

Determinants of Discordance Between Patients and Physicians in Their Assessment of Lupus Disease Activity

JIM C. YEN, MICHAL ABRAHAMOWICZ, PATRICIA L. DOBKIN, ANN E. CLARKE, RENALDO N. BATTISTA, and PAUL R. FORTIN

ABSTRACT. Objective. We evaluated different methods for quantifying patient-physician discordance and identified factors associated with discordance in the assessment of lupus disease activity.

Methods. Data from 208 female patients who had a comprehensive annual examination were extracted from the Montreal General Hospital Lupus Registry. Discordance was measured by the difference between the patient self-reported 10 cm visual analog scale (VAS) and the physician VAS for global disease activity (VASDIFF). Multiple linear regression was used to identify the correlates of discordance, e.g., SF-36™ scales, Systemic Lupus Activity Measure (SLAM) components, etc. Four regression models were estimated using: (1) all patients; (2) only patients who evaluated disease activity higher than their physician's assessment; (3) only patients who evaluated disease activity lower than their physician's assessment; and (4) all patients, with the absolute value of VASDIFF as the dependent variable.

Results. Of the 208 observations, 150 (72%) of the VASDIFF scores were within ± 2.5 cm on a 10 cm scale, indicating absence of marked discordance; 43 (20.7%) were from patients overscoring and 15 (7.2%) from patients underscoring their physician by at least 2.5 cm. Higher SF-36 role physical score, more bodily pain, and lower role emotional score in addition to the SLAM-skin component were independently associated with higher discordance. SF-36 social function and mental health scores as well as SLAM-neurological and kidney components were correlated with discordance in some subanalyses. Bodily pain was the most important variable for predicting "clinically relevant" discordance, followed by SLAM-skin and kidney components.

Conclusion. Discordance between patients and physicians may result from patients scoring their disease activity based on their psychological and physical well-being, whereas physicians score disease activity based on the clinical and physical signs and symptoms of lupus. (J Rheumatol 2003;30:1967-76)

Key Indexing Terms:

PATIENT-PHYSICIAN RELATIONSHIP DISCORDANCE LUPUS DISEASE ACTIVITY

Discordance between patients and physicians occurs when the patients and physicians assign different values to a health trait of interest¹. Discordance between patients and physicians has been documented in the assessment of the patient's desire for information², outcome of surgery³, disability, and quality of life⁴. Discordance can negatively

affect patient care, adherence to treatment, and health outcome of disease⁵.

Physicians may not be fully able to evaluate the effect of a systemic disease such as systemic lupus erythematosus (SLE) on their patient's quality of life⁶; further, patients may not be aware of clinically important signs of disease activity

From the Division of Clinical Epidemiology, The McGill University Health Centre; Department of Epidemiology and Biostatistics, McGill University; Faculty of Medicine, McGill University; Division of Clinical Immunology and Allergy, The McGill University Health Centre; Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé (AÉTMIS), Montreal, Quebec; and Division of Rheumatology, Toronto Western Hospital, University Health Network, Toronto, Ontario, Canada.

Dr. Abrahamowicz is a James McGill Professor and a Scientist of the Canadian Institutes of Health Research (CIHR). Dr. Dobkin is a Senior Scientist of the Fonds de la Recherche en Santé de Québec. Dr. Clarke is a Canadian Institutes of Health Research (CIHR) Investigator. Dr. Fortin is a Senior Research Scholar of The Arthritis Society and Director of Clinical Research, Arthritis Centre of Excellence, University Health Network.

J.C. Yen, BScPhm, PhD, Division of Clinical Epidemiology, The McGill

University Health Centre; M. Abrahamowicz, PhD, Division of Clinical Epidemiology, The McGill University Health Centre, Department of Epidemiology and Biostatistics, McGill University; P.L. Dobkin, PhD, Division of Clinical Epidemiology, The McGill University Health Centre, Faculty of Medicine, McGill University; A.E. Clarke, MD, MSc, Division of Clinical Epidemiology, Division of Clinical Immunology and Allergy, The McGill University Health Centre; R.N. Battista, MD, ScD, Department of Epidemiology and Biostatistics, McGill University, Agence d'Évaluation des Technologies et des Modes d'Intervention en Santé; P.R. Fortin, MD, MPH, Division of Rheumatology, Toronto Western Hospital, University Health Network.

Address reprint requests to Dr. P. Fortin, Division of Rheumatology, Room MP-10-304 Toronto Western Hospital, 399 Bathurst Street, Toronto, Ontario M5T 2S8, Canada.

Submitted June 25, 2002; revision accepted February 24, 2003.

(e.g., kidney involvement), thereby increasing the likelihood for discordance. Discordance is undesirable, as it may lead to patients neglecting to follow the treatment plan or to physicians misjudging the effect of disease on their patients, thereby increasing use of health services and costs. To prevent the negative influence of discordance on health outcomes, the magnitude of discordance between patients with SLE and their physicians must be measured in a valid manner, and determinants of discordance need to be identified.

Different scales are used by patients and physicians to measure health status in SLE. Traditionally, physicians complete disease activity and damage indices, whereas patients complete quality of life scales, rendering comparisons between patients' and physicians' ratings difficult. The visual analog scale (VAS) is a brief, simple means of measuring disease activity that can be used by both patients and physicians. Recent studies in SLE that included VAS for global assessment of disease activity completed by both patients and physicians revealed discordance between their ratings⁷⁻¹⁰. However, most of these studies used only correlations to compare the patients' and physicians' VAS scores⁷⁻⁹.

Neville, *et al*¹⁰ were the first to use the difference between patient and physician VAS scores, recorded during the same visit, as a measure of discordance in SLE. In 6% of the cases, the physician's score was greater than the patient's by at least 2.5 cm on a 10 cm VAS. In 16% of the cases, the patients scored higher than their physicians by at least 2.5 cm. In multivariable analyses the authors found that patients based their ratings on both physical and psychological well-being, whereas physicians relied mostly on physical indices¹⁰. Yet in order to avoid confusion between discordance where the patients scored higher activity than their physicians and discordance where physicians scored higher activity than patients, Neville, *et al* excluded cases where patients scored substantially lower disease activity than their physicians in the multivariable analyses, which may have affected both the results and their interpretation¹⁰. Further, they did not investigate potential effect modifiers¹⁰. It is possible that the effect of a given variable on discordance depends on patient characteristics. For example, ethnicity may modify the effect of bodily pain on discordance. In this study, we provide a more comprehensive analysis of different aspects of patient-physician discordance and the factors associated with it in SLE. The primary objectives of this study were (1) to determine the extent to which discordance exists between patient and physician assessed global disease activity in patients with SLE; and (2) to identify factors associated with discordance and investigate potential effect modifiers. The secondary objective was methodological: we explored the effect of using different methods to quantify discordance.

MATERIALS AND METHODS

Data source and study population. All data were extracted from the Lupus Registry maintained at the Montreal General Hospital. This registry

contains specific patient data and disease indices that were collected as part of a standard annual visit from patients attending the Lupus Clinics at the Montreal General Hospital. Between January 1, 1992, and December 31, 1999, 227 patients had at least one "annual" visit with a specialist at the Lupus Clinic. During the visit, a standard protocol was used to obtain informed consent and to collect data for the registry. Patients were asked to complete a set of questionnaires, including a Medical Outcomes Study Short Form-36 items (SF-36TM) Questionnaire¹¹. Patients also described the level of their overall disease activity over the past month on a 10 cm VAS, with the anchors 0 cm representing no activity and 10 cm representing the most activity. At the same time, the physician completed the Systemic Lupus Activity Measure (SLAM-R)⁷, the SLE Disease Activity Index (SLEDAI)¹², and the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SLICC/ACR SDI)¹³. Physicians also completed a VAS for global disease activity that was analogous to the patient VAS. The patients' and physicians' VAS assessments were completed independently and neither party had access to the other's score. This study focused on the first visit, where both the patient and the physician VAS scores for disease activity were recorded simultaneously.

It is known that the distribution of various psychological measures differs between men and women¹⁴. Further, disease can affect men and women differently¹⁵. For these reasons, we considered it not appropriate to pool the data from both sexes and to assume, *a priori*, that correlates of discordance are the same for women and men. Given that there were too few male patients (19 out of 227 subjects, 8.4%) to allow the modeling of interactions between sex and several other patient characteristics, or to carry out separate multivariable analyses on men only, the analyses were restricted to female patients.

Potential correlates of discordance. Patient demographic variables available from the Lupus Registry include age, duration of disease, education level, and ethnic background. Education was recoded as a categorical variable: up to grade 11 education (reference group), 12–15 years of education (junior college and some university), and 16 years or more of education (university education). Ethnicity was grouped into 3 categories: Caucasians (reference group), Blacks, and Asians and Others (including Natives). In SLE, several studies have shown that the patient-completed SF-36 scales measure domains that are different from the physician-measured disease activity (e.g., SLAM-R) and damage (SDI) scales^{6,16-18}. Therefore, we included the original eight SF-36 scales of physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health, as independent variables.

In the Lupus Registry, SLAM-R⁷ and SLEDAI¹² scores are recorded during the annual visit. Neville, *et al*¹⁰ used selected components (integument/skin, neurologic, kidney, constitution, and musculoskeletal) of the SLAM-R and found an association between kidney involvement and discordance. For model parsimony, we selected *a priori* the same SLAM-R components as Neville *et al.* (Appendix I), which were represented as binary variables (presence/absence of involvement of a specific organ/system). In addition to the relationships between discordance and each of the 5 selected components, we assessed the association with the total SLAM-R score analyzed as a continuous variable. Similarly, we tested the relationship between discordance and the SLEDAI, but due to the strong correlation between the SLAM-R and the SLEDAI ($r = 0.69$), these variables were never included in the same multivariable model. The SDI, a measure of disease damage, was investigated simultaneously with the SLAM-R or the SLEDAI because they measure different constructs^{7,16}. Any missing value for an independent variable was replaced by the mean score for that variable calculated from all visits of all the patients. A single missing SLAM-kidney component score was replaced by 0, as we assumed that there was no kidney involvement.

All data for the Lupus Registry were entered and maintained in Medlog¹⁹. Data were exported into SAS for statistical analyses²⁰.

Overview of analyses. First, the VAS scores for patients and physicians were examined separately using multiple linear regression to determine

which factors were associated with the patient or physician VAS disease activity scores. Next, discordance scores were calculated on a continuous scale as the difference between the patient's VAS and the physician's VAS for global disease activity (VASDIFF = patient VAS – physician VAS). Multiple linear regression was used to determine which patient demographics and disease related measures were important in explaining the patient-physician discordance. Potential effect modification was also investigated by testing *a priori* selected 2-way interactions. In addition, multiple logistic regression was performed to identify characteristics of patients who scored lower or higher disease activity than their physicians.

To reduce the risk of inflated type I error due to multiple testing, the first step of all regression analyses was to test the statistical significance of the entire multivariable model with k independent variables. This step relied on an overall F-test with k degrees of freedom (df) or a k -df chi-square likelihood ratio test, respectively, for multiple linear and logistic models. The statistical significance of the model was a necessary and sufficient condition to carry out subsequent tests of each of k independent variables, as such overall tests control the overall type I error²¹. However, to account for using several different regression models, the significance level of each overall test was reduced to 0.01.

Multiple linear regression analyses for the patient VAS and the physician VAS. To gain insight into the factors that may be important in rating global disease activity, we performed separate multiple linear regression analyses using either the patient's or the physician's VAS as the dependent variables. Independent variables included patient SF-36 scales, SLAM-R components, and total SDI. The final multivariable model was selected using a stepwise procedure and interaction terms were tested as described below.

Multiple linear regression analyses of discordance scores.

Representation of discordance. The main dependent variable in multiple linear regression was the difference in the VAS scores (crude discordance, VASDIFF). However, the interpretation of this VAS difference varies depending on whether its value is positive or negative. For example, a positive VAS difference (e.g., 1.5 cm) indicates that the patient scores more disease activity than the physician. An increase in the VAS difference by 0.5 cm means that discordance between the patient and physician increases, i.e., the new VAS difference would be 2.0 cm. However, if the initial VAS difference was negative, e.g., -1.5 cm, then an increase of 0.5 cm means that the discordance is actually *decreased* (VAS difference = -1.0 cm). Therefore, 3 different analyses of crude discordance were carried out: (1) all patients; (2) only patients who scored higher activity than their physicians (positive discordance; VASDIFF > 0); and (3) only patients who scored lower activity than their physicians (negative discordance; VASDIFF < 0), because we expected that determinants of more discordance might vary depending on whether the patient scored higher or lower disease activity than her physician. The fourth analysis included all patients, but the dependent variable was defined as the absolute discordance (ABSDIFF) to investigate the variables that correlated with the magnitude of discordance, regardless of its direction.

Multivariable model selection. To obtain a parsimonious model, stepwise model selection was used with $p < 0.15$ as the criterion for variable entry into the model and $p > 0.25$ for elimination. In subsequent models, patient demographic variables (age, disease duration, education, and ethnic background) were adjusted for regardless of their statistical significance, because these variables may have an effect on disease activity and patient-reported measurements²²⁻²⁷.

Potential effect modifiers. Plausible interaction terms were identified *a priori* and their statistical significance was tested. We expected that age could modify the effects of SF-36 mental health and general health scales on disease activity and discordance. Education was tested for interactions with the SF-36 bodily pain, mental health, and general health scales. Further, education could interact with the SLAM-kidney component, since it may affect patient comprehension of the impact of disease severity. Ethnicity was tested for interactions with the SLAM-kidney component and the SDI, as the influence of disease activity and damage could vary

differently across ethnic groups^{22,28,29}, as well as with SF-36 bodily pain, general health, and mental health^{23,27,30}.

A 2-step approach was used to identify statistically significant interactions. In step 1, the various interaction terms were tested one at a time in a reduced regression model that included only the 2 variables involved in the interaction. Multiple-partial tests were used to determine the joint significance of all interaction terms corresponding to categorical variables such as education and ethnicity. Only significant or marginally nonsignificant ($p < 0.3$) were included in the next phase of testing. Second, the interactions selected in step 1 were tested while adjusting for other covariates. All the covariates that were previously selected were forced into the multivariable model and stepwise selection was used to identify interaction terms that were significant at $\alpha = 0.10$.

Multiple logistic regression to predict the direction of discordance. In addition to the above analyses that used a continuous dependent variable representing the *magnitude* of discordance, multiple logistic regression was performed to investigate which factors influence the direction of patient-physician differences. For these analyses, the observations with VASDIFF = 0 were eliminated and the dependent variable was recoded as a binary (positive/negative discordance) variable. Thus, we modeled the probability that the physician scores higher activity than the patient versus the opposite. The independent variables and model selection criteria were the same as for multiple linear regression.

RESULTS

The dataset. The Lupus Registry contained 208 first-visit observations for female patients occurring between January 1, 1992, and December 31, 1999, that had both the patient VAS and the physician VAS recorded simultaneously. Figure 1 shows the distribution of the crude discordance for the 208 observations. Most patients and their physicians do not differ too greatly in their scores, as 72% of the absolute VAS differences were below 2.5 cm, which was considered a cutoff for clinically meaningful discordance by Neville, *et al*¹⁰. There were 43 (20.7%) cases where the patient scored a higher VAS than the physician by at least 2.5 cm and 16 (7.7%) cases where the physician scored higher disease activity than the patient by 2.5 cm or more. Interestingly, discordance > 6 cm occurred only in situations (6 patients, 2.9%) where the patient scored higher activity than the physician.

Of the 208 female patients, 11 (5%) scored the VAS exactly the same as the physicians. Among other patients, 129 (62%) patients scored higher activity than the physicians and 68 (33%) patients scored lower. Tables 1a and 1b show the univariate statistics for all 208 female patients and a comparison of these 2 groups. Generally, patients who scored higher activity than their physicians reported lower scores (poorer quality of life score) on all SF-36 scales, with the difference in the bodily pain scores being statistically significant (Table 1a), but had lower disease activity according to both SLAM-R ($p = 0.03$) and SLEDAI ($p = 0.07$).

Separate multivariable analyses of patient VAS and physician VAS. Multiple linear regression, using stepwise model selection, identified 5 variables as statistically significant determinants of the patient-completed disease activity VAS. Longer disease duration, less bodily pain, better general

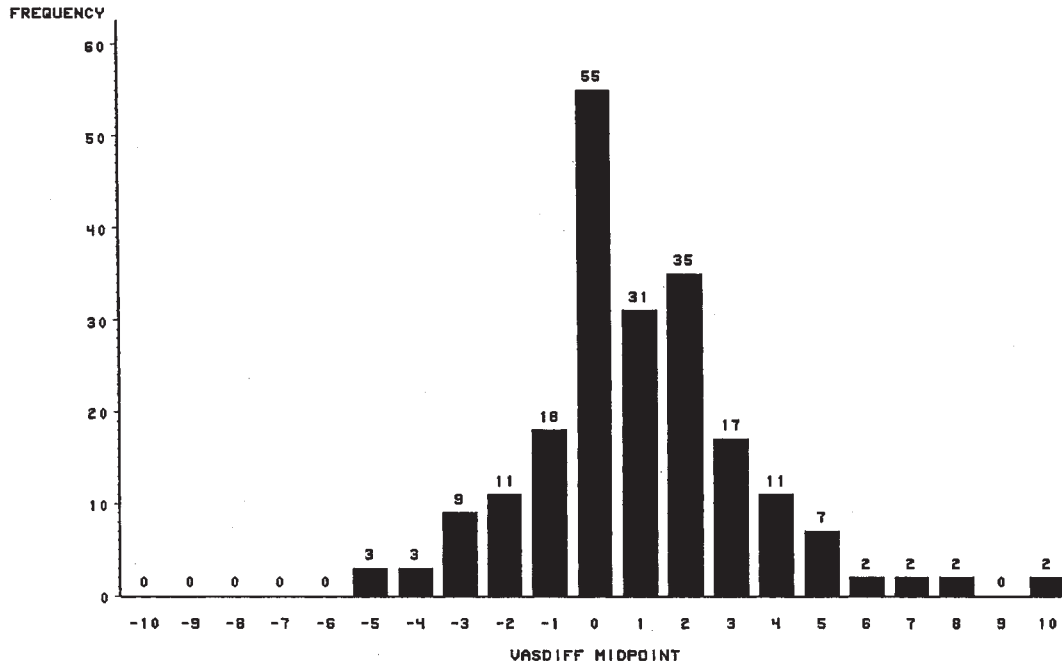


Figure 1. Distribution of VAS difference between patients and physicians for the first visits only (208 female subjects). There are actually 16 values where the VASDIFF discordance score was ≤ -2.5 cm. One value of exactly -2.5 was included in the -2 category by the graphing program. Crude discordance: VASDIFF = patient VAS – physician VAS for global disease activity.

Table 1A. Summary of continuous variables from first-visit data for female subjects from the Montreal General Hospital Lupus Registry. Description of all data and comparison of VASDIFF > 0 with VASDIFF < 0 subsets.

Variable	All VASDIFF, n = 208, mean (SD)	VASDIFF > 0 , n = 129, mean (SD)	VASDIFF < 0 , n = 68, mean (SD)	t Test ^a p
Patient VAS, cm	3.12 (2.68)	4.27 (2.62)	1.40 (1.51)	NA
Physician VAS, cm	2.22 (2.13)	2.05 (2.01)	2.88 (2.27)	NA
VASDIFF, cm	0.89 (2.43)	2.22 (1.94)	-1.48 (1.36)	NA
ABSDIFF, cm	1.86 (1.80)	2.22 (1.94)	1.48 (1.36)	NA
Age, yrs	42.00 (13.21)	42.83 (13.72)	39.85 (12.34)	0.1244
Disease duration, yrs	9.38 (8.52)	8.47 (7.75)	10.79 (9.79)	0.0927
SF-36 Scales				
Physical function	66.82 (26.34)	63.70 (27.58)	69.90 (24.15)	0.1050
Role physical	49.24 (43.49) ^b	44.96 (43.00)	50.62 (44.10) ^b	0.3855
Bodily pain	59.96 (27.25)	54.70 (26.33)	67.19 (27.15)	0.0023
General health	52.93 (22.42) ^c	50.29 (22.20)	54.80 (21.82) ^b	0.1720
Vitality	49.58 (22.68)	46.33 (23.13)	52.08 (20.93)	0.0791
Social function	71.23 (25.73)	69.38 (26.96)	72.98 (24.20)	0.3420
Role emotional	67.48 (40.78) ^c	62.50 (42.04)	71.64 (39.03) ^b	0.1318
Mental Health	66.20 (19.79)	64.70 (21.06)	66.88 (17.85)	0.4444
SLAM-R	8.04 (5.11)	7.72 (4.87)	9.18 (5.59)	0.0720
	median = 7	median = 7	median = 8	
SLEDAI	6.48 (6.48) ^c	5.80 (5.73) ^d	8.12 (7.77) ^b	0.0337
	median = 4	median = 4	median = 6	
SDI	1.36 (1.93)	1.22 (1.57)	1.63 (2.50)	0.2240
	median = 1	median = 1	median = 1	

^a T test for comparing VASDIFF > 0 with VASDIFF < 0 ; ^b One missing value; ^c Two missing values; ^d Three missing values; ^e Four missing values. VAS: Visual analog scale for global disease activity; SLAM-R: Systemic Lupus Activity Measure - revised; SLEDAI: SLE Disease Activity Index; SDI: Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; VASDIFF: difference between patient VAS and physician VAS for global disease activity; ABSDIFF: absolute value of VASDIFF; VASDIFF > 0 : patient scores higher disease activity than physician; VASDIFF < 0 : patient scores lower disease activity than physician.

Table 1B. Summary of categorical variables for first-visit data for female subjects from the Montreal General Hospital Lupus Registry. Description of all data and comparison of VASDIFF > 0 with VASDIFF < 0 subjects.

Variable	All VASDIFF,		VASDIFF > 0,		VASDIFF < 0,		Chi-square ^{a,b} p
	n	%	n	%	n	%	
Education, yrs	192		117		64		
1–11	70	36.5	40	34.2	27	42.2	0.2866
12–15	69	35.9	42	35.9	24	37.5	0.8304
16+	53	27.6	35	29.9	13	20.3	0.1618
Ethnic origin	208		129		68		
Caucasian	164	78.9	100	77.5	56	82.4	0.4269
Black	20	9.62	14	10.9	5	7.4	0.6125 ^a
Asian, native/other	24	11.5	15	11.6	7	10.3	1.000
SLAM-R components ^c (% with involvement)	208		129		68		
Skin	114	54.8	71	55.0	40	58.8	0.6106
Neurological	70	33.7	47	36.4	22	32.4	0.5681
Kidney	82 ^d	39.6	45 ^d	35.2	33	48.5	0.0687
Constitution	148	71.2	94	72.9	49	72.1	0.9036
Musculoskeletal	134	64.4	85	65.9	42	61.8	0.5650

^a Chi-square statistic calculated for comparison of VASDIFF > 0 with VASDIFF < 0; ^b Fisher's exact test was used if the cell value was small (≤ 5); ^c Categories are not mutually exclusive, i.e., subjects can have multiple SLAM-R component involvement; ^d One missing value. VASDIFF: difference between patient and physician VAS for global disease activity; VASDIFF > 0: patient scores higher disease activity than physician; VASDIFF < 0: patient scores lower disease activity than physician.

health, and better role emotional were independently associated with lower patient-scored disease activity (Table 2). However, the associations were weak, e.g., a 20 point increase in the bodily pain scale (indicating considerably less pain) was associated with a decrease of the patient VAS by 0.54 cm, corresponding to only 20% of the sample standard deviation (SD = 2.6). The presence of skin (integument) involvement, recorded in SLAM-R, corresponded to an increase in the disease activity score by 0.7 cm. Overall, the 5 variables explained 29% of the total variance in patients' VAS scores (Table 2).

Similar multivariable analysis of physician-completed VAS identified 6 statistically significant variables ($p \leq 0.05$). Patient's age (older) and higher SF-36 role physical and general health scores were independently associated with lower physician VAS scores. A higher SF-36 mental health score was associated with a higher physician VAS, but the result was only marginally significant ($p < 0.10$). SLAM-skin, kidney, and constitution involvements were associated with higher physician-scored disease activity scores (Table 2). Together, these 6 variables explained a substantially higher proportion of the variance in physicians' VAS than the 5 correlates of the patient VAS (43% versus 29%). The above results did not change materially after adjusting for patients' age, disease duration, education, and ethnicity.

Regression analyses of VAS differences. Table 3 shows the results of the stepwise selection in 4 analyses that focused on the discordance between patient's and physician's

disease activity VAS scores. SF-36 role physical, bodily pain, role emotional, and SLAM-skin, as well as patient age and disease duration, were identified as independent correlates of crude VASDIFF in all patients (2nd column of Table 3). However, the interpretation of the parameters in this model is not straightforward. For example, the negative coefficient for bodily pain score indicates that less pain (i.e., a higher score) is associated with a lower VASDIFF, but this may mean less agreement or more agreement depending on whether the patient scores below or above the physician. To facilitate the interpretation of the results of the analyses using all VASDIFF scores, we arbitrarily selected the physician VAS score as a "reference" to which the patient VAS score is compared. Using this interpretation, between 2 patients with the same physician VAS score, a 10-point increase in the bodily pain scale is associated with a decrease in the patient's disease activity score by 0.27 cm on a 10 cm VAS, regardless of whether the original VASDIFF score was positive or negative. Older patients were more likely to score higher than their physicians, but with longer duration of disease, patients' scores tended to decrease, which may reflect "habituation" or "acceptance" of the disease's effect on the patient's health (2nd column of Table 3).

To avoid the above difficulty of interpreting the regression parameters for all VASDIFF, separate regression models were estimated for positive and negative VASDIFF. It was expected that correlates of discordance might be different for patients who scored higher than their physicians than for those patients who scored lower than their physicians. In the case where the patient scored higher

Table 2. Results of automated stepwise model selection for multivariable regression with patient VAS and physician VAS for global disease activity, separately, as the dependent variables.

Variable	Patient VAS, n = 208	Physician VAS, n = 208
Age		-0.0214**
Disease duration	-0.0505**	
Education		
Education-d1		
Education-d2		
Ethnicity		
Ethnicity-d1		
Ethnicity-d2		
SF-36		
Physical function		
Role physical		-0.0129**
Bodily pain	-0.0269**	
General health	-0.0184**	-0.0138**
Vitality		
Social function		
Role emotional	-0.00877**	
Mental health		0.0108*
SLAM-R		
Skin	0.732**	1.244**
Neurological		
Kidney		0.680**
Constitution		1.133**
Musculoskeletal		
SDI		
Adjusted R ²	0.2868	0.4271
Overall F test, p	< 0.0001	< 0.0001

* $0.05 < p \leq 0.10$. ** $p \leq 0.05$. Stepwise regression criteria for variables to enter model was $p \leq 0.15$ and criteria for staying in model was $p \leq 0.25$. Blank cells indicate that variable was not selected or included in the model. Education: reference category = 1–11 years, d1 = 12–15 years, d2 = ≥ 16 years of education. Ethnicity: reference category = Caucasians, d1 = Blacks, d2 = Asians, natives, or other. Example for interpretation of results: For Patient VAS, the parameter estimate for bodily pain is -0.0269. This means that for a 10 point increase (an improvement) in bodily pain score, the patient VAS for disease activity decreases by 0.27 cm, given all other variables being equal. For Physician VAS, the parameter for SLAM-skin is 1.244. Thus, involvement of the skin component according to the SLAM-R is associated with an increase of 1.244 cm in the Physician VAS for disease activity, all other variables being equal.

disease activity than the physician (VASDIFF > 0), only SLAM-kidney ($0.05 < p < 0.10$) was selected as a correlate of higher discordance, after adjusting for patient demographic characteristics (3rd column of Table 3). However, no variables in the model were statistically significant at the 0.05 level, suggesting that the amount by which the patient over-rated disease activity, compared to the physicians, does not depend systematically on any of the patient's characteristics considered.

On the other hand, the regression model for situations where the physician scores higher disease activity than the patient (VASDIFF < 0) explained 44% of the variance, and several statistically significant correlates of the amount of discordance were identified. Better social functioning was associated with less discordance, and better mental health

with higher discordance. Ethnicity was a significant predictor: on average, Black patients tended to under-rate their disease activity by about 1 cm more than Caucasians. Similarly, involvement of neurological or kidney components was associated with more marked underestimation of disease activity by patients compared to their physicians, suggesting that patients might be unaware of the importance of some of their underlying problems.

In the analysis focusing on the amount of discordance regardless of whether the patient or physician scores higher (ABSDIFF), only higher level of bodily pain and the presence of kidney involvement were associated with an increase in discordance (last column of Table 3).

Investigation of potential effect modifiers. Of 11 clinically plausible interactions, only 4 were selected into the multivariable models (data not shown). For crude discordance scores, ethnicity modified the effects of general health and SLAM-kidney. In general, kidney involvement tended to lower the discordance, except for Blacks, where those with kidney involvement had higher VASDIFF. For Blacks, Asians, and others, discordance tended to increase as the general health status improved: this increase was more evident for Asians and others. In contrast, for Caucasians, discordance decreased as the general health improved.

Among patients who scored lower than their physicians, education was found to modify the effect of general health. For patients with junior college education there was a gradual decrease in discordance as the general health score increased, whereas for those with university level (≥ 16 years) education, better general health was associated with higher discordance. A similar pattern was seen in the model for absolute discordance (ABSDIFF), with the exception that education modified the effect of mental rather than general health and in the opposite direction. For example, for patients with junior college education, discordance increased with better mental health. Finally, the effect of kidney involvement on increased absolute discordance in Asians or others, and especially in Blacks, was significantly stronger than in Caucasians.

Sensitivity analyses using logistic regression. We used logistic regression to explore which factors influenced whether the patient scored lower disease activity than the physician (VASDIFF < 0). In the univariate regressions, lower level of bodily pain and higher SLEDAI scores ($p < 0.05$), as well as higher vitality score and kidney involvement ($p < 0.10$) were all associated with an increased likelihood that the patient would score lower activity than her physician (data not shown). After adjusting for age, disease duration, education level, and ethnicity in the multivariable logistic regression, the variables that entered the model were role physical, bodily pain, and SLAM-skin (data not shown). An increase of 10 points on the SF-36 role physical score (i.e., less role limitation due to physical involvement) was associated with an 8% decrease in the probability that

Table 3. Comparison of factors associated with different definitions of patient-physician discordance at the time of first visit using stepwise multivariable linear regression, adjusted for age, disease duration, education, and ethnicity.

Variable	VASDIFF, n = 208	VASDIFF > 0, n = 129	VASDIFF < 0, n = 68	ABSDIFF, n = 208
Age	0.0365**	0.0163	-0.00799	0.0121
Disease duration	-0.0479**	-0.0263	-0.00059	-0.0201
Education				
Education-d1	0.304	-0.516	0.225	-0.278
Education-d2	0.616	-0.164	-0.338	-0.0635
Ethnicity			Prob > F = 0.0775	
Ethnicity-d1	0.338	0.0692	-1.055**	0.528
Ethnicity-d2	0.193	0.633	-0.574	0.517
SF-36				
Physical function				
Role physical	0.0151**			
Bodily pain	-0.0268**			-0.0104**
General health				
Vitality				
Social function			0.0200**	
Role emotional	-0.0124**			
Mental health			-0.0287**	
SLAM-R				
Skin	-0.765		-0.469	
Neurological			-0.937**	
Kidney		0.576	-0.591**	0.554**
Constitution				
Musculoskeletal				
SDI				
Adjusted R ² ,	0.1002	-0.0046	0.4384	0.0390
Overall F test, p	0.0005	0.4968	< 0.0001	0.0423

* 0.05 < p ≤ 0.10. ** p ≤ 0.05. Stepwise regression criteria for variables to enter model was p ≤ 0.15 and criteria for staying in model was p ≤ 0.25. Multiple-partial tests (probability > F) for dummy variables representing categorical variables were not statistically significant unless otherwise noted. Blank cells in table indicate that variable was not selected or included in the model. Education: reference category = 1–11 years, d1 = 12–15 years, d2 = ≥ 16 years of education. Ethnicity: reference category = Caucasians, d1 = Blacks, d2 = Asians, natives, or other. VASDIFF: difference between patient and physician VAS for global disease activity; VASDIFF > 0: patient scores higher disease activity than physician; VASDIFF < 0: patient scores lower disease activity than physician. ABSDIFF: absolute difference between VAS scores.

the patient scored less disease activity than the physician (OR 0.92, 95% CI 0.91, 0.93), whereas a 10 point decrease in bodily pain was associated with a 34% increase in the probability that patient under-scored disease activity compared to the physician (OR 1.34, 95% CI 1.32, 1.37). Finally, the presence of skin manifestations was associated with 77% greater odds that the patient under-rated her disease activity compared to her physician (OR 1.77, 95% CI 0.88, 3.54).

DISCUSSION

Understanding the reasons for patient-physician discordance is important because discordance may affect patient care, adherence to treatment, and outcomes. Yet discordance between patients and physicians can be defined in different ways and operationalization of the construct may affect the results. We analyzed several approaches to representing discordance, based on the difference between the patient

VAS and the physician VAS scores for global disease activity. Through multivariable analysis of all VAS difference scores, measured on a continuous scale, we found that the patient-scored SF-36 role physical, bodily pain, and role emotional scales played a role in explaining the extent to which the patient and physician scores diverged. Further, the physician-scored skin component of the SLAM-R was also an important explanatory variable. Nonetheless, when we restricted the analyses to the subgroup of patients who scored higher activity than their physicians, we found that the magnitude of discordance could not be adequately explained by any of the characteristics we explored. On the other hand, in the cases of the physicians reporting higher disease activity than the patients, almost 50% of the total variance in the amount of discordance could be explained by social function and mental health, as well as SLAM-skin, neurological, and kidney.

The variable that was most systematically associated

with discordance was bodily pain, with more pain being associated with higher discordance between patients' and physicians' perceptions. Pain is a subjective experience and physicians may not fully appreciate the effect it has on their patients. Indeed, bodily pain alone explained 20% of the variance for the patient disease activity score, but it was not identified as a significant predictor of physician scores (Table 2). The skin and kidney components of the SLAM-R are important physician-scored variables associated with discordance. Kidney involvement was not an independent predictor of the patient VAS score (Table 2), suggesting that patients may not be aware of the presence of kidney involvement or its importance in disease activity.

The SF-36 scales of role physical, social function, and role emotional, as well as SLAM-R neurological and constitution components also played a role in explaining some aspects of discordance. We found that patient sociodemographic variables need to be adjusted for in all multivariable analyses. The LUMINA (Lupus in Minority Populations, Nature versus Nurture) studies, and studies by other investigators, have noted that lupus has effects that differ across various groups of lupus patients^{22-24,28,29}. We found that the impact of organ involvement on discordance varied across ethnic groups.

The patient and physician measures used in this study were recorded at the same point in time, which avoids problems of variability in disease activity that can change rapidly in a short span of time³¹. Neville, *et al*¹⁰ first reported the use of patient-physician difference as a measure for discordance in SLE and found that the SF-36 mental health and the SLAM-kidney component were associated with discordance. While our patient population originated from the same source as Neville's study, the methodology was different, as are the results. We had a larger sample size and restricted the analyses to female patients. Further, we used various strategies to model the factors associated with various aspects of discordance. We also performed various sensitivity analyses to verify our findings. This allowed us to identify different correlates of discordance. For example, we found that bodily pain was an important correlate regardless of the definition of discordance or the statistical analysis used. On the other hand, we found that kidney involvement was mostly associated with important difference and was associated with the physician scoring higher activity than the patient. We also found that SLAM-skin was important, in contrast to Neville, *et al*¹⁰, who found no statistically significant association between skin involvement and discordance. Although both patients and physicians consider the involvement of skin in their assessments, it affected the physicians' score more than the patients'. This indicates the need to educate patients on the importance of skin disease in SLE³². In addition, we tested for effect modifiers and found some potentially important and statistically significant interactions, such as between education and

general health, and between ethnicity and kidney involvement.

There are a limited number of studies that used the difference of scores between patients and physicians as a measure of discordance in SLE. Aranow, *et al*³³ presented an abstract comparing patient versus physician VAS in flare assessment. They found that agreement was more likely to occur when the patient believed they were in an active disease state or when the physician believed the disease was in remission. A few studies examined the correlation between the patient VAS and physician VAS⁶⁻⁹, and found that it varied from as low as 0.37 to as high as 0.91, but none examined which patient characteristics influenced discordance.

To aid interpretation of the results, we chose the physician VAS as a frame of reference when comparing the patient and physician VAS. This is not meant to imply a right or wrong answer¹ — both the patient and physician VAS assess their respective points of view. The physician VAS showed less variability (SD 2.13) than the patient VAS (SD 2.68). When a patient scores the disease activity, she is basing it on her personal experience/perception, whereas the physician scores are based on the sum of his/her practice experience as well as on report from this particular patient. On the other hand, even though the physician experience may be broader, discordance still exists, since the physician cannot completely understand and precisely value how the patient feels at the time of the evaluation.

The OMERACT IV meeting dealt with identifying core sets of outcome domains for epidemiologic studies in SLE and recommended consideration of other domains in addition to patient health-related quality of life, disease activity, disease damage indices, and toxicity/adverse events³⁴. Our study supports the need to investigate these other domains, since simply including the SF-36, SLAM-R/SLEDAI, and SDI could not fully explain the variance in discordance. Measures of physical disability, psychosocial factors (e.g., psychological distress), comorbidity, and fatigue may be useful in understanding the discordance between patients and physicians. Clearly, more work is needed to understand what disease activity means to patients and physicians and how these interpretations translate into the way they score disease activity.

Some limitations of this study should be noted. Given that we used patient data that were collected in a registry, a lack of information on patients who choose not to enter the Lupus Registry may induce some participation bias. We investigated only the first visit that was recorded in the registry, and therefore we are unable to assess how discordance may vary when the patient and physician develop a closer relationship over time. However, we did adjust for disease duration, which may affect how patients view their diseases over time⁶. We also did not have an explicit measure of the patient's socioeconomic status; therefore we

could only adjust for the patient's education level and their ethnicity. Clearly, there are other possible correlates of discordance that could have been investigated in this exploratory analysis. For example, we did not have data on physician factors such as sex, years of practice, etc. Also, because data were collected in a tertiary care setting, the results may not be representative of the general lupus population. There were some missing values that we arbitrarily replaced using the group average (see footnotes in Table 1). However, this should not materially affect our results, since the proportion of missing values was small (maximum of 4 values, i.e., 1.9%); if at all, this would only have biased the association toward the null.

This study provides insights into patients' characteristics that are associated with different aspects of discordance in SLE at a single point in time. Future research in discordance should clearly state the definition used for discordance, as different criteria for discordance can affect the results. Future studies could follow the patients over time and examine if discordance changes as the patient-physician relationship develops. Also, future studies could add measures of physical disability, psychological distress, stress and coping mechanisms, comorbidity, and fatigue, as these variables may be important in understanding how patients are affected by their disease. Discordance can affect patient satisfaction^{35,36}, and this can influence adherence to treatment regimens and use of alternative therapies. Research into how discordance may affect patient care and outcome is needed. In addition, the framework that we used in this study for identifying different aspects of discordance may be applied to studies of other diseases, which, in turn, may help us to better understand patient-physician discordance in SLE.

In summary, discordance exists between patients' and physicians' assessments of disease activity in SLE because their assessments are based on different experiences. Patients that perceive their disease activity as lower than their physician may adhere less to their treatments and this may lead to recurrent flares and poor overall control of their lupus. On the other hand, patients that perceive their disease activity as higher than their physician may seek nontraditional treatments that may not be appropriate and may be costly. Identification of the factors associated with discordance will allow development of interventions that can be evaluated to test if decreasing patient-physician discordance is associated with better adherence to treatment and a rational use of complementary medicines. Clinicians need to be aware of the possibility of discordance between themselves and their patients, and they should also be sensitive to patients' complaints of physical symptoms, since this appears to influence their patients' perceptions of disease. Some patients need help in learning to cope better with their illness as well as understanding the importance of organ involvement in their condition; this may be achieved

through a multidisciplinary approach and good patient-physician communication. These recommendations apply equally well to other chronic medical conditions.

APPENDIX

SLAM-R Components

All components were scored 0 for not present and 1 for present

Skin/Integument

- Presence of oral/nasal ulcers or rash (SLAM-R question 4)
- And/or presence of alopecia (question 5)
- And/or presence of rash or lesions (question 6)
- And/or presence of vasculitis (question 7)

Neurological

- Presence of stroke syndrome (SLAM-R question 18)
- And/or presence of seizure (question 19)
- And/or presence of cortical dysfunction (question 20)
- And/or presence of headache (question 21)
- And/or presence of myalgia/myositis (question 22)

Kidney

- Presence of increased serum creatinine (question 30)
- And/or presence of urine sediment abnormality (question 31)

Constitutional

- Presence of weight loss (question 1)
- And/or presence of fatigue (question 2)
- And/or presence of fever (question 3)

Musculoskeletal

- Presence of myalgia/myositis (question 22)
- And/or presence of joint pain or swelling (question 23)

Note: SLAM-R question 22 appears twice in this scoring system; it is measured in both the Neurological component and the Musculoskeletal component.

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. Waitzkin H. Doctor-patient communication. Clinical implications of social scientific research. *JAMA* 1984;252:2441-6.
3. Lieberman JR, Dorey F, Shekelle P, et al. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *J Bone Joint Surg Am* 1996;78:835-8.
4. Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ* 1997;314:1580-3.
5. Starfield B, Wray C, Hess K, Gross R, Birk PS, D'Lugoff BC. The influence of patient-practitioner agreement on outcome of care. *Am J Public Health* 1981;71:127-31.
6. Fortin PR, Abrahamowicz M, Neville C, et al. Impact of disease activity and cumulative damage on the health of lupus patients. *Lupus* 1998;7:101-7.
7. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in SLE. *Arthritis Rheum* 1989;32:1107-18.
8. Morand EF, Miller MH, Whittingham S, Littlejohn GO. Fibromyalgia syndrome and disease activity in systemic lupus

- erythematosus. *Lupus* 1994;3:187-91.
9. Fortin PR, Abrahamowicz M, Danoff D. Small changes in out-patient lupus activity are better detected by clinical instruments than by laboratory tests. *J Rheumatol* 1995;22:2078-83.
 10. 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.
 11. 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.
 12. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH, The Committee on Prognosis Studies in SLE. Derivation of the SLEDAI. A disease activity index for lupus patients. *Arthritis Rheum* 1992;35:630-40.
 13. Gladman DD, Ginzler E, Goldsmith C, et al. The development and initial validation of the SLICC/ACR damage index for SLE. *Arthritis Rheum* 1996;39:363-9.
 14. Hopman WM, Towheed T, Anastassiades T, et al. Canadian normative data for the SF-36 health survey. Canadian Multicentre Osteoporosis Study Research Group. *CMAJ* 2000;163:265-71.
 15. Swaak AJG, Nossent JC, Bronsveld W, et al. Systemic lupus erythematosus. I. Outcome and survival: Dutch experience with 110 patients studied prospectively. *Ann Rheum Dis* 1989;48:447-54.
 16. Hanly JG. Disease activity, cumulative damage and quality of life in systemic lupus erythematosus: results of a cross-sectional study. *Lupus* 1997;6:243-7.
 17. Stoll T, Gordon C, Seifert B, et al. Consistency and validity of patient administered assessment of quality of life by the MOS SF-36; its association with disease activity and damage in patients with systemic lupus erythematosus. *J Rheumatol* 1997;24:1608-14.
 18. Stoll T, Stucki G, Malik J, Pyke S, Isenberg DA. Association of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index with measures of disease activity and health status in patients with systemic lupus erythematosus. *J Rheumatol* 1997;24:309-13.
 19. Medlog Systems. [Computer program]. Incline Village, NV: Medlog Systems; 2000.
 20. SAS for Windows. [Statistical program]. Cary, NC: SAS Institute; 1999.
 21. Kleinbaum DG, Kupper L, Muller KE, Nizam A. Applied regression analysis and multivariable method. 3rd ed. Pacific Grove, CA: Cole Publishing Company; 1998.
 22. Alarcon GS, Roseman J, Bartolucci AA, et al. Systemic lupus erythematosus in three ethnic groups: II. Features predictive of disease activity early in its course. LUMINA Study Group. *Arthritis Rheum* 1998;41:1173-80.
 23. Friedman AW, Alarcon GS, McGwin G Jr, et al. Systemic lupus erythematosus in three ethnic groups. IV. Factors associated with self-reported functional outcome in a large cohort study. LUMINA Study Group. *Arthritis Care Res* 1999;12:256-66.
 24. Alarcon GS, Rodriguez JL, Benavides G Jr, Brooks K, Kurusz H, Reveille JD. Systemic lupus erythematosus in three ethnic groups. V. Acculturation, health-related attitudes and behaviors, and disease activity in Hispanic patients from the LUMINA cohort. LUMINA Study Group. *Arthritis Care Res* 1999;12:267-76.
 25. Rivest C, Lew RA, Welsing PM, et al. Association between clinical factors, socioeconomic status, and organ damage in recent onset systemic lupus erythematosus. *J Rheumatol* 2000;27:680-4.
 26. Gilboe IM, Kvien TK, Husby G. Disease course in systemic lupus erythematosus: changes in health status, disease activity, and organ damage after 2 years. *J Rheumatol* 2001;28:266-74.
 27. Sternfeld B, Cauley J, Harlow S, Liu G, Lee M. Assessment of physical activity with a single global question in a large, multiethnic sample of midlife women. *Am J Epidemiol* 2000;152:678-87.
 28. Waters TM, Chang RW, Worsdall E, Ramsey-Goldman R. Ethnicity and access to care in systemic lupus erythematosus. *Arthritis Care Res* 1996;9:492-500.
 29. Reveille JD, Moulds JM, Ahn C, et al. Systemic lupus erythematosus in three ethnic groups: I. The effects of HLA class II, C4, and CR1 alleles, socioeconomic factors, and ethnicity at disease onset. LUMINA Study Group. *Arthritis Rheum* 1998;41:1161-72.
 30. Andresen EM, Brownson RC. Disability and health status: ethnic differences among women in the United States. *J Epidemiol Community Health* 2000;54:200-6.
 31. Petri M, Genovese M, Engle E, Hochberg MC. Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. *Arthritis Rheum* 1991;34:937-44.
 32. Tebbe B, Orfanos CE. Epidemiology and socioeconomic impact of skin disease in lupus erythematosus. *Lupus* 1997;6:96-104.
 33. Aranow C, Soliman N, Ginzler E. Lupus flares: impressions by patients and physicians [abstract]. *Arthritis Rheum* 1997;40 Suppl:S208.
 34. Smolen JS, Strand V, Cardiel M, et al. Randomized clinical trials and longitudinal observational studies in systemic lupus erythematosus: consensus on a preliminary core set of outcome domains. *J Rheumatol* 1999;26:504-7.
 35. Sutherland HJ, Lockwood GA, Minkin S, Trichler DL, Till JE, Llewellyn-Thomas HA. Measuring satisfaction with health care: a comparison of single with paired rating strategies. *Soc Sci Med* 1989;28:53-8.
 36. Wartman SA, Morlock LL, Malitz FE, Palm E. Impact of divergent evaluations by physicians and patients of patients' complaints. *Public Health Rep* 1983;98:141-5.