< 0.0001**
Skin score	**–0.069 (–0.86, –0.052)**	**< 0.0001**
Fingertip ulcers	0.000 (–0.068, 0.068)	0.99
Swollen joints	0.017 (–0.058, 0.092)	0.66
Tender joints	–0.015 (–0.062, 0.032)	0.52
Creatinine	–0.102 (–0.192, –0.013)	0.025
Deviance explained	37.95%	

physician assessments depended on the sample (data not shown).

DISCUSSION

We found some similarities but also important differences in how patients and physicians rate disease severity in SSc. On average, patients rated disease severity as worse than physicians did. Patient and physician severity ratings were associated with both physician-rated skin scores and patient-reported shortness of breath and pain in their assessments of severity, although skin was more strongly associated for physicians than patients and pain was a more robust correlate for patients than physicians. Patient severity assessments were also significantly influenced by self-reported estimates of the severity of RP, GI symptoms, and fatigue, while physician global severity ratings were influenced by disease duration and creatinine.

Our report demonstrated that, using global assessments, patients and physicians rate disease severity differently in magnitude and are influenced by different factors. The implications of our findings are 2-fold. First, our findings suggest that traditional biomedical assessments of disease status in SSc (e.g., physician assessments of skin involvement or laboratory tests such as creatinine) may be supplemented by patient-derived information. In other words, patient-reported severity allows for more aspects of the disease to be captured than physician-reported assessments. In fact, it is striking that the predictors of importance for patients but not physicians were indeed in relation to symptoms for which good outcome measures in SSc are currently lacking (in particular GI symptoms and fatigue) or where patient reports are the only means of obtaining the information (in particular RP).

Second, in the absence of a gold standard to measure disease severity in SSc, both patient and physician global assessments of disease severity could be used together, to better approximate "true" disease severity. Indeed, in a study of RP in patients with SSc, both physician and patient assessments of RP activity were found to be valid and reliable and the authors recommended that both be included in the core set of measures for use in future clinical trials in this area[20]. Similarly, although definitive validation of patient and physician global assessments of disease severity in SSc has yet to be done, our data suggest that the 2 measures may provide complementary data and both should be considered as outcome measures in this highly heterogeneous disease.

There are limitations that should be considered in interpreting the results of our study. First, patients in the CSRG registry are a convenience sample of patients with SSc. Their median disease duration since the onset of non-RP symptoms was 10 years, suggesting a sample of patients with generally stable disease. Moreover, patients with very severe SSc who were too sick to participate or who died earlier in their disease course were not included. This may have resulted in an overrepresentation of healthier patients in our SSc sample (survival cohort), and results may therefore not be generalizable to the full spectrum of SSc. Despite these limitations, the demographic and clinical characteristics of the CSRG Registry patients in this study were consistent with other outpatient SSc samples that have been reported in the research literature[21].

Second, it is possible that the strong association between patient-assessed severity and symptoms (e.g., pain, fatigue, severity of RP) occurred because both outcome and predictors were self-reported and the relationship reflects, to some degree, characteristics of the patient that influence how distress is reported on self-report questionnaires[22]. As a result, the relationships between outcome and predictors may be overstated in the models for patient-assessed severity reported in our study. On the other hand, there are currently no good substitutes for patient-reported symptoms such as pain, fatigue, and severity of RP, and this limitation is thus largely inevitable.

Finally, both samples of patients were composed of predominantly white, female patients with SSc. Consequently, this limits the generalizability of our results as far as patients with SSc from other ethnic groups or men are concerned.

The strength of our study lies in its large, multicenter sample of Canadian patients and validation of the results in an independent sample of patients with SSc.

We showed that patients and physicians rate SSc disease severity differently in magnitude and are influenced by different factors. Thus, patient-assessed and physician-assessed measures of severity should be considered as

complementary and should be used together in future studies of SSc.

APPENDIX

Investigators of the Canadian Scleroderma Research Group: M. Baron, Montreal, Quebec; J. Pope, London, Ontario; J. Markland, Saskatoon, Saskatchewan; D. Robinson, Winnipeg, Manitoba; N. Jones, Edmonton, Alberta; N. Khalidi, Hamilton, Ontario; P. Docherty, Moncton, New Brunswick; E. Kaminska, Hamilton, Ontario; A. Masetto, Sherbrooke, Quebec; D. Smith, Ottawa, Ontario; E. Sutton, Halifax, Nova Scotia; J-P. Mathieu, Montreal, Quebec; M. Hudson, Montreal, Quebec; S. Ligier, Montreal, Quebec; T. Grodzicky, Montreal, Quebec; S. Mittoo, Winnipeg, Manitoba; M. Fritzler, Advanced Diagnostics Laboratory, Calgary, Alberta.

REFERENCES

1. Yen JC, Neville C, Fortin PR. Discordance between patients and their physicians in the assessment of lupus disease activity: relevance for clinical trials. Lupus 1999;8:660-70.
2. Hanly JG, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. J Rheumatol 1996;23:1531-8.
3. Nicolau G, Yogui MM, Vallochi TL, Gianini RJ, Laurindo IM, Novaes GS. Sources of discrepancy in patient and physician global assessments of rheumatoid arthritis disease activity. J Rheumatol 2004;31:1293-6.
4. Neville C, Clarke AE, Joseph L, Belisle P, Ferland D, Fortin PR. Learning from discordance in patient and physician global assessments of systemic lupus erythematosus disease activity. J Rheumatol 2000;27:675-9.
5. Alarcon GS, McGwin G Jr, Brooks K, Roseman JM, Fessler BJ, Sanchez ML, et al. Systemic lupus erythematosus in three ethnic groups. XI. Sources of discrepancy in perception of disease activity: a comparison of physician and patient visual analog scale scores. Arthritis Rheum 2002;47:408-13.
6. Spoorenberg A, van Tubergen A, Landewe R, Dougados M, van der Linden S, Mielants H, et al. Measuring disease activity in ankylosing spondylitis: patient and physician have different perspectives. Rheumatology 2005;44:789-95.
7. Kucharz EJ, Grucka-Mamczar E, Mamczar A, Brzezinska-Wcislo L. Acute-phase proteins in patients with systemic sclerosis. Clin Rheumatol 2000;19:165-6.
8. Medsger TA Jr. Assessment of damage and activity in systemic sclerosis. Curr Opin Rheumatol 2000;12:545-8.
9. Medsger TA Jr, Silman AJ, Steen VD, Black CM, Akesson A, Bacon PA, et al. A disease severity scale for systemic sclerosis: development and testing. J Rheumatol 1999;26:2159-67.
10. Van Tubergen A, Debats I, Ryser L, Londoño J, Burgos-Vargas R, Cardiel MH, et al. Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. Arthritis Rheum 2002;47:242-8.
11. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
12. van Gestel A, Haagsma C, van Riel P. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. Arthritis Rheum 1998;41:1845-50.
13. Steen VD, Medsger TA Jr. 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.
14. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
15. Ware JE Jr, Sherbourne CD. The MOS 36 item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
16. Ware J, Kosinski M, Bjorner J, Turner-Bowker D, Gandek B, Maruish M. User's manual for the SF-36v2 health survey. 2nd ed. Lincoln, RI, USA: QualityMetric Inc.; 2007.
17. Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. Arthritis Rheum 2007;57:429-39.
18. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2008.
19. Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc 1995; 90:773-95.
20. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum 2002;46:2410-20.
21. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum 2008;37:223-35.
22. Meehl P. Why summaries of research on psychological theories are often uninterpretable. Psychol Rep 1990;66:195-244.