Minimal Clinically Important Difference for 7 Measures of Fatigue in Patients with Systemic Lupus Erythematosus

EWAN C. GOLIGHER, JACQUES POUCHOT, ROLLIN BRANT, RAHEEM B. KHERANI, J. ANTONIO AVIÑA-ZUBIETA, DIANE LACAILLE, ALLEN J. LEHMAN, STEPHANIE ENSWORTH, JACEK KOPEC, JOHN M. ESDAILE, and MATTHEW H. LIANG

ABSTRACT. ***Objective.*** To determine the minimal clinically important difference (MCID) for 7 measures of fatigue in patients with systemic lupus erythematosus (SLE).

Methods. Study subjects completed 7 fatigue instruments [Fatigue Severity Scale (FSS), Multidimensional Assessment of Fatigue (MAF), Multidimensional Fatigue Inventory (MFI), Vitality scale of the MOS-SF-36, Chalder Fatigue Scale (CFS), Functional Assessment of Chronic Illness Therapy-Fatigue, and a global Rating Scale (RS)] and then participated in a series of interviews with other study participants comparing their fatigue with one another. Each interview participant rated the difference in their fatigue levels on a 7-point transition scale. The MCID was estimated from the mean difference in fatigue scores between each pair of interview participants based on their subjective rating of fatigue contrast. The MCID was also estimated using linear regression modeling.

Results. Eighty patients with SLE participated. Patients reported significant levels of fatigue [mean normalized (0 = none, 100 = maximum) fatigue scores for the 7 instruments ranged from 49.8 (CFS) to 71.1 (FSS)]. The MCID of "a little more" fatigue tended to be greater than the MCID for a "little less fatigue" and differed significantly for FSS and MAF. The MCID of normalized scores estimated by linear regression ranged from 7.0 (CFS) to 14.3 (MFI).

Conclusion. Fatigue is a common and debilitating component of SLE. Estimates of MCID will help to interpret changes observed in a fatigue score and will be critical in estimating sample size requirements for clinical trials including fatigue as an outcome. (First Release Mar 1 2008; J Rheumatol 2008;35:635–42)

Key Indexing Terms:

SYSTEMIC LUPUS ERYTHEMATOSUS
MINIMAL CLINICALLY IMPORTANT DIFFERENCE
SAMPLE SIZE REQUIREMENT
RESPONSIVENESS
QUESTIONNAIRE
FATIGUE

Patients with systemic lupus erythematosus (SLE) report that fatigue is a common and debilitating symptom of their illness. The majority of patients experience significant levels of fatigue[1]. The presence and severity of fatigue have been ascribed to a variety of causes including disease activity, lack of aerobic fitness, psychological depression, sleep

From the Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; Department of Internal Medicine, Hôpital Européen Georges Pompidou, Université René Descartes – Paris V, Paris, France; Arthritis Research Centre of Canada, Division of Rheumatology, University of British Columbia; Centre for Community Child Health Research, Child and Family Research Institute, University of British Columbia; Department of Educational Studies, University of British Columbia; and Department of Health Care and Epidemiology, University of British Columbia, Vancouver, BC, Canada.

Supported in part by NIH grant AR47782 and a grant from Rheuminations, Inc. Dr. Goligher was recipient of a Kirkland Summer Student Award. Dr. Aviña-Zubieta holds doctoral fellowship awards from the Canadian Institutes of Health Research, the Michael Smith Foundation for Health Research, The Canadian Arthritis Network, CONACyT-Mexico, and receives support from a Kirkland Scholar Award. Dr. Lacaille holds an Investigator Award from The Arthritis Society. Mr. Lehman was supported by doctoral fellowship awards from the Canadian Institutes of Health Research, the Michael Smith Foundation for Health Research, and the Pfizer Research Fellowship in Arthritis. Dr. Kopec is a Senior Scholar of the Michael Smith Foundation for Health Research.

Dr. Esdaile is a Kirkland Scholar. Dr. Liang is the Molson Foundation Arthritis Scholar.

E.C. Goligher, MD, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School; J. Pouchot, MD, Arthritis Research Centre of Canada and Department of Internal Medicine, Hôpital Européen Georges Pompidou, Université René Descartes – Paris V; R. Brant, PhD, Centre for Community Child Health Research, Child and Family Research Institute; R.B. Kherani, MD; J.A. Aviña-Zubieta MD, MSc; D. Lacaille, MD, MHSc; S. Ensworth, MD; J.M. Esdaile, MD, MPH, Arthritis Research Centre of Canada; A.J. Lehman, MA, Arthritis Research Centre of Canada and Department of Educational Studies, University of BC; J. Kopec, MD, PhD, Arthritis Research Centre of Canada and Department of Health Care and Epidemiology, University of BC; M.H. Liang, MD, MPH, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School and Arthritis Research Centre of Canada.

Address reprint requests to Prof. J. Pouchot, Service de Médecine Interne, Hôpital Européen Georges Pompidou, 20 rue Leblanc, 75908 Paris Cedex 15, France. E-mail: jacques.pouchot@egp.aphp.fr

Accepted for publication November 28, 2007.

disturbances, and comorbid medical conditions such as fibromyalgia[2-6]. While the etiology of fatigue is not always clear in the individual patient, it is associated with reduced quality of life[7] and is often difficult to manage.

Including fatigue as an outcome in clinical studies of SLE requires standardized measures. Several instruments have been developed to assess different aspects of fatigue[8-11], and some are already in use in SLE[8,10]. The psychometric properties of these fatigue instruments have been studied to varying degrees. Knowledge of such properties is essential to their application and interpretation in clinical studies of disease-related fatigue.

One such important but understudied psychometric property is the minimal clinically important difference (MCID), which is defined as "the smallest difference in score in the domain of interest (i.e., fatigue) which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"[12]. It establishes how much change in an outcome variable is necessary to provide a noticeable improvement or worsening in disease outcome, whether objective or subjective. It provides an estimate of the required effect size for clinical studies and facilitates sample size and statistical power calculations. Also, the MCID estimation is important to interpret the magnitude of longitudinal changes for individual patients during routine care or differences when comparing therapeutic strategies. We estimated the MCID of 7 measures of fatigue using a cross-sectional approach described by Redelmeier and Lorig[13] and employed an innovative statistical approach[14] to refine the results.

MATERIALS AND METHODS

Patients. The study was conducted at the Mary Pack Arthritis Centre, Vancouver, BC, and at the Robert B. Brigham Musculoskeletal Sciences Center, Brigham and Women's Hospital, Boston, MA. Approval for the study was obtained from the institutional review boards of the University of BC, the Mary Pack Arthritis Centre, and Brigham and Women's Hospital.

Patients were eligible to participate if they had SLE as defined by the American College of Rheumatology (ACR) criteria[15] and were older than age 18 years. Patients were excluded if they were admitted to hospital, were unable to complete self-administered questionnaires, or were unable to read or converse in English.

In Vancouver, patients with SLE were initially contacted from patient lists of rheumatologists affiliated with the Mary Pack Arthritis Centre, and volunteers were enrolled over a 3-month period (April to June 2004). In Boston, patients were recruited by advertisement with the local Lupus Foundation in Boston. Patients were enrolled between April and November 2004. A total of 80 patients with SLE were recruited. Mean age, disease duration, and disease activity were in the same range for both patient populations; 61% were Caucasian.

Questionnaires. We used 7 validated self-administered fatigue instruments. Six fatigue questionnaires were selected from an extensive literature search. These included the Fatigue Severity Scale (FSS)[8,16], the Vitality scale (VT) of the Medical Outcome Study 36-item short-form health survey (SF-36)[17], the Multidimensional Assessment of Fatigue (MAF)[18], the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F)[9], the Chalder Fatigue Scale (CFS)[10], and the Multidimensional Fatigue Inventory (MFI)[11]. Selection was based on the content of the instrument, documented psychometric validity, availability in English, previous use in inflammatory rheumatic diseases and especially in SLE, the ability to be self-administered, and the number of items. Pertinent characteristics of the questionnaires are summarized in Table 1. Not all the selected questionnaires had previously been used in studies of SLE. In addition to these 6 questionnaires, we employed a global assessment of fatigue using an 11-point numerical rating scale (RS); the anchors were 0 = "no fatigue at all" and 10 = "fatigue as bad as it could be." Because of the cross-sectional study design, each questionnaire was standardized to refer to the previous week's experience, even if the original version used a different timeframe.

Data collection. Study participants were consecutively divided into groups of 6 to 8 people. Each group met together in a single session. During this session, patients completed the 6 fatigue questionnaires and a global fatigue severity rating scale. In addition, participants completed the Systemic Lupus Activity Questionnaire (SLAQ)[19] and a global self-assessment of their disease activity on an 11-point numerical rating scale (0 = "no activity" and 10 = "the most activity"). The SLAQ is a validated self-administered questionnaire that has been developed to assess SLE activity through 24 clinical items derived from the Systemic Lupus Activity Measure[20] and is amenable to self-report[19]. Its score ranges from 0 (no activity) to 47 (maximum activity)[19].

Immediately after completing the questionnaires, each member of the group met successively with 5 other group members in a series of individual conversations. Participants were instructed to discuss issues and concerns they felt were important with respect to their fatigue, focusing on their experience over the preceding week to be consistent with the frame of reference of the fatigue instruments. Specifically, they were encouraged to ask one another, "Are you easily fatigued?", "Do you have a lot of energy?", "Does fatigue interfere with your activities — family, social, etc.?". The one-on-one conversations lasted for 10 minutes. At the end of the conversation, participants confidentially rated their fatigue in contrast to their impression of their conversation-partner's fatigue. The fatigue contrast was rated using a single-item Likert scale with 7 response categories: "Much more fatigue," "Somewhat more fatigue," "A little bit more fatigue," "About the same fatigue," "A little bit less fatigue," "Somewhat less fatigue," and "Much less fatigue." Participation was complete once all participants had been involved in at least 5 conversations.

Statistical analysis. Patient demographics and disease activity were assessed. Fatigue instrument scores were computed according to the scoring system of the questionnaires and were then normalized (0 = no fatigue, 100 = maximum fatigue) for appropriate comparisons between questionnaires.

The MCID were estimated using a paired approach[13] and an unpaired linear regression approach[14]. Following the paired approach, the differences in normalized fatigue instrument scores between each pair of interview participants were calculated. For each fatigue instrument, the mean and 95% confidence intervals of these score differences were calculated, stratifying by the reported fatigue contrast. The MCID was estimated by calculating the mean difference between the fatigue instrument scores of interviewees reporting "a little bit more" fatigue and their interview partners (for convenience we will refer to this as the "Greater fatigue" MCID). The MCID was also estimated by calculating the mean difference in fatigue instrument scores between interviewees reporting "a little bit less" fatigue and their interview partners (for convenience we will refer to this as the "Less fatigue" MCID). These estimates were adjusted by subtracting the mean difference for "about the same" contrast rating category.

The data were also analyzed using a 2-step regression approach. Based on the concept that a single contrast score actually represents the difference of 2 individual fatigue levels, we sought to separate each contrast into 2 subject-specific fatigue scores. This was implemented by fitting a linear mixed-effects model that produced a single fatigue value for each study participant. This model required the assumption that the relative degree of difference in fatigue was roughly equal between the different levels of

Table 1. The 7 self-administered fatigue instruments employed in this study.

Instrument	No. of Items	Response Format	Score Range*	Concepts of Interest	Prior Use	No. Citations in Studies of SLE**
FSS	9	7-point Likert scale	1–7 (positive)	Impact on ADL	SLE	19
MAF	16	10-point numerical scale and multiple choice	1–50 (positive)	Severity, distress, ADL, timing	RA	2
MFI	20	5-point Likert scale	20–100 (positive)	General/physical fatigue, activity/motivation level, mental fatigue		1
CFS	11	4-point Likert scale	0–33 (Positive)	Physical/mental fatigue	Primary SS and SLE	4
VT	4	6-point Likert scale	0–100 (negative)	Fatigue and energy level	Chronic rheumatic conditions	1
FACIT-F	13	5-point Likert scale	0-52 (negative)	General fatigue	Cancer	0
Global rating scale***	1	10-point scale	0–10 (positive)	Global fatigue	SLE, RA, primary SS	6

FSS: Fatigue Severity Scale; MAF: Multidimensional Assessment of Fatigue; MFI: Multidimensional Fatigue Inventory; CFS: Chalder Fatigue Scale; VT: Vitality scale of SF-36; FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue scale; SLE: systemic lupus erythematosus; RA: rheumatoid arthritis; SS: Sjögren's syndrome. ADL: activities of daily living. * Positive refers to increasing fatigue with increasing score, negative to decreasing fatigue with increasing score. ** Until July 2006. *** Anchored in various wording.

fatigue contrast reported (i.e., the difference between "a little bit more" and "somewhat more" was the same as the difference between "somewhat more" and "much more"). Then the individual fatigue scores were used as the predictor variable in regression models for each instrument. The MCID and its statistical significance were obtained by calculating the slope of the regression line. The mixed-effects model was also used to adjust the MCID for potential self-reference bias in interview participants and bias relating to interview-order effects. Standardized MCID values for each questionnaire were obtained by dividing the slope of the regression line by the standard deviation (SD) of the participant scores for that instrument. The standardized MCID can be interpreted as an estimate of responsiveness (defined as an effect size = mean change/SD at baseline) for an intervention that would shift the contrast score results for a patient by one category ("on average").

Data management and statistical analysis were performed using SAS statistical software (SAS Institute, Cary, NC, USA) and R[21].

RESULTS

The mean age of study participants was 47.8 years and 96% were female. The average disease duration among participants was 12.8 years (Table 2).

The mean raw and normalized scores for each measure of fatigue are displayed in Table 3. The majority of study participants reported significant levels of fatigue. Fatigue instrument scores were correlated with self-report disease activity as measured by SLAQ scores (Pearson's correlation coefficient r varied between 0.38 and 0.59 for the various questionnaires, $p < 0.001$) and the global assessment of disease activity (r varied between 0.24 and 0.56 for the various fatigue instruments, $p < 0.05$; Table 3). Fatigue severity was not associated with age ($r = –0.07$, $p = 0.6$) or disease duration ($r = 0.19$, $p = 0.11$).

Table 2. Clinical characteristics of the 80 participants with SLE.

Clinical Characteristics	Mean (SD) or N (%)	Range
Age, yrs	47.8 (12.5)	22–75
Women	77 (96)	
Disease duration, yrs	12.8 (8.8)	0.4–47.1
Disease activity*	5 (2.7)	0–10
SLAQ score**	16.1 (17.8)	3 to 39

SLAQ: Systemic Lupus Activity Questionnaire score. * Disease activity assessed by a global 11-point numerical rating scale (0 = no activity, 10 = most activity. ** The SLAQ could range between 0 (no activity) and 47 (highest activity score).

Mean differences (95% confidence intervals) in fatigue scores between pairs of interview participants for each of the 7 instruments in relation to the contrast categories are displayed in Figure 1. The mean difference in fatigue instrument scores for patients reporting "about the same" fatigue was not statistically significantly different from zero for any of the 7 instruments. The degree of difference in fatigue scores was generally consistent with the subjective report of fatigue contrast, although some of these differences appeared to be "out of order" (i.e., patients feeling "a little more" fatigue than their conversational partners scoring greater differences in measured fatigue than patients with "somewhat more fatigue"). However, the mean difference in fatigue instrument scores for patients reporting "a little more" fatigue was not statistically significantly greater than the mean difference in fatigue instrument scores for patients reporting "somewhat more fatigue." This finding applied in all such cases where the difference in fatigue instrument scores appeared to be "out of order."

The MCID for each fatigue instrument were first estimated from these data by the paired approach. The results are displayed in Table 4. "Greater fatigue" MCID were larger than the "Less fatigue" MCID for all 7 fatigue instruments; this difference was statistically significant for the FSS and MAF.

The MCID were then estimated using a regression

Table 3. Fatigue scores for 7 fatigue measurement scales in the 80 participants with SLE, and correlation of fatigue with disease activity.

Instrument	Raw Mean (SD)*	Normalized Mean (SD)**	Correlation with Disease Activity***	
			SLAQ (p)	Global Assessment (p)
FSS	5.3 (1.5)	71.1 (24.4)	0.46 (0.001)	0.42 (< 0.001)
MAF	31.1 (11.4)	56.7 (26.2)	0.57 (0.001)	0.53 (< 0.001)
MFI	62.8 (19.8)	53.5 (24.7)	0.53 (0.001)	0.29 (0.009)
CFS	16.4 (6.6)	49.8 (20.0)	0.38 (0.001)	0.24 (0.032)
VT	37.9 (24.6)	62.1 (24.6)	0.46 (0.001)	0.37 (< 0.001)
FACIT-F	25.7 (12.0)	50.6 (23.0)	0.59 (0.001)	0.49 (< 0.001)
Global rating scale	5.6 (2.7)	55.9 (27.0)	0.58 (0.001)	0.56 (< 0.001)

Definitions as in Table 1. SLAQ: the Systemic Lupus Activity Questionnaire score could range between 0 (no activity) and 47 (highest activity score). * For all but VT and FACIT-F, higher raw scores in the original scoring of the instruments indicate higher level (severity or impact) of fatigue (score range for each instrument is indicated in Table 1). ** Raw scores were converted to 0–100 scale, higher scores indicating higher fatigue levels. *** Pearson's correlation coefficients.

approach. Latent subject-specific contrast scores were calculated and regressed against fatigue instrument scores. A sample graph of the linear regression of the FSS scores versus the estimated fatigue contrast scores is shown in Figure 2. The estimated MCID and standardized MCID of all fatigue instruments are displayed in Table 5. The standardized MCID varied between 0.36 and 0.59. The MCID values are adjusted for a small but statistically significant ($p = 0.02$) self-reference bias, which did not greatly affect the estimates. Adjusting for lack of statistical independence and self-reference bias did not significantly affect the calculated MCID values.

DISCUSSION

This study was designed to estimate the quantity of difference in clinical fatigue instrument scores required for a patient with SLE to experience a noticeable difference in their subjective level of fatigue in comparison with someone else. We found that this quantity, the minimum clinically important difference (MCID), is readily measurable using a between-patients approach. The measured MCID for each fatigue instrument are fairly variable, with important implications for the design of observational studies and future clinical trials that will use fatigue as an outcome measure.

The level of fatigue reported by the group of patients studied was high but was comparable to other studies of fatigue in SLE[1,4]. In our study, fatigue was associated with disease activity, and represented a source of significant disability. Data on several important confounding factors for fatigue such as sleep disorders, depression, or fibromyalgia were not collected; therefore we could not assess their contribution to the burden of fatigue in our study population.

We employed 2 different approaches to the calculation of the MCID. The paired approach is a simple intuitive approach to the analysis. Using this technique, we found that the "Greater fatigue" MCID varied between 9.7 and 19.7 for the various fatigue instruments (normalized scores of range 0 to 100), whereas the "Less fatigue" MCID varied between –2.1 and 12. The variation in the MCID between instruments probably reflects the differing constructs of fatigue examined by the different questionnaires. These findings provide an interesting insight into the psychological aspects of patient's self-assessment. The differences (in some cases statistically significant) between the "Greater fatigue" and "Less fatigue" MCID indicate a self-reference bias on the part of the study participants: the tendency to perceive their conversation-partner's fatigue as greater than their own. This self-reference bias was clearly evident in the regression model for the individual scores (the first step in the 2-step unpaired approach). However, the estimated magnitude of this bias estimated by the regression modeling that we used was only 0.14 (on the 7-point Likert scale, corresponding to a small fraction of a "step"), not nearly large enough to significantly bias the results. Interestingly, a similar tendency was reported in 2 previous studies of the MCID using the same "between-patients" methodology[13,22].

The paired technique is subject to a number of methodological problems including a lack of statistical independence between data points (multiple observations from the same subject), self-reference bias (subjects lack an objective basis for comparison of fatigue levels), within-interview correlation (subjects interviewing one another probably tend to provide reciprocal scores, further reducing the independence of observations), and interview order effects (subjects may assess their own fatigue levels differently after comparing themselves with certain other subjects). We therefore employed a linear mixed-effects model to analysis of these data. Applying this statistical model allowed adjustment for the above sources of bias and the calculation of a more precise estimate of the MCID[14].

The standardized MCID (obtained by dividing the estimated MCID by the standard deviation of the participant scores for a given instrument) allows for meaningful comparison between fatigue instruments because it reflects the average amount of change required for a clinically noticeable difference and the variability in this required amount of change between patients for a given instrument. Using this approach, the standardized MCID varied between 0.36

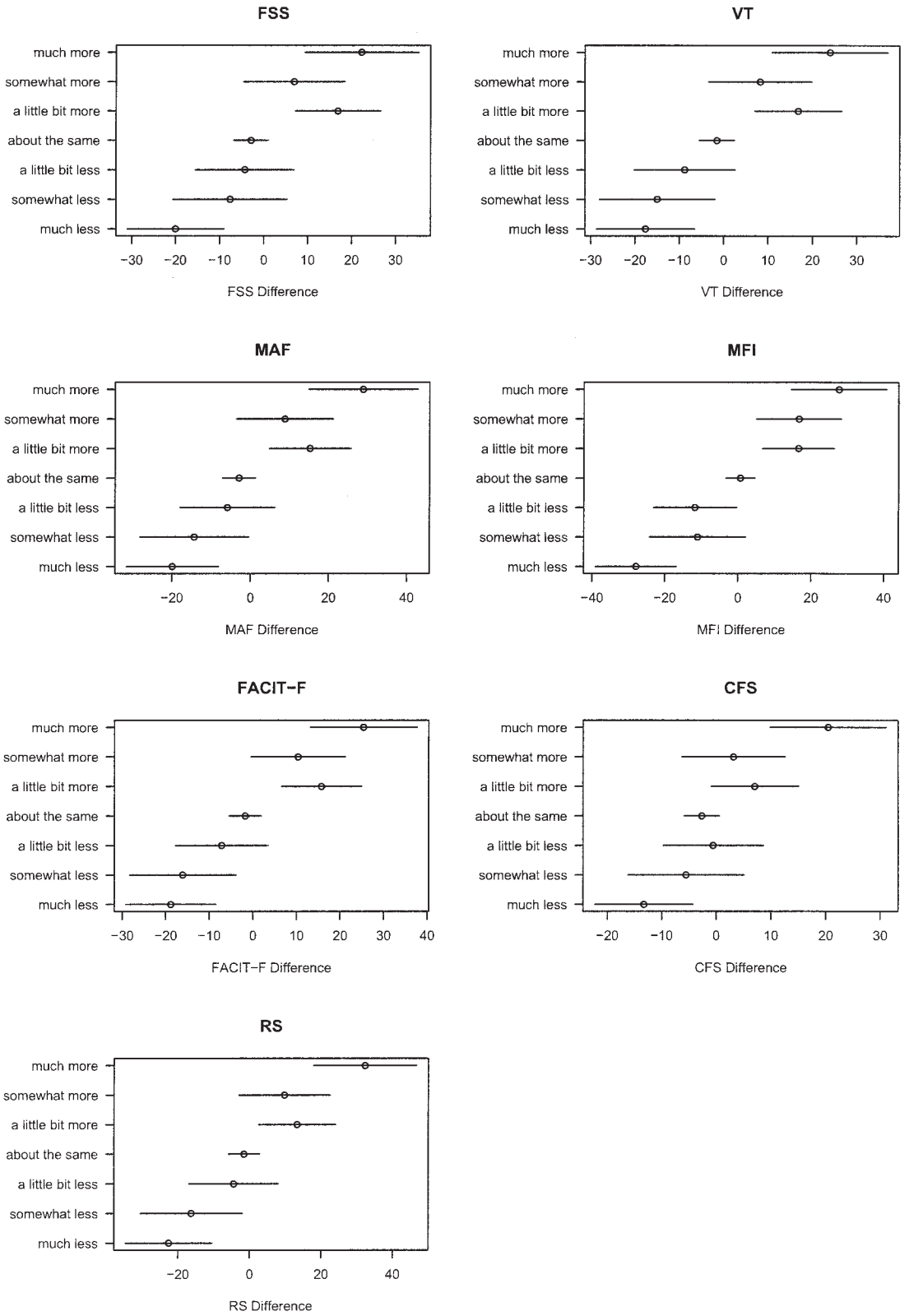

Figure 1. Mean paired differences in 7 fatigue instrument scores for each fatigue contrast rating in 80 patients with SLE. Error bars represent 95% confidence intervals. FSS: Fatigue Severity Scale; MAF: Multidimensional Assessment of Fatigue; MFI: Multidimensional Fatigue Inventory; CFS: Chalder Fatigue Scale; VT: Vitality scale of the SF-36; FACIT-F: Functional Assessment of Chronic Illness Therapy-Fatigue scale; RS: global rating scale for fatigue.

Table 4. Mean paired differences in normalized fatigue instrument scores by fatigue contrast rating in SLE (80 study participants)*.

Instrument	Much More Fatigue, N = 48	Somewhat More Fatigue, N = 38	Little More Fatigue**, N = 52	About the Same Fatigue, N = 104	Little Less Fatigue***, N = 36	Somewhat Less Fatigue, N = 38	Much Less Fatigue, N = 74
FSS	25.1 (12.6, 37.6)	9.8 (–1.1, 20.7)	19.7 (10.3, 29.2)	–2.8 (–6.4, 0.7)	–1.4 (–13.1, 10.2)	–4.8 (–17.7, 8.1)	–17.1 (–28, –6.3)
VT	25.4 (13, 38)	9.7 (–1.1, 20.6)	18.3 (8.9, 27.7)	–1.5 (–5.0, 2.0)	–7.3 (–18.9, 4.3)	–13.5 (–26.3, –0.6)	–16.1 (–26.9, –5.3)
MAF	31.8 (18.6, 45)	11.7 (0.2. 23.3)	18.2 (8.3, 28.1)	–2.9 (–6.6, 0.9)	–2.9 (–15.2, 9.4)	–11.4 (–25, 2.2)	–17 (–28.4, –5.6)
MFI	27 (15.6, 38.3)	16 (6.1, 25.9)	16 (7.3, 24.4)	0.7 (–2.5, 3.9)	–12 (–23, –1.9)	–12 (–23.4, 0)	–29 (–38.4, –18.8)
FACIT-F	27.2 (16.1, 38.3)	12.2 (2.6, 21.8)	17.5 (9.2, 25.9)	–1.8 (–4.9, 1.4)	–5.3 (–15.6, 4.9)	–14.3 (–25.7, –2.9)	–17 (–26.5, –7.5)
CFS	23.1 (12.6, 33.7)	5.8 (–3.4, 15)	9.7 (1.8, 17.6)	–2.6 (–5.6, 0.4)	2.1 (–7.7, 11.9)	–2.9 (–13.8, 7.9)	–10.6 (–19.8, –1.5)
Global rating scale	33.8 (20.6, 47)	11.3 (–0.2, 22.7)	14.8 (4.9, 24.7)	–1.5 (–5.3, 2.2)	–2.9 (–15.2, 9.4)	–14.8 (–28.3, –1.2)	–21 (–32.4, – 9.6)

Definitions as in Table 1. * Number of fatigue contrast rating. A contrast was defined as the subjective comparison rating obtained at the end of a one-on-one conversation, between both participants of a pair (each one-on-one conversation providing 2 contrasts). Results were adjusted by subtracting the "about the same" value from the raw mean paired difference for each fatigue contrast rating category. Results are mean (95% confidence intervals). ** Equivalent to "Greater Fatigue" MCID. *** Equivalent to "Less Fatigue" MCID.

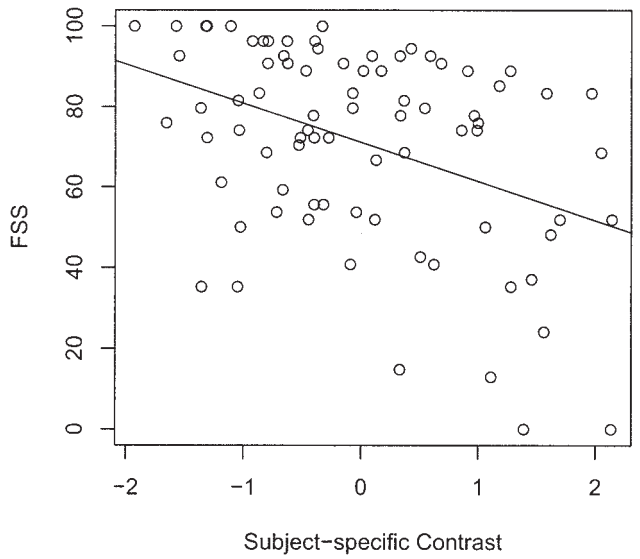

Figure 2. Sample linear regression of Fatigue Severity Scale (FSS) scores against latent subject-specific fatigue contrast scores using a mixed-effects model. The MCID is obtained from the slope of the regression line.

(CFS) and 0.59 (MFI). These findings imply that of the fatigue instruments we studied, the MFI, is the most sensitive to subjectively detectable differences in fatigue levels. Consequently, when employed in studies and clinical trials, the MFI will require a smaller sample size to detect differences in fatigue compared with the other fatigue instruments. The procedure for using the MCID in sample size calculations is discussed elsewhere[23]. However, the choice of a fatigue instrument should not rely only on the sample size computation; other instrument properties including conceptual content, length of the questionnaire, and face validity must also be considered.

Of note, the linear regression model applies the assumption of linearity to the data (i.e., positive and negative MCID ought to be approximately equal), which may in fact not be the case for some of the fatigue instruments, given the results of the paired analysis. Other sources of error in the MCID include the significant variability in the differences in fatigue measured for each level of subjective contrast in fatigue. This reflects in part significant variability in the interviews conducted between participants. The contrast in fatigue reported by each participant based on each interview was likely a function of the cohesiveness of the interview, the narrative styles of the individual participants, the quality of communication, and the degree of adherence to the structured questions provided for the interview, as well as the actual severity of fatigue experienced by each patient. Unexpectedly, "little more fatigue" and "somewhat more fatigue" mean differences were inverted (Table 4). Sample size may have played a role as evidenced by the overlapping 95% CI. Also, it is possible that people do not interpret the difference between "somewhat more" and "a little more" in a consistent way. The wording of the transition question should probably be changed in future studies to correct this.

The method of measuring the MCID employed in this study is a cross-sectional between-patients approach based on the method described by Redelmeier and Lorig[13]. The other major approach is a longitudinal within-patients approach developed by Jaeschke, *et al*[12]. This method has been applied to several chronic diseases to estimate the MCID[24-26]. The estimation of MCID is based on the intrapersonal variation of the outcome score (i.e., fatigue) between the onset and the end of a therapeutic intervention that is related to a single transition question about the change that occurred. The MCID for improvement is then defined as the difference between the mean effect of the intervention assessed by the instrument of those who rated themselves as slightly better and those who rated themselves as about the same. The disadvantage of the longitudinal approach, particularly for highly subjective constructs such as fatigue, is the presence of a recall bias in making a retro-

Table 5. Estimation of minimal clinically important differences for 7 measures of fatigue in SLE*.

Instrument	MCID (95% CI) (normalized scaling)	Standardized MCID (95% CI)**	MCID (95% CI) (original scaling)
FSS	9.7 (4.9, 14.6)	0.41 (0.2, 0.57)	0.6 (0.3, 0.9)
VT	10.7 (5.9, 15.5)	0.44 (0.25, 0.60)	–10.7 (–15.5, –5.9)
MAF	11.5 (6.4, 16.7)	0.45 (0.25, 0.61)	5.0 (2.8, 7.2)
MFI	14.3 (10.0, 18.7)	0.59 (0.42, 0.72)	11.5 (8.0, 15.0)
FACIT-F	11.3 (6.9, 15.6)	0.50 (0.31, 0.65)	–5.9 (–8.1, –3.6)
CFS	7.0 (2.9, 11.1)	0.36 (0.15, 0.53)	2.3 (1.0, 3.7)
Global RS	12.9 (7.7, 18.1)	0.49 (0.3, 0.64)	1.3 (0.8, 1.8)

Definitions as in Table 1. * Estimations using linear regression modeling. ** Standardized MCID were obtained as MCID divided by the standard deviation of the participant scores for the corresponding instrument.

spective judgment about the severity of previous levels of fatigue especially in longterm trials. The minimum change reported by a patient may be biased by her or his expectations for the therapeutic outcome; patients anticipating major improvement will be less likely to consider minor improvements to be meaningful changes in their condition. Another method of assessing clinically significant change was attempted using clinician-rated change based on retrospective reviews of medical records, but significant variability was observed between clinicians in terms of the degree of change perceived to be clinically significant[27]. The cross-sectional approach is easier to implement and is not subject to a recall bias. It is logistically more feasible than longitudinal methods of estimating the MCID. However, it also has several limitations including a lack of statistical independence between data points, a self-reference bias, a within-interview correlation, and an interview order effect. Although one study comparing both methods found no significant difference in the resulting MCID values[28], it is possible that clinically meaningful differences between patients may not be the same as differences within patients[12]. Both longitudinal and cross-sectional methods provide group-based estimates of MCID in contrast to individual-based estimates[29].

As Redelmeier and Lorig[13] point out, a subjectively detectable difference is not necessarily the same as a clinically significant difference. Indeed, the term "MCID" as it is usually employed in the literature applies to the detection of a minimally "detectable" or "perceptible" change/difference without consideration of whether the change/difference is actually "important" or "significant" from the patient's perspective[30]. This is particularly relevant when the MCID is applied to the analysis of clinical trial data, where patients may experience a measurable change in fatigue equal to the MCID in fatigue; however, they may not consider this change in fatigue "worthwhile," given the possible adverse effects of trial interventions or their own expectations for therapeutic outcome.

In summary, the MCID is an important parameter for clinical research and practice. It can provide a basis for estimating desired effect sizes in the design of clinical trials, and hence can facilitate sample size calculation. It allows clinicians to estimate the amount of benefit required for an intervention to produce at least a perceptible benefit to patients. In this regard, measurement of the MCID is especially useful when considering patient-reported outcomes such as fatigue. The estimates of the MCID for each of the fatigue instruments assessed in this study provide useful information for application to research in amelioration of the significant fatigue and associated disability experienced by patients with SLE. The standardized MCID is useful in comparing the sensitivity of the various instruments, and based on our results the MFI and the FACIT-F have the most advantageous MCID of fatigue in SLE. However, before those instruments can be recommended, further studies should confirm our results.

Future research should examine the relationships between the MCID and truly meaningful (not merely detectable or perceived) changes for patients, similar to the information obtained using recently developed concepts such as the low disease activity state or the patient acceptable symptom state[31]. Also, the MCID should be validated against functional improvements, mood improvements, and objective disease activity measures.

REFERENCES

1. Zonana-Nacach A, Roseman JM, McGwin G Jr, et al. Systemic lupus erythematosus in three ethnic groups. VI: Factors associated with fatigue within 5 years of criteria diagnosis. LUMINA Study Group. LUpus in MInority populations: NAture vs Nurture. Lupus 2000;9:101-9.
2. Tench C, Bentley D, Vleck V, McCurdie I, White P, D'Cruz D. Aerobic fitness, fatigue, and physical disability in systemic lupus erythematosus. J Rheumatol 2002;29:474-81.
3. Tench CM, McCurdie I, White PD, D'Cruz DP. The prevalence and associations of fatigue in systemic lupus erythematosus. Rheumatology Oxford 2000;39:1249-54.
4. Bruce IN, Mak VC, Hallett DC, Gladman DD, Urowitz MB. Factors associated with fatigue in patients with systemic lupus erythematosus. Ann Rheum Dis 1999;58:379-81.
5. McKinley PS, Ouellette SC, Winkel GH. The contributions of

disease activity, sleep patterns, and depression to fatigue in systemic lupus erythematosus. A proposed model. Arthritis Rheum 1995;38:826-34.
6. Jump RL, Robinson ME, Armstrong AE, Barnes EV, Kilbourn KM, Richards HB. Fatigue in systemic lupus erythematosus: contributions of disease activity, pain, depression, and perceived social support. J Rheumatol 2005;32:1699-705.
7. Alarcon GS, McGwin G Jr, Uribe A, et al. Systemic lupus erythematosus in a multiethnic lupus cohort (LUMINA). XVII. Predictors of self-reported health-related quality of life early in the disease course. Arthritis Rheum 2004;51:465-74.
8. Schwartz JE, Jandorf L, Krupp LB. The measurement of fatigue: a new instrument. J Psychosom Res 1993;37:753-62.
9. 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.
10. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. J Psychosom Res 1993;37:147-53.
11. Smets EMA, Garssen B, Bonke B, De Haes JCJM. The Multidimensional Fatigue Inventory (MFI): psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315-25.
12. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
13. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
14. Brant R, Sutherland L, Hilsden R. Examining the minimum important difference. Stat Med 1999;18:2593-603.
15. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus [letter]. Arthritis Rheum 1997;40:1725.
16. Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The Fatigue Severity Scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989;46:1121-3.
17. 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.
18. Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. J Rheumatol 1995;22:639-43.
19. Karlson EW, Daltroy LH, Rivest C, et al. Validation of a Systemic Lupus Activity Questionnaire (SLAQ) for population studies. Lupus 2003;12:280-6.
20. Bae SC, Koh HK, Chang DK, Kim MH, Part JK, Kim SY. Reliability and validity of systemic lupus activity measure-revised (SLAM-R) for measuring clinical disease activity in systemic lupus erythematosus. Lupus 2002;10:405-9.
21. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2005. [Internet. Accessed January 10, 2007.] Available from: http://www.R-project.org
22. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh BJ, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. J Rheumatol 1993;20:557-60.
23. Schulz KF, Grimes DA. Sample size calculations in clinical trials: mandatory and mystical. Lancet 2005;365:1348-53.
24. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life instrument. J Clin Epidemiol 1994;47:81-7.
25. Ringash J, Bezjak A, O'Sullivan B, Redelmeier DA. Interpreting differences in quality of life: the FACT-H&N in laryngeal cancer patients. Qual Life Res 2004;13:725-33.
26. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. Cancer 2007;110:196-202.
27. American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria. The American College of Rheumatology response criteria for systemic lupus erythematosus clinical trials: measures of overall disease activity. Arthritis Rheum 2004;5011:3418-26.
28. Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. J Clin Epidemiol 1996;49:1215-9.
29. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. J Rheumatol 2001;28:406-12.
30. Wright JG. The minimal important difference: who's to say what is important? J Clin Epidemiol 1996;49:1221-2.
31. Tubach F, Wells GA, Ravaud P, Dougados M. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. J Rheumatol 2005;32:2025-9.