INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

An Evaluation of the Responsiveness of Two Systemic Lupus Erythematosus Disease Activity Indices

Erika Chang
Department of Epidemiology and Biostatistics
McGill University, Montreal

July 2000

A thesis submitted to the faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science (M.Sc.)

ABSTRACT

Objectives: 1) To measure the responsiveness of SLAM-R and SLEDAI to meaningful changes in SLE activity; 2) to determine how strongly activity in specific organ systems affects SLAM-R and SLEDAI responsiveness.

Methods: A secondary analysis was performed on blinded data of SLE patients. Sensitivity of SLAM-R and SLEDAI to change were assessed with traditional measures. Also, perceived change in disease activity was modelled as a function of change in overall instrument scores, and of change in organ system subscores.

Results: Both SLAM-R and SLEDAI were responsive to changes perceived by physicians. However, only SLAM-R was sensitive to changes reported by patients. The relevance of type of organ involvement depended on whether the assessor was the patient or the physician.

Conclusion: The differences between the type of change relevant to physicians and patients may account for SLAM-R's better ability to reflect patients' judgments.

RÉSUMÉ

But: 1) Mesurer la sensibilité du SLAM-R et SLEDAI aux changements importants de l'activité du lupus erythémateux disséminé (LED); et 2) évaluer l'importance des changements de l'activité des systèmes organiques individuels.

Méthodes: La sensibilité aux changements et le pouvoir de prédire les améliorations ou les exacerbations sont mesurés par une analyse secondaire. Les changements aperçus dans l'activité du LED sont modelés comme les effets des changements aux scores des systemes organiques individuels du SLAM-R et SLEDAI.

Résultats: SLAM-R et SLEDAI sont les deux instruments les plus sensibles aux changements aperçus par les médecins. Mais, SLAM-R est le seul à être sensible aux changements qui sont importants aux personnes atteintes du lupus. La pertinence de l'implication de l'organe dépend de si le médecin où la personne atteinte du LED est celui qui fait l'évaluation de l'activité du LED.

Conclusions: Les différences entre l'implication des organes individuels qui sont importants aux médecins et aux personnes atteintes du lupus semblent expliquer pourquoi SLAM-R peut distinguer mieux que SLEDAI les changements aperçus par les patients.

PREFACE

This thesis includes two manuscripts intended for publication. This option is permitted by the McGill University Faculty of Graduate Studies and Research. Accordingly, they have established the following guidelines:

> Candidates have the option of including, as part of the thesis, the text of one or more papers submitted or to be submitted for publication, or the clearly-duplicated text of one or more published papers. These texts must be bound as an integral part of the thesis.
>
> If this option is chosen, connecting texts that provide logical bridges between the different papers are mandatory. The thesis must be written in such a way that it is more than a mere collection of manuscripts; in other words, results of a series of papers must be integrated.
>
> The thesis must still conform to all other requirements of the "Guidelines for Thesis Preparation". The thesis must include: A Table of Contents, an abstract in English and French, an introduction which clearly states the rationale and objectives of the study, a review of the literature, a final conclusion and summary, and a thorough bibliography or reference list.
>
> Additional material must be provided where appropriate (e.g. in appendices) and in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.
>
> In the case of manuscripts co-authored by the candidate and others, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. Supervisors must attest to the accuracy of such statements at the doctoral oral defense. Since the task of the examiners is made more difficult in these cases, it is in the candidate's interest

to make perfectly clear the responsibilities of all the authors of the co-authored papers.

CONTRIBUTIONS OF AUTHORS

This thesis contains the text of two papers intended for publication. The authors of both papers are, in this order: Erika Chang, BSc, Michal Abrahamowicz, PhD, Diane Ferland, RN, BScN, and Paul Fortin, MD, MPH. The research questions were formalized through discussions between Erika Chang, Michal Abrahamowicz, and Paul Fortin. The data used in the analyses were collected as part of a larger study, SMILE, funded by The Arthritis Society grant #95072, of which Dr. Paul Fortin is the Principal Investigator and Dr. Michal Abrahamowicz is one of the co-investigators. Original data was collected and entered in the computer by Diane Ferland. All data manipulation and analyses reported in this thesis were carried out by Erika Chang. The text of both manuscripts, as well as of all other sections of this thesis, was written by Erika Chang, with text revisions suggested by Drs. Michal Abrahamowicz and Paul Fortin. Interpretation of the results was based on discussions between Erika Chang, Dr. Michal Abrahamowicz, and Dr. Paul Fortin.

STATEMENT OF ORIGINALITY

I investigated the responsiveness of two systemic lupus erythematosus (SLE) disease activity measures, SLAM-R and SLEDAI, to relevant changes in SLE activity. Although previous work has been done in this area, my approach contains several novel elements. In addition to evaluating the overall responsiveness of the instruments, I analyzed the responsiveness of item subgroups to relevant change. This has not been studied before, and may further our understanding of how the differences in structure and content of SLAM-R and SLEDAI affect their psychometric properties. Second, in contrast with much of the existing literature, which concentrates on changes relevant to physicians alone, I focused on changes in SLE activity important to patients as well. This has allowed us to relate patient response to changes in the score of each instrument, and understand better the implications of such score changes in clinical trials and other research.

I employed both traditional and non-traditional methods to assess responsiveness. In the former instance, I calculated 95% confidence intervals for each measure, using bootstrapping techniques. Because the sampling distributions of the measures are unknown, their confidence intervals cannot be calculated parametrically and so are rarely seen in the literature. By using non-parametric methods instead, I was able to produce estimates that could be interpreted more meaningfully.

As well as the traditional methods for assessing responsiveness, I have used the generalized estimating equation (GEE) approach to logistic regression to measure the ability of SLAM-R and SLEDAI to predict meaningful change in disease activity. This technique enables us to increase the power of our analysis by including repeated measures of patients in our analysis without introducing bias caused by intra-patient correlation, thus making research on rare diseases such as SLE more feasible. Also, by measuring the predictive value of the instruments, we may enhance our interpretation of their score changes in future studies.

ACKNOWLEDGEMENTS

There are many people without whom completion of this thesis would not have been possible. First and foremost, I would like to thank my supervisor, Dr. Michal Abrahamowicz, for his encouragement, patience, and inspiration over the past year and a half. His feedback taught me a great deal and helped me every step of the way, and his humour made the entire learning process an enjoyable experience. I am also grateful to Dr. Paul Fortin, who, as my other thesis committee member, provided endless amounts of encouragement and insight on SLE and measurement of its activity.

Some parts of this work were supported financially through a grant from the Canadian Arthritis Network to Dr. Abrahamowicz.

I am greatly indebted to the investigators, research assistants, and patients involved in the Study of Methotrexate in Lupus Erythematosus. None of this work would have been possible without the devotion of their time and energy, and I hope that in return they benefit from my work in some way.

My thanks go out to the staff and members of the Clinical Epidemiology Division of the Montreal General Hospital, in particular Diane Ferland, Roxane du Berger, and Jennifer Gardner. Diane provided me with the data for my analyses as well as the study documentation, and aided me in translating my thesis abstract into French; Roxane assisted me with the bootstrapping steps of my analyses; and Jennifer provided valuable secretarial support, especially in the final weeks.

I also wish to thank the Department of Epidemiology and Biostatistics, in particular the faculty for the valuable instruction they have given me, the staff for all of their support and assistance, and my friends and classmates.

My time at McGill would not have been nearly as enjoyable without the likes of the "Epi gang", the McGill Choral Society, Cantare, and all my other friends. In particular, I would like to thank Matt Tracy for all his support and encouragement.

Finally, I wish to thank my mom, my dad, Lucas, and Judy, who rooted for me and believed in me all the way.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic, incurable autoimmune disease that can affect all parts of the body. It is characterized by unpredictable periods of activity (flare) and remission. The clinical manifestations are due to antibody-mediated inflammation and may include arthritis, fever, weight loss, fatigue, alopecia, rash, serositis, and renal or central nervous system (CNS) involvement. However, these signs and symptoms vary over time and between patients, thus making diagnosis difficult.

The prevalence of SLE varies from study to study, probably partly due to different methods of ascertainment and to time trends (Uramoto, Michet, Thumboo, et al., 1999). Estimates range from 14.6 to 50.8 cases per 100,000 persons in the United States (Hochberg, 1993; Klippel, 1997). Uramoto et al. (1999) calculated a prevalence in a Rochester, Minnesota, population of 1.22 per 1,000 (95% CI 0.97 - 1.47) or 122 per 100,000 as of January 1993, after adjusting their study population for age and sex to the 1970 U.S. white population. This varies by sex, age, and ethnicity; approximately 80% of cases are seen in women of childbearing age (women:men ratio = 9:1), and the disease appears to be more common in Black and Asian women than in Caucasian women (Hochberg, 1993). Although the prevalence of SLE has not been formally measured in Canada, there were estimated to be between 15,000 and 50,000 cases in the early 1990s (Senécal, 1991). Estimates of incidence also vary by study. Uramoto et al. (1999) found an average age- and sex-adjusted incidence of 5.56 per 100,000 per year in a cohort followed from 1980 to 1992. They also reported an increase in SLE incidence over time, but suggested that this may be due partly to increasing detection of milder cases. This study, though, may be generalizable only to the White population, which comprised 94% of the source population. The annual incidence estimated by Klippel (1997) was similar, ranging from 5 to 7 per 100,000.

The etiology of SLE is unknown, but may include both genetic and environmental factors. The role of genes is suggested by the higher concordance of SLE cases in monozygotic twins than in dizygotic twins or other pairs of siblings. Hormones have also been suspected, since the disease is much more common in women of childbearing age than in men or in older and younger age groups. Environmental factors have been implicated as well. People and even pets living in close proximity to SLE patients appear to be at a higher

risk than the general population for the disease (Panush, Levine, & Reichlin, 2000). Drugs such as procainamide have induced some cases of SLE. In addition, ultraviolet radiation may play a role in the pathogenesis, since SLE patients are frequently photosensitive (Hochberg, 1993).

Treatment for SLE depends on the severity of the manifestations. Non-steroidal anti-inflammatory drugs (NSAIDS) may be used for mild cases, but the most common therapy involves administration of corticosteroids such as prednisone. Other drugs that may be used include anti-malarials, such as hydroxychloroquine, or cytotoxic agents, such as cyclophosphamide and azathioprine. The side effects of these drugs are undesirable; corticosteroid use, for example, may result in disfigurement due to fat redistribution, osteoporosis, or opportunistic infections, and it may also lead to myocardial infarction (Senécal, 1991).

In the past, lupus was often rapidly fatal. The 5-year survival rate among SLE patients in 1950 was estimated at around 50% (Schroeder & Euler, 1997); a study of a cohort followed from 1950 to 1979 found average 5- and 10-year survival rates of 75% and 50% respectively (Uramoto et al., 1999). However, improvements in treatment and recognition of milder cases have resulted in an increase of the 5 and 10-year survival rates to as high as 91% and 76% respectively (Jacobsen et al., 1998). The prognosis appears to be worse for men than for women (Chang, Chang, Kuo, Chu, & Chang, 1998; Kiss, Regeczy, & Szegedi, 1999; Xie, Feng, & Fu, 1998). Death is often caused by infections associated with use of immunosuppressants, cardiovascular disease, and renal or central nervous system involvement (Hochberg, 1993; Mills, 1994; Klippel, 1997).

The improved prognosis for SLE has resulted in greater interest in outcomes other than death, such as disease activity (Liang, Stern, & Esdaile, 1988b). Disease activity, which includes all reversible organ dysfunction (Bombardier, Gladman, Urowitz, Caron, & Chang, 1992), has traditionally been assessed subjectively and objectively by physicians. Though standardization is needed to facilitate comparisons of studies, the great variability of SLE manifestations, and the incomplete knowledge of biological mechanisms involved in the disease pathway, have made it very difficult to develop a measure of SLE activity that both provides a global, numeric score, and satisfies all researchers. As a result, over 60 measures with varying psychometric characteristics exist (Liang, Socher, Roberts, & Esdaile, 1988a).

Two commonly-used indices in North America are the revised SLE Activity Measure (SLAM-R) (Liang, Socher, Larson, & Schur, 1989) and the SLE Disease Activity Index (SLEDAI) (Bombardier et al., 1992).

CHAPTER 2: LITERATURE REVIEW

This literature review comprises three sections. In the first section, I will review selected measures of SLE activity. In the second section, I will discuss methods of measuring responsiveness of instruments to clinically significant change. In the third section, I will discuss research on the responsiveness of the two most popular measures of lupus activity: SLAM-R and SLEDAI.

2.1. A review of selected SLE activity measures.

The improving survival rate of SLE patients (Schroeder et al., 1997) has resulted in an increased interest in studying other outcomes related to the evolution of the patient's health over time, such as lupus disease activity. Measuring disease activity in SLE is important because, if not properly managed, the inflammation can lead to permanent tissue damage and death (Gladman et al., 1996; Neville et al., 2000). Since the subjective global assessment of disease activity often used by clinicians does not allow researchers to compare patients from different trials, or seen by different physicians, a standardized measure of SLE activity is desirable (Bombardier et al., 1992; Liang et al., 1988b). Unfortunately, SLE activity is difficult to quantify. Because of the multi-systemic nature of the disease, and the heterogeneity of manifestations between patients and over time, there is no single clinical or laboratory result that can be used as a gold standard for global activity (Fortin, Abrahamowicz, & Danoff, 1995; Petri, Hellmann, & Hochberg, 1992). More importantly, there is little agreement among researchers on what constitutes disease activity. Consequently, over 60 different instruments possessing a variety of scoring systems and content exist (Liang et al., 1988a). Three of the most widely-used instruments are the revised Systemic Lupus Activity Measure (SLAM-R) (Liang et al., 1989), the SLE Disease Activity Index (SLEDAI) (Bombardier et al., 1992), and the British Isles Lupus Assessment Group (BILAG) score (Symmons et al., 1988).

SLAM-R (Liang et al., 1989; revised in 1991) includes 23 clinical and 7 laboratory items representing 11 organ systems (constitutional, integument, eye, reticuloendothelial, pulmonary, cardiovascular, gastrointestinal, neuromotor, and musculoskeletal, renal, and hematological), and one "miscellaneous" item for scoring manifestations not listed

elsewhere. Members of the American Rheumatology Association's Council on SLE determined the content of the instrument, basing their decisions on the frequency of appearance, ability to measure, and ease of operationalizing the manifestations. Each item is described, and *ad hoc* ascertainment and scoring rules for the "miscellaneous" item are recorded by the physician. Item scores depend on both the absence/presence and the severity of organ involvement. Thus, if a manifestation is absent, the corresponding item score is zero, and if it is present, the score varies from 1-3, with 3 signifying greatest severity. The maximum possible score is 84 if the miscellaneous item is included, and 81 if it is not. SLAM-R covers manifestations occurring during the preceding month. A copy of this instrument can be found in Appendix A of this thesis.

SLEDAI (Bombardier et al., 1992) consists of 16 clinical and 8 laboratory items representing 9 organ systems: central nervous system (CNS), vascular, musculoskeletal, renal, integument, serosal, immunological, constitutional, and hematological. The group who developed this questionnaire sought to model the physician's thought process in estimating global SLE activity. A list of thirty-seven clinical manifestations of SLE found in the literature was given to 15 rheumatologists, who individually rated the importance of each for disease activity evaluations; only the 24 items deemed most important were retained. Clinicians assessed 58 profiles of real patients and 35 hypothetical "paper" profiles generated from combinations of the 24 items, and rated the overall disease activity level of each profile on a scale of 0-10. The independent effects of involvement of a particular organ system on the clinician's SLE activity level rating were estimated with multiple regression modelling, with the dependent variable defined as the overall SLE activity rating. In the final version of SLEDAI, item weights are based on the regression parameter for the corresponding organ system. In contrast with SLAM-R, only the presence or absence of a manifestation is recorded, but the maximum possible score for a given item varies according to the perceived seriousness of the sign. For example, CNS-related items each score 8, whereas renal items each score 4 and hematological items each score only 1. In contrast to SLAM-R, only events occurring within the 10 days up to and including the day of measurement are recorded. The maximum possible SLEDAI score is 105, but physicians rarely see scores higher than 46 (Bombardier et al., 1992). A copy of SLEDAI is also included in Appendix A.

The BILAG scoring system (Symmons et al., 1988) is a computerized index developed in the 1980s by a group of British rheumatologists. It records manifestations related to eight organs/systems (mucocutaneous, central nervous sytem, renal, musculoskeletal, cardiovascular and respiratory, vasculitis, hematological, and non-specific items) and, as with SLAM-R, includes only those occurring during the four weeks prior to measurement. Unlike SLAM-R and SLEDAI, it assigns letter grades to each system and does not generate a cumulative score. The letter grades reflect both presence and severity of organ involvement, with "A" indicating activity requiring immediate treatment, and "D" indicating inactivity. Brunner, Feldman, Bombardier, et al. (1999) also use a category "E" in their study to signify inactivity with lack of past involvement. The classification of the manifestations by letter grade was agreed upon by the rheumatologists involved in the development of this instrument. Although BILAG was not intended originally for quantifying overall disease activity, researchers have proposed various letter-to-number conversion formulae for this purpose (Gladman et al., 1994; Liang et al., 1989; Stoll, Stucki, Malik, Pyke, & Isenberg, 1996). BILAG scores were not recorded in this study and so the instrument will not be included in this analysis or discussed further in this thesis.

The content of SLAM-R and SLEDAI differs in several ways. First, although both instruments include many of the same organ systems, gastrointestinal, reticuloendothelial, and pulmonary involvement appear only in SLAM-R, and immunological activity only in SLEDAI. Second, SLAM-R rates manifestations directly observable only by the patient, such as fatigue and joint, abdominal, and chest pain, and places more emphasis than SLEDAI on the effect of organ involvement on the patient, differentiating between perceptible but bearable symptoms, and limiting or incapacitating symptoms (Ward, Marx, & Barry, 2000). In contrast, SLEDAI includes only those items that can be measured or observed directly by the physician. Third, some items are defined differently in each instrument. Such is the case for vasculitis, which is scored by total body surface area involved in SLAM-R and by presence of various internal events in SLEDAI, and for eye involvement, which includes optic neuritis in SLEDAI but not in SLAM-R (Brunner et al., 1999). Fourth, some items are combined in one instrument and not in the other. Cortical dysfunction in SLAM-R encompasses psychosis and organic brain syndrome, which are individual items in SLEDAI; urinary casts, haematuria, proteinuria, and pyuria are separated in SLEDAI but combined

under "urine sediment" in SLAM-R; and the three eye items in SLAM-R are grouped under a single descriptor in SLEDAI (Brunner et al., 1999).

SLAM-R and SLEDAI also differ in their emphases. Whereas SLAM-R weights items by severity of the manifestation regardless of the organ system involved, SLEDAI weights items by organ system, giving more weight to some than others (Fortin et al., 2000). For example, CNS-related items constitute 50% and 17%, respectively, of the maximum possible SLEDAI and SLAM-R scores, even though they comprise 25% and 17%, respectively, of the total number of items in each. As a result of these differences, it is quite conceivable that the two instruments might differ in their responsiveness to relevant change in SLE activity.

2.2. Methods for measuring responsiveness.

Responsiveness, or sensitivity to change, is the ability of an instrument to detect clinically important change (Vliet Vlieland, Zwinderman, Breedveld, & Hazes, 1997; Kazis, Anderson, & Meenan, 1989; Guyatt, Walter, & Norman, 1987; Deyo, Diehr, & Patrick, 1991; Fortin, Stucki, & Katz, 1995). Evaluation of this property is important for several reasons. First, changes in score will be meaningless unless researchers know how to interpret them and know whether a given score change indicates a change in actual disease activity or functional status. Second, disease-specific instruments may not be equally appropriate for all patients, and evaluating responsiveness is one way to decide on the best instrument to use in a given circumstance. Third, poor responsiveness of a scale may suggest the need to modify the instrument.

Though it is to some extent dependent on reliability and validity, responsiveness is distinct from both properties. Reliability, a measurement of how consistent a scale will be when used on the same patient under fixed conditions, can affect responsiveness in that poor reliability increases the difficulty in distinguishing meaningful score changes from random changes (Fortin et al., 1995). An instrument with poor validity may or may not show good responsiveness, depending on what it is actually measuring (Guyatt et al., 1987). For example, an instrument may not be valid for measuring back pain, but it might be responsive if the attribute it actually measures changes when the pain changes. On the other hand, an instrument might be very valid but unresponsive. For example, if the changes in disease

activity required to produce a score change are much larger than what patients normally experience, then responsiveness will be low. The same effect might occur if the items in the scale are not relevant to a specific patient population (Guyatt, Deyo, Charlson, Levine, & Mitchell, 1989).

Unfortunately, there is no consensus on the best method for assessing responsiveness. The ones described below lie in two main categories: calculations of "signal-to-noise" ratios, and comparisons of instrument score changes with some external criterion, typically a change in health or function that is noted by either physician or patient.

2.2.1. Effect size (Kazis et al., 1989)

Effect size is defined as the ratio of mean score change to the standard deviation of the baseline score:

$$ES = \frac{mean\ score\ change}{standard\ deviation\ of\ baseline\ score}$$

Standardizing score changes by taking the ratio of mean score change to a standard deviation is common in "signal-to-noise" responsiveness statistics, and makes it possible to compare the responsiveness of instruments possessing very different scoring ranges. However, the use of the baseline score standard deviation when calculating effect size results in a ratio that is not very meaningful. Supposing, for example, that there are two study populations, one with a much wider spectrum of true initial disease activity than the other, and that over time the mean disease activity in each group changes equally. If an instrument reflects the variation in baseline activity properly, the standard deviation of baseline instrument scores will differ for the two groups, and the effect size and therefore responsiveness will seem greater in the group with the narrower range of true disease activity. Therefore, effect size values depend partly on the characteristics of the population for which they are estimated.

2.2.2. Standardized Response Mean (Liang, Fossel, & Larson, 1990)

The standardized response mean (SRM) is the ratio of mean score change to the standard deviation of change for those subjects:

$$SRM = \frac{mean\ score\ change}{standard\ deviation\ of\ change}$$

It has been argued that, since the SRM is related to the paired t-test, it is "a natural statistical index for evaluating response magnitude" (Liang et al., 1990). In addition, because the SRM uses the SD of the score change rather than the SD of the baseline score for standardization, it gives us a better indication than ES of the relevance of the change. Guidelines for interpreting SRM values exist; according to these, an SRM of 0.3 or less indicate low responsiveness, an SRM of 0.6 indicates moderate responsiveness, and an SRM of 1.0 or more indicates high responsiveness (Cohen, 1988). In addition, by squaring the ratio of SRMs, we can find the ratio of sample sizes necessary in different instruments for the detection of a statistically significant change in scores (Liang, 1995). However, as with ES, SRM values are dependent on the characteristics of the population used. If the true change in disease activity is homogeneous and the instrument reflects this, then the standard deviation of score change should be small. However, if some patients improve, some worsen, and some remain the same, then the mean change in score will tend toward zero, and the standard deviation will be large, thereby decreasing SRM and making the instrument look unresponsive.

Because the sampling distribution of SRM is not known, Liang et al. (1990) suggest using a jackknife procedure to find the point estimate and 95% confidence interval, and to do hypothesis testing.

2.2.3. Control-Standardized Response Mean (Guyatt et al., 1987; Deyo et al., 1991)

The control-standardized response measure (CSRM) is the ratio of mean score change to standard deviation of change in the "control" group, with the "control" group being those patients categorized as "unchanged" according to an external criterion:

$$CSRM = \frac{mean\ score\ change}{standard\ deviation\ of\ change\ in\ stable\ group}$$

Variations in score changes in the control group represent the score change that would occur randomly. Therefore, any change beyond this is regarded as significant (Guyatt

et al., 1987). However, this theory is based on the assumption that the mean score change in stable patients is zero. If, for example, the score changes equally in patients who improve and those who don't, then the instrument might appear responsive according to the CSRM value, even though it is unable to distinguish between the two groups of patients. Deyo et al. (1991) suggest adjusting for this by subtracting the control group's mean score change from that of the group of interest.

Another problem with the CSRM is that if patient characteristics affecting responsiveness of a measure differ between the control group and the group of interest, then the CSRM may not be valid. For example, if the patients who improve generally have much greater initial disease activity than those who do not change, and the index being assessed behaves differently among sicker patients, then the standard deviation of score changes in the control group may not represent "random" score changes, i.e. changes expected in the group of interest when there is no improvement.

2.2.4. Relative Efficiency (Liang, Larson, Cullen, & Schwartz, 1985)

Instead of providing a direct estimate of responsiveness, relative efficiency (RE) compares the efficiency of two instruments for detecting change in a patient's condition. Liang et al. (1985) define RE as the squared ratio of two t-tests. A RE of <1 indicates that the instrument with the denominator t-statistic is more efficient, whereas a RE of >1 indicates that the other instrument is more efficient. Although the inclusion of sample sizes in a responsiveness statistic may seem inappropriate, Liang et al. (1985) argue that it is necessary when the sample sizes for data from two instruments is unequal.

2.2.5. Correlation

Responsiveness is sometimes expressed as the strength of the correlation between the score change in the instrument of interest and the change in an external criterion. An example of this is the comparison by Meenan, Anderson, Kazis, et al. (1984) of clinical and health status measures in their study on rheumatoid arthritis patients. However, because the correlation coefficient is affected by the ranges of the variables of interest, a narrow scale in either variable may decrease the absolute value of the coefficient (Matthews & Farewell, 1988) and, consequently, the apparent responsiveness of the instrument. In addition, the

correlation coefficient does not indicate the magnitude of the relationship between the two variables. As a result, by measuring correlation alone we are unable to determine the value of an instrument score change expected when a clinically relevant change is observed, or the expected probability that a relevant change will be reported at a given instrument score change.

2.2.6. Time-path diagrams (Stucki, Liang, Stucki, Katz, & Lew, 1999)

Stucki et al. (1999) use time-path diagrams to visualize the relationship between health status indicator scores and transition scores. Baseline and final scores are graphed on two vertical axes and the line connecting the points is coded according to the transition scores (e.g. solid line = perceived improvement and dotted line = no change or perceived decline). By using this technique, we can observe directly any trends in instrument score changes at given values of the transition scale. For example, a time-path diagram of health status scores recorded on an instrument sensitive to change should show lines with a consistently positive (or negative) slope when there is relevant reported improvement, flat lines when there is no relevant change, and lines with a consistently negative (or positive) slope when there is relevant deterioration. In comparison, the time-path diagram for an unresponsive instrument might show lines with little or no slope regardless of the direction of reported change. We can also determine with this technique if floor or ceiling effects are influencing the range of score changes, resulting, for example, in scores clustering at one end of the scale. However, as this method does not quantify responsiveness, interpretation of the graphs is somewhat subjective, and a comparison of the responsiveness of different measures may be difficult.

2.2.7. Receiver Operating Characteristic (ROC) curves (Deyo & Centor, 1986)

Deyo and Centor (1986) proposed treating the instrument being evaluated as a diagnostic test and plotting sensitivity vs. 1-specificity for score changes of various magnitudes. The standard of comparison is a binary external criterion, such as the physician's judgment of whether there was a change in patient health or function. When sensitivity vs. 1-specificity for hypothetical "tests" corresponding to various cut-offs is graphed, we can calculate the area under the curve (AUC), which represents the "probability

of correctly identifying the improved patient from randomly selected pairs of improved and unimproved patients" (Deyo et al., 1986). The AUC ranges from 0.5, when correct identification of the improved patient is more or less random, to 1.0, when correct identification occurs every time.

As with the time-path diagrams, the visual nature of ROC curves may appeal to some people. In addition, unlike SRM, ES, and CSRM, ROC curves reflect the ability of the instrument to identify those who remain stable since the curves measure specificity as well as sensitivity to change. However, because the external criterion must be dichotomous (Deyo & Centor, 1986), separate ROC curves must be constructed to measure responsiveness to improvement and deterioration in the patient's condition.

2.2.8. Minimal Clinically Important Difference (MCID) (Jaeschke, Singer, & Guyatt, 1989)

It is important that an instrument not only detects changes in disease activity, but also that the detectable changes are clinically important (Fortin et al., 1995). Therefore, a useful method for measuring responsiveness may be to determine the minimum score difference corresponding with changes in the domain of interest that either are relevant to the assessor, or result in changes to the patient's management (Jaeschke et al., 1989). Jaeschke et al. (1989) used consensus in a group of experts to determine the MCIDs for several different questionnaires. When the score changes were compared to transition scores provided by patients, the authors found that the MCIDs corresponded with small changes in patient function. This approach to assessing responsiveness may be of more interest to clinicians than methods such as ES, SRM, or CSRM, since it provides an explicit guideline for judging the relevance of score changes.

2.2.9. Ability to predict relevant change (Fortin et al., 2000)

When assessing responsiveness, an alternative to measuring the magnitude of instrument score change associated with relevant change in the external criterion is to evaluate the probability that changes in the instrument score correspond with clinically relevant changes. This allows clinicians to understand the significance of score changes without having to refer to a particular cut-off point, which is especially useful when an

instrument is interpreted by a person inexperienced in its use (Jaeschke et al., 1989). This approach was taken in a recent study, in which a technique for flexible polytomous regression (regression involving a discrete dependent variable with more than two values) adapted from one developed by Abrahamowicz and Ramsay (1992) was used to determine the probabilities of three relevant outcomes simultaneously, at given changes in instrument scores (Fortin et al., 2000).

The wide variety of measures available for assessing responsiveness, coupled with the lack of consensus over the most appropriate one to use, has resulted in the responsiveness of the same instrument being evaluated with several different approaches, sometimes within the same study. On one hand, if different techniques all provide the same results, this may increase confidence in the robustness of the findings. However, differences between the methods can affect the responsiveness estimates and occasionally result in inconclusive findings, such as when one method suggests that one instrument is more responsive to change than another while a second method does not (Brunner et al., 1999). However, until responsiveness is operationalized in a uniform manner, it may be best to assess this characteristic with several different techniques that all appear reasonable on conceptual grounds.

2.3. Research on the responsiveness to change of SLAM-R and SLEDAI.

The reliability and validity of both SLEDAI and SLAM-R have been well established. Ward et al. (2000) used pooled time series regression analysis to measure construct validity, defined as the correlation between instrument scores and scores for physician global assessment of activity that were marked on a 15-cm visual analogue scale (VAS). They found statistically significant correlation coefficient estimates of 0.54 for SLAM and 0.52 for SLEDAI. Their study population consisted of 23 patients seen every two weeks over a 40-week period. Petri et al. (1992) measured both validity and reliability of SLEDAI and two other SLE disease activity measures in a prospective cohort containing 150 patients. Validity was evaluated with the Pearson correlation between instrument and physician global assessment scores for all patients. Reliability was assessed with a nested analysis of variance on instrument scores for six patients who were each seen twice by nine

physicians. The three factors included in the nested ANOVA were subject, visit, and physician or measurement error. Although the test-retest and inter-rater reliability seemed lower for SLEDAI than for the other two indices, all three instruments appeared to have high validity. Liang et al. (1989) evaluated SLAM, SLEDAI, BILAG, and three other instruments in a group of 25 patients each seen twice by two physicians. They also used the nested ANOVA, but in addition to measuring test-retest and inter-rater reliability, they used the technique to assess convergent validity between instruments. They found that SLAM, SLEDAI, and BILAG exhibited the best inter-rater and test-retest reliability, and that all instruments demonstrated good convergent validity.

In comparison, research on the responsiveness to change of SLAM-R and SLEDAI in SLE activity has produced conflicting results. Gladman et al. (1994) found, in a small study on SLEDAI, SLAM, and BILAG, that only SLEDAI was responsive to changes in disease activity. Eight physicians completed SLAM, SLEDAI, and BILAG, using chart data from three clinic visits by each of eight patients. The responsiveness assessment consisted of a four-way analysis of variance that included visit, patient, order, and assessor as the factors. Of the three instruments, significant variation between visits was seen only for SLEDAI. However, by including "visit" as one of the factors, the investigators assumed implicitly that change in disease activity between visits would be the same for every patient, which is clearly refuted by the first table in the article describing their study. Therefore, these results may not be valid.

In contrast, Fortin et al. (2000) separated patients according to the type of change perceived by the physician at each visit. They calculated ES, SRM, and CSRM, and plotted ROC curves for each subset of observations, and found that for both relevant improvement and deterioration, SLAM-R was systematically more sensitive to change than SLEDAI. However, although they used repeated measures of patients, they did not pool the responsiveness estimates over all the visits. Also, they did not use statistical inference on their estimates.

Brunner et al. (1998) studied the sensitivity of SLEDAI, BILAG, and SLAM to changes in 35 patients with childhood-onset SLE. The researchers abstracted data from the patient charts and, for each instrument, recorded disease activity at diagnosis, at 6 months after, at the first flare, and at 6 months after the flare. Instead of using any global assessment

of changes in disease activity as the external criterion, they assumed that disease activity decreased between time of diagnosis and six months after, and between the first flare and six months after, and that it increased between six months after diagnosis and first flare. The researchers calculated the effect size (ES), effect size index (ESI), standardized response mean (SRM), responsiveness statistic (RS), and relative efficiency index (REI) for each pair of visits. Although no instrument seemed consistently more responsive than the others, the researchers felt SLAM tended to be less responsive than either SLEDAI or BILAG. However, this conclusion was based mostly on comparisons of point estimates of responsiveness. The 95% CIs that were generated for SRM were based on the assumption that SRM is normally distributed, which is questionable given that the ratio of two normally distributed estimates does not have a normal distribution. Moreover, the CIs of SLEDAI and SLAM both overlapped and included zero. Another problem with this study is that the researchers did not verify their assumption that disease activity changed in certain directions over given time periods. A heterogeneous mix of changes of disease activity could have occurred, resulting in underestimation of responsiveness, and possibly biased comparisons between different SLE activity measures.

Ward et al. (2000) evaluated 23 patients every two weeks for up to 40 weeks, and collected scores for SLAM, SLEDAI, BILAG, and two other instruments, as well as physician and patient global assessments of disease activity rated on a 15-cm visual analogue scale (VAS). For each patient they then selected the pairs of consecutive visits exhibiting the greatest changes in physician and patient VAS scores, regardless of direction, and calculated the SRMs of the absolute change in instrument score between the corresponding visits. The results suggest that SLAM was more responsive to changes in either physician or patient assessments (SRM = 0.62 and 0.61 respectively) than SLEDAI, which was responsive to changes in physician but not patient assessments (SRM = 0.48 and –0.01 respectively). However, because their sample size was so small they chose not to compare the instruments statistically. They also combined visits showing improvement with visits showing deterioration, but it might have been better to analyze the two outcomes separately, as it is possible that the responsiveness of these instruments depended on the direction of relevant change being assessed.

The studies discussed above all concentrated on the magnitude of instrument score changes when changes in disease activity were assumed to occur. An alternate approach to analyzing responsiveness is to determine how score changes should be interpreted. Specifically, it would be useful to determine the cut-offs such that a score change larger than the cut-off may be interpreted as being important. Again, various methods have been employed.

For SLEDAI, Gladman et al. (2000) suggested score changes of greater than +3 and less than -3 points as signifiers of relevant increase and decrease of disease activity, respectively. They based their recommendations on the median score changes seen in five clinician-defined categories of disease activity, two of which involve relevant improvement or worsening. However, because they do not provide ranges of score changes observed in each category, we cannot tell how likely it is that misclassification will occur if we use these guidelines.

The predictive power of SLAM-R and SLEDAI was measured in a study by Fortin et al. (2000). Three physicians completed SLAM-R and SLEDAI for 95 patients over five consecutive monthly visits. At each visit the physicians recorded a transition score, i.e. whether they felt the patient had improved, stayed the same, or worsened. The researchers grouped the observations by these transition categories, and for every pair of consecutive visits, they used polytomous regression to estimate simultaneously the probability of occurrence of all three outcomes as a non-parametric function of the score changes, based on the method developed by Abrahamowicz and Ramsay (1992). The curves for the outcomes were well-separated for SLAM-R, except for the first pair of visits, but were less distinct for SLEDAI, suggesting that less misclassification would occur if changes in disease activity were judged by SLAM-R scores. For SLAM-R, they concluded that changes of about 5 points in either direction signified relevant change. Because of the complexity of the method, though, the models from each pair of visits could not be aggregated.

A limitation common to most of these studies is that they do not address the responsiveness of SLAM-R and SLEDAI to changes in SLE activity changes relevant to patients. Patient judgment of SLE activity has been receiving greater attention recently, especially for sample size calculations in clinical trials (Naylor & Llewellyn-Thomas, 1994), and as indicators of lack of compliance with treatment regimens (Neville et al., 2000).

Discordance between patient and physician assessments of disease activity at a single point in time has been described previously in both rheumatoid arthritis and SLE patients (Kwoh et al., 1992; Neville et al., 2000), and it is possible that the same phenomenon occurs when changes in disease activity are rated. One possible reason for the discordance is that patients might consider factors such as emotional state and personal experience in their rating of disease activity, while physicians might consider physical findings, laboratory results, and patient-reported effects (Neville et al., 2000). This disagreement means that clinician opinion should not be considered as an accurate proxy for patient opinion of progress.

Both intra- and inter-patient ratings have been proposed for ascertaining relevant changes in disease activity. Intra-patient ratings involve asking the patient whether he or she feels that disease activity has changed since the last evaluation, and establishing thresholds for perceived changes. Inter-patient ratings involve using a standardized protocol for discussions between patients about their health, and comparisons by the patients afterwards between their own disease activity and that of their conversation partners (Redelmeier & Lorig, 1993). Although inter-patient ratings eliminate effects of faulty memories of disease activity, intra-patient ratings are more feasible and realistic. Although follow-up is necessary for intra-patient ratings, this can be done through regular clinic visits. In comparison, inter-patient comparisons necessitate the simultaneous presence of large groups of patients (Wright, 1996). Also, a tendency towards optimism has been observed in inter-patient ratings in a group of rheumatoid arthritis patients (Wells et al., 1993). Finally, over the course of a treatment, patients will most likely evaluate their progress in terms of remembered disease activity. Using intra-patient comparisons to establish threshold relevant score changes may thus generate a more realistic profile of how a patient actually rates the importance of increases and decreases in their disease activity.

Another area that requires further study is the impact of changes in specific organ systems on the relevance of score changes. Indeed, original scoring systems for SLAM-R and SLEDAI were developed based on the instruments' abilities to discriminate between patients with different levels of disease activity, at a fixed point in time (Bombardier et al., 1992). It is therefore not obvious if the same scoring is able to optimally reflect changes over time in the activity of the same patient. For example, a change of three points due to rash and alopecia may not be viewed in the same way as a three-point change due to joint pain or

renal activity. In addition, activity in certain organs has been shown to explain some of the disagreement between physician and patient ratings of disease activity. A study by Neville et al. (2000) found greater discordance between patient and physician ratings when patients had more kidney involvement. This may be because renal involvement can be detected through laboratory tests before it directly affects the patient (Neville et al., 2000). If instruments place emphasis on organ systems or manifestations relevant to physicians but less so to patients, these instruments may show lower responsiveness to patient assessment of relevant change in disease activity than to physician assessment.

In addition to providing more detailed insight into responsiveness of SLE activity measures, the identification of organ-specific changes with a strong impact on patients' and/or physicians' global assessment of change may help us in understanding the differences in the results of studies carried out on different populations. Several researchers have observed different patterns of organ involvement between ethnic and age groups, and between centres (Thumboo et al., 1998; Brunner et al., 1999; Symmons et al., 1988). If organ-specific changes have dissimilar impacts on the global impression of change in the patient's status, these differences may result in the variation of the levels of responsiveness observed in different populations.

In summary, previous research on the responsiveness of SLAM-R and SLEDAI has limitations that need to be addressed before we can interpret the relevance of score changes. Firstly, because SLE is infrequent, most studies relied on small sample sizes that made estimation of responsiveness quite imprecise. Secondly, in spite of some recent findings suggesting that SLAM-R may be systematically more responsive than SLEDAI, the uncertainty around the difference in their responsiveness was never formally quantified. Thirdly, there is continuing disagreement over methodology. Finally, few studies have looked at the relevance to patients of changes in lupus disease activity. In the next three chapters, I will attempt to address some of these limitations and evaluate the responsiveness of SLAM-R and SLEDAI, when physician judgment and when patient judgment are the external criteria. In addition, I will examine the effect of type of organ involvement on the relevance of changes in disease activity.

The data I have used in this analysis was obtained from an ongoing Canadian multicentre randomized controlled trial on SLE patients, the Study of Methotrexate in Lupus

Erythematosus (SMILE). This analysis will aid the investigators in their interpretation of the trial results. Ethical approval of the original study was granted by the Research Ethics Boards of each participating centre, and no new data was collected from the study participants for this analysis. The documentation showing board approval is included in Appendix B, and examples of consent forms for study participants can be found in Appendix C.

2.4. Study Objectives

1) To estimate the predictive power of changes in the total scores of SLAM-R and SLEDAI.
2) To evaluate the predictive power of changes in organ-system-based subscores of SLAM-R and SLEDAI.
3) To compare the responsiveness of SLAM-R and SLEDAI to changes in disease activity which are clinically relevant to the physician and the patient.

CHAPTER 3: MANUSCRIPT OF FIRST ARTICLE TO BE SUBMITTED FOR PUBLICATION

Preface to Article 1

Four methodological limitations are frequently encountered in the literature concerning the responsiveness of SLE activity instruments to important change in disease activity:

1) Most studies on SLAM-R and SLEDAI responsiveness measure their sensitivity to relevant change. However, this property does not by itself let us predict whether or not a score change of a particular magnitude will actually be seen as meaningful. This aspect of SLAM-R and SLEDAI responsiveness has been addressed in only one previous study (Fortin et al., 2000).

2) Because the sampling distributions of most responsiveness statistics are unknown, the precision of the estimates cannot be calculated analytically. Consequently, only one study in the literature includes 95% confidence intervals for any of the responsiveness index estimates, and this study relies on the assumption of an underlying parametric distribution (Brunner et al., 1999).

3) Although there is growing interest in the use of patient assessment of change as a study outcome, only one study on SLAM-R and SLEDAI responsiveness has examined the behaviour of these instruments with respect to global change important to patients (Ward et al., 2000). If sample sizes in future studies are to be based on the change in disease activity meaningful to patients, it will be necessary to establish first what are the corresponding magnitudes of SLAM-R and SLEDAI score changes.

4) The SLAM-R and SLEDAI responsiveness statistics are often estimated from data sets containing one score change per patient (Ward et al., 2000). When two or more score changes were included per patient, the estimates are usually calculated for subsets of observations grouped by visit (Brunner et al., 1999; Fortin et al., 2000). Because SLE is such a rare disease, recruitment of large numbers of patients can be difficult. In this context, repeated measures on the same patients may be useful for enhancing the precision of the estimates.

To address these issues, in the first paper we have analyzed data from a study population that was assessed monthly, for up to eighteen months. As well as measuring sensitivity to change of SLAM-R and SLEDAI, we have evaluated the strength of the association between change in instrument scores, and changes in disease activity that are important to physicians and patients.

ARTICLE ONE

Comparison of the responsiveness of SLAM-R and SLEDAI to changes in SLE activity relevant to patients and physicians

Erika Chang[1, 3], BSc, Michal Abrahamowicz[1, 2], PhD, Diane Ferland[2, 4], RN, BScN, Paul R. Fortin[3, 4], MD, MPH, for CaNIOS investigators[4]

[1] Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada.

[2] Division of Clinical Epidemiology, The Montreal General Hospital, Montreal, Quebec, Canada.

[3] The Arthritis Center of Excellence, the Division of Rheumatology, The Toronto Western Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada.

[4] CaNIOS investigators are listed in Appendix E.

Supported in part by grants from the Canadian Arthritis Network and from The Arthritis Society (Grant # 95072). Dr. Fortin is a Senior Research Scholar of The Arthritis Society and Director of Clinical Research, Arthritis Center of Excellence of the University Health Network. Dr. Abrahamowicz is supported by a Health Scientist Award from the Medical Research Council of Canada.

Address reprint requests to Michal Abrahamowicz, Division of Clinical Epidemiology, Montreal General Hospital, 1650 Cedar Ave., Montreal, Quebec, Canada, H3G 1A4

Conflicts of interest declared: none

ABSTRACT

Background: Both the revised Systemic Lupus Activity Measure (SLAM-R) and the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) are valid and reliable measures of SLE disease activity. However, more study of their responsiveness, especially that to changes in disease activity relevant to patients, is needed,.

Objective: To measure the responsiveness of SLAM-R and SLEDAI to disease activity changes relevant to physicians and patients.

Methods: Eighty-six patients participating in the Study of Methotrexate in Lupus Erythematosus (SMILE) were evaluated monthly for up to 18 months. At each visit, the physician completed SLAM-R and SLEDAI. Also, patients and physicians assessed, independently of each other, whether the disease activity had improved, worsened, or remained unchanged since the previous visit. Effect size (ES), standardized response mean (SRM), and control SRM (CSRM) were calculated separately for each response category, with 95% CIs. The probability of relevant change in disease activity was modelled using logistic regression, as a function of instrument score change, and the regression parameters were converted into ORs.

Results: Eighty-one patients contributed 754 observations for the analysis of responsiveness to physician-reported changes, and 755 observations for the analysis of responsiveness to patient-reported changes. When the response categories were based on physician evaluation, the CSRM for SLAM-R and SLEDAI were –0.60 vs. –0.36 respectively for visits showing improvement, 0.02 vs. –0.01 respectively for visits showing no change, and 0.79 vs. 0.39 for visits showing deterioration. The 95% CIs for SLAM-R and SLEDAI excluded zero when either improvement or deterioration were detected, indicating that both instruments were responsive to physicians. Similar results were seen for ES and SRM. When the response categories were based on patient judgment, the CSRM for SLAM-R and SLEDAI were –0.32 vs. –0.13 respectively for visits showing improvement, -0.09 vs. –0.01 for visits showing no change, and 0.39 vs. 0.05 for visits showing deterioration. Only the 95% CIs for SLAM-R excluded zero when improvement or deterioration were detected, indicating that only SLAM-

R was responsive to patients. Again, similar results were found for ES and SRM. As well, the ORs for association between instrument score change and relevant change in disease activity were greater for SLAM-R than for SLEDAI, and greater for physicians than for patients. The 95% CIs of the ORs for improvement and deterioration relevant to patients included zero when change in SLEDAI was modelled, and excluded zero for all other ORs.

Conclusion: Both SLAM-R and SLEDAI are responsive to changes in SLE disease activity important to physicians, but SLAM-R appears more responsive. Only SLAM-R is responsive to changes important to patients. These differences may result from the inclusion of subjective SLE manifestations in SLAM-R but not SLEDAI.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic, autoimmune disease characterized both by unpredictable flares and remissions, and by the great variety of manifestations that may be observed in different patients and over time. Although it can affect people of all ages and both sexes, it primarily occurs in women of childbearing age (Hochberg, 1993). Although it was frequently fatal in the past, new treatment practices and the identification of milder cases have resulted in greatly improved 5- and 10-year survival rates . As a consequence, there is increasing interest in studying outcomes other than death, such as disease activity. However, the heterogeneous nature of SLE manifestations has complicated attempts to quantify its activity.

Currently, over 60 different measures of SLE activity with varying psychometric properties exist (Liang et al., 1988a). Two that are commonly used in North America are the revised SLE activity measure (SLAM-R) (Liang et al., 1989; revised in 1991) and the SLE Disease Activity Index (SLEDAI) (Bombardier et al., 1992). These have both been shown to be valid and reliable (Petri et al., 1992; Liang et al., 1989; Ward et al., 2000; Bombardier et al., 1992), but studies on their responsiveness to changes in SLE activity have produced conflicting results. Gladman et al. (1994) and Brunner et al. (1999) found SLEDAI to be more sensitive to change than SLAM-R, while Ward et al. (2000) and Fortin et al. (2000) found SLAM-R to be more responsive than SLEDAI. This inconsistency may have been due in part to differences in analytical approaches. Moreover, because the sampling distributions of the responsiveness measures used are unknown, the studies were not able to assess to what extent the observed differences between the responsiveness of the two instruments might have been due simply to chance.

The most common method of assessing responsiveness is by measuring the magnitude of instrument score change that corresponds to clinically relevant increases and decreases in disease activity (Petri, Genovese, Engle, & Hochberg, 1991; Brunner et al., 1999; Ward et al., 2000; Fortin et al., 2000; Gladman et al., 2000). However, this does not allow us to evaluate the likelihood that a given score change will be perceived as important. An alternative approach used by Fortin et al. (2000) involves modelling the probability of perceived relevant change as a function of the instrument score changes.

Another limitation common to most studies on the responsiveness of SLAM-R and SLEDAI is that they concentrate on physician rather than patient evaluations of change in disease activity. There is evidence that patients and physicians do not agree on assessments of disease activity at a single point in time (Neville et al., 2000). It is therefore possible that they also differ in their perceptions of the importance of changes in activity. Because these disagreements may lead to lower compliance with treatment, it is important to understand to what extent observed changes in objectively scored SLE activity measures are able to reflect patients' subjective judgments. This may, in turn, enhance the relevance of results of clinical studies that use these instruments as outcome measures.

Our objectives in this study are: 1) to evaluate the responsiveness of SLAM-R and SLEDAI, defined as both their sensitivity to change and predictive ability; and 2) to assess the precision of these estimates to enable formal statistical inference about the differences between the responsiveness of the particular instruments.

A major problem for many SLE studies is that the rareness of the disease makes recruitment of large numbers of patients unfeasible, resulting in limited precision and low statistical power of the estimates. We have overcome this problem by incorporating repeated measures of patients in our analysis, while using published techniques to adjust for intra-patient dependence of observations.

PATIENTS AND METHODS

Design for the original SMILE study. A secondary analysis was performed on blinded data obtained from an ongoing Canadian multi-centre randomized controlled trial involving SLE patients. Approval for the RCT was granted by the research ethics boards of all participating centres, and informed consent was given by enrolled patients. The participants were blindly assigned to one of two treatment arms for one year and followed up through monthly visits. After the blinded portion of the trial, they were then invited to take methotrexate for six months unblinded. The patients continued to visit the outpatient clinics each month during the open-level phase of the trial. At the time of enrolment, because of the inclusion criteria, all study participants had a SLAM-R score of 8 or more.

All analyses reported in this article were blinded with respect to the randomization status of the patients since the study is ongoing at this time.

Data collection. At each follow-up visit, both the patient and the physician evaluating the patient recorded, independently of each other, their responses to the question "Over the past month, has the lupus been...," on a five-point transition scale ranging from "much better" to "much worse". After this, the physician completed the non-laboratory components of SLAM-R and SLEDAI. To ensure that the physician attending the patient was blinded with respect to the patient's treatment status, another physician reviewed all laboratory measures of SLAM-R and SLEDAI. Data were entered in Medlog (Medlog, 1995)and analyzed with SAS v6.12 (SAS, 1996). The visits for which all instrument scores had been obtained were used to assess responsiveness to change in the present study.

Statistical analysis of the sensitivity to clinically relevant changes in disease activity. "Clinically relevant change" was defined as any change reported by the physician or the patient on the transition scales (Fortin et al., 2000). Because few respondents reported the lupus being "much better" or "much worse" at any of the visits, the two categories were combined with "better" and "worse" respectively, making a total of three transition categories—"better", "same", and "worse".

The sensitivity of SLAM-R and SLEDAI to important change was measured with effect size (ES) (Kazis et al., 1989), standardized response mean (SRM) (Liang et al., 1990), and control-standardized response mean (CSRM) (Guyatt et al., 1987). Because individual patients differed with respect to the presence and/or direction of the reported changes in disease activity, visits were classified by transition category, and the responsiveness measures were calculated separately for each of the three subsets of visits.

Since participants were assessed repeatedly and recruited into the study at different points in calendar time, the number of observations contributed by each patient varied. In addition, it was possible that characteristics unique to each patient influenced either the patient's or the physician's perception of the relevance of change in disease activity, inducing dependence between subsequent transition responses for the same patient. Yet, conventional methods for assessing responsiveness to change are not able to account for the dependence of the observations. Therefore, we adopted an approach that allowed us to employ conventional responses while, at the same time, pooling the results from different visits. The general idea

is first to pool together the observations from the repeated visits, and then to use a suitable modification of the bootstrap technique to account for the possible dependence of the observations from the same patients. Two alternative approaches were considered at the step of data pooling. Using the first approach, we first estimated responsiveness indices separately for each pair of subsequent visits, each time using observations on all patients who had scores on both relevant visits. Thus, when estimating each visit-specific responsiveness parameter the independence assumption was met. Then, we estimated the overall responsiveness index as the weighted average of the visit-specific indices, with weights corresponding to the number of patients available for each visit. We will refer to this approach as the "average responsiveness" estimation. The alternative approach, referred to as "pooled responsiveness," consisted of simply pooling all the data from all the visits, including multiple observations from the same patients, and then estimating a single responsiveness index. Because the average responsiveness approach reduces the concern about the dependence of observations, it was considered the primary approach *a priori*.

Because previous studies have been unable to quantify the uncertainty around the estimates of responsiveness, the precision of their reported values of ES, SRM, or CSRM remains unknown, making it difficult to compare results of different studies, across different populations or activity measures. The complexity of assessing precision for these estimates lies in the fact that each responsiveness measure is defined as a ratio of two estimates, resulting in distributions that do not conform with conventional analytical models. When repeated measures are used, matters are further complicated because of the dependence of observations within the same patient. To address these issues, in this study we employ a repeated-measures modification of the bootstrap technique proposed by Abrahamowicz, Fortin, du Berger, et al. (1998). Bootstrapping is a computationally intensive resampling technique that allows for direct quantification of the uncertainty of estimates through generation of a large number of random "mutations" of the original sample, each obtained by sampling with replacement from the original data (Efron & Gong, 1983). The estimation procedure is repeated for each bootstrap sample. Because the distribution of the sample estimates approximates well the unknown "true" distribution of the estimator of interest, it can be used to evaluate precision. For example, 95% confidence intervals are estimated by the interval between the 2.5^{th} and 97.5^{th} percentiles of the observed distribution of the

bootstrap estimates. In the case of independent data, sampling with replacement is applied to individual observations, but in the case of data that are clustered within patients, the sampling unit corresponds to a patient rather than a single data point (Abrahamowicz et al., 1998). Using this approach, we created 1,000 bootstrap samples by sampling with replacement individual patients. Accordingly, if a patient was excluded from a given sample, we eliminated the entire block of observations related to all repeated measures on this patient. This results in a conservative estimate of the variance of our estimates and accounts for the dependence of the within-patient observations.

To provide a comparison between the responsiveness statistics of SLAM-R and SLEDAI, we used bootstrapping to obtain the means and 95% confidence intervals for the differences between the corresponding estimates for each instrument.

Statistical analysis of the association between change in SLAM-R and SLEDAI scores and clinically relevant change. Two of the limitations of measures such as ES, SRM, and CSRM are that their values depend partly on the population being studied, and they provide no indication of how score changes of a particular magnitude should be interpreted. To determine the association between changes in SLAM-R and SLEDAI scores and clinically relevant changes, we modelled the probability of relevant change in disease activity, as assessed by the global transition score by either physician or patient, as a function of the change in score of each instrument. SLAM-R and SLEDAI were assessed in separate models, and the instrument score from the previous visit was included as a covariate, to adjust for possible effects of recent levels of disease activity on the patients' or the physicians' perceptions. Separate analyses were carried out to contrast (i) decreased activity vs. unchanged activity, and (ii) increased activity vs. unchanged activity. The reason for this separation was to avoid assuming *a priori* that the relationship between the change in score and probability of a "relevant change" is the same regardless of the direction of change.

Because the data set consisted of repeated measures of patients, the regression parameters were estimated with the generalized estimating equations (GEE) approach (Liang & Zeger, 1986), using an autoregressive order one (AR(1)) correlation structure (Jennrich & Schluchter, 1986). In this type of structure, it is assumed that the correlation between any two visits is dependent on their proximity, and equals a single coefficient to the power of the

number of visits separating the two of interest. The results were summarized in terms of odds ratios and corresponding 95% confidence intervals, and in terms of estimated probabilities of relevant change being perceived at given changes in instrument scores.

RESULTS

Study population. Eighty-six patients enrolled in SMILE at the time the data were obtained contributed 1125 visits. However, because instrument scores were not recorded at several visits and some patients had not yet returned for follow-up, score changes from approximately 755 pairs of adjacent visits, contributed by 81 patients, were used to assess responsiveness. Baseline characteristics of the patients included in the analysis are shown in Table 1. Patients reported improvement at 275 visits (36.4%), no change in 309 visits (40.9%), and deterioration in 171 visits (22.6%). In comparison, physicians reported improvement, no change, and deterioration at 261 (34.6%), 340 (45.1%), and 153 visits (20.3%), respectively.

-INSERT TABLE 1 HERE-

Sensitivity of instruments to relevant change. Table 2 summarizes the results of the "average responsiveness" and shows the means and 95% confidence intervals of the bootstrap estimates of ES, SRM, and CSRM, for SLAM-R and SLEDAI. The left part of Table 2 analyzes patients' responses and the right part focuses on physicians' responses. As expected, in all instances the responsiveness indices were negative when the lupus appeared to improve, positive when it seemed to worsen, and close to zero when no obvious change had occurred. The 95% CIs of estimates for both SLAM-R and SLEDAI excluded zero when physicians perceived improvement or deterioration, and included zero when the activity remained the same, demonstrating that both instruments were sensitive to change relevant to physicians. On the other hand, only SLAM-R showed consistent statistically significant sensitivity to changes important to patients, although comparison of the corresponding estimates indicates that it is less responsive to patients' than physicians' assessments. In the case of SLEDAI, the 95% CIs for the responsiveness statistics included zero for all three transition categories, suggesting that the instrument was not responsive to changes reported by patients. As expected, when there is true change, ES tends to yield lower estimates than

the two other measures. The reason is that ES assesses mean change relative to the standard deviation of the baseline scores, which is both less relevant and overly conservative compared with SRM and CSRM, which use the standard deviation of changes (Fortin et al., 2000). CSRM tended to yield the largest estimates of the three, which was also expected, given that variation in the score change should be small when no change in disease activity is perceived.

-INSERT TABLE 2 HERE-

Table 3 shows the bootstrap estimates, together with the 95% confidence intervals, for the differences in the corresponding responsiveness measures for SLAM-R and SLEDAI. The difference may be considered statistically significant at the conventional 0.05 significance level if the confidence interval excludes zero. In the cases when true change occurred, point estimates of the differences show that SLAM-R was more responsive than SLEDAI. Although some of the 95% CIs for the differences overlap zero, this overlap is very slight, and the point estimates are consistent with each other.

-INSERT TABLE 3 HERE-

Association between SLAM-R and SLEDAI score changes and clinically relevant change. Results from the GEE generalization of the logistic regression models for repeated measures are shown in Tables 4 and 5. The odds ratios (ORs) for score changes unadjusted and adjusted for previous score are presented, with unadjusted ORs in Table 4 and adjusted ORs in Table 5. These reflect the strength of the association found between change relevant to either the physician or the patient, and change in the score of the instrument being evaluated. Both SLAM-R and SLEDAI appeared to be statistically significant predictors of change relevant to physicians. However, only SLAM-R appeared to have a statistically significant association with patient-reported changes. Furthermore, the association between score change of both instruments, and perceived change, was lower when the assessor was the patient, indicating that both instruments were less responsive to patient judgment of change in disease activity than to that of physicians.

-INSERT TABLES 4 AND 5 HERE-

In Figures 1 to 4, the estimated relationship between instrument score change and the probability of relevant change is depicted. The curves have been calculated from the GEE

regression models in which no adjustment was made for the instrument score of the previous visit. The adjustment did not change the curves materially. On each graph, the X-axis represents the score change of one of the instruments, and the Y-axis represents the estimated probability of either relevant improvement vs. no change, or relevant deterioration vs. no change, as assessed by physicians and patients. The ranges of the SLAM-R and SLEDAI score changes included were based on the ranges of score changes observed for each instrument in our sample. In Figure 1, we see the likelihood of relevant improvement from the previous visit being perceived by physicians and patients, at given score changes for SLAM-R. It can be seen that when the score change is zero, the estimated probability of perceptible decrease in SLE activity, vs. no change, is greater than 0.40 for improvement reported by both physicians and patients. It exceeds 0.50 when SLAM-R decreases by more than two points, signifying that with this score decrease, the probability of improvement being reported is estimated to be greater than the probability of no change being perceived. Also, the curve is steeper for physicians than for patients, reflecting the greater responsiveness of SLAM-R to physician-reported improvement than to patient-reported improvement. A decrease of 5 points is associated with the probability of physicians reporting improvement being two times higher than that of physicians reporting no change. However, for patients, a 5-point decrease leaves the two alternatives equally possible.

-INSERT FIGURE 1 HERE-

When relevant deterioration vs. no change at a given change in SLAM-R is graphed (Figure 2), the estimated probability of either physician or patient perceiving deterioration vs. no change when SLAM-R remains the same from the previous visit is lower, ranging from approximately 0.29 to 0.37. Also, these probabilities do not exceed 0.50 until SLAM-R increases by 5 and 8 points respectively, for physician- and patient-reported deterioration. Again, the line corresponding to the estimated probability of worsening in activity relevant to physicians is steeper than that for patients.

-INSERT FIGURE 2 HERE-

When the estimated probability of perceived improvement is graphed as a function of change in SLEDAI scores (Figure 3), we see that, as in Figure 1, the estimated probabilities for physician- and patient-perceived improvement lie between 0.40 and 0.50 when change in SLEDAI from the previous visit equals zero. Also, a decrease in SLEDAI by about 5 points

corresponds to an estimated probability greater than 0.50 that either patients or physicians will perceive improvement. Again, the line representing physician assessment is steeper than that representing patient assessment, demonstrating that SLEDAI is more responsive to improvement relevant to physicians than it is to improvement reported by patients.

-INSERT FIGURE 3 HERE-

Figure 4 shows the estimated probability of deterioration relevant to patients and physicians, vs. no important change, as functions of change in SLEDAI. As with Figure 2, the estimated values lie between 0.30 and 0.40 when the SLEDAI score remains unchanged from the previous visit. However, the estimated curve for patients is almost completely flat, indicating lack of a relationship between change in SLEDAI and patient perception of increase lupus disease activity. In addition, the estimated probability for relevant worsening of activity, which exceeds 0.50 for physicians when SLEDAI increases by about 15 points, does not exceed 0.50 for patients at any point.

-INSERT FIGURE 4 HERE-

DISCUSSION

We evaluated the ability of SLAM-R and SLEDAI to both detect and correlate with clinically meaningful change. Sensitivity was measured by effect size, standardized response mean, and control-standardized response mean, while predictive value was assessed using the GEE generalization of logistic regression. We found that in all instances, SLAM-R was more sensitive to and more strongly associated with important change, and that of the two instruments, only SLAM-R was responsive to changes detected by patients. Our findings agree with those from a previous study on a different group of patients (Fortin et al., 2000), and with those of Ward et al. (2000).

Some limitations of this study should be mentioned. First, because of our use of logistic regression, we were able to compare only two outcomes simultaneously, and thus could not calculate the true probability that a given score change would be interpreted as relevant. Although we could have compared the probability of no change to relevant change in either direction, we would have had to assume *a priori* that the predictive ability of instrument score changes for improvement and for deterioration were equal. A method simultaneously comparing all three outcomes was employed in another study (Fortin et al.,

2000), but it was computationally intensive and too complex to allow for the use of repeated measures or for the estimation of confidence intervals.

Second, it is possible that participation in the trial from which data in this analysis was obtained affected either patients' or physicians' perceptions of the relevance of changes in disease activity. If trial participation produced a sense of optimism in an assessor, that person might have been more likely than usual to detect decreases in disease activity. This would have led to overestimation of the predictive ability of SLAM-R and SLEDAI for relevant improvement. Such a phenomenon may explain the asymmetry observed between Figures 1 and 2, and between Figures 3 and 4. It may be useful to perform a similar analysis on data from routine clinic visits, to determine whether such an effect had occurred in the RCT.

Third, we did not adjust for inter-dependence of observations within physicians. Because more than one clinician was involved in evaluating SMILE participants at the follow-up visits, it is possible that some of the variation in the responsiveness to change of SLAM-R and SLEDAI may have been caused by differences in physician judgment. However, there were too few participating clinicians for us to assess the effects of this inter-dependence.

Finally, in the interest of maximizing the sample size used in the analyses, we eliminated only those observations missing change in instrument scores, instrument scores from the previous visit, or transition scores for the assessor of interest, i.e. patient transition scores if they were the external criteria and physician transition scores if they were the external criteria. We feel that this should not affect comparability of the results since 751 observations, or approximately 99.5% in each set, overlapped between the two sets.

Our study may be the first to demonstrate statistical significance of the difference in the responsiveness of the two instruments. However, our results differ from those of Gladman et al. (1994), and of Brunner et al. (1999), both of whom found that SLEDAI was more sensitive than SLAM-R to change in disease activity assessed by physicians. These disparities may be explained by the differences in analytical approaches in each case. Gladman et al. (1994) used a four-way analysis of variance to determine the statistical significance of the relationship between level of disease activity at a single point in time, and instrument scores, while adjusting for the effects of patient, observer, and order in which

instruments were completed. Although they purposely selected a heterogeneous group of patients with respect to levels of disease activity and selected three consecutive visits from each patient, they did not account for varying patterns of change between patients. This would have diluted the apparent responsiveness of the instruments. Indeed, a very responsive instrument may yield an ES, SRM, or CSRM very close to zero if these indices are estimated from a group of patients in whom some have decreased while others have increased their activity (Fortin et al., 2000). In the study by Brunner et al. (1999), changes in disease activity levels over three different time intervals were analyzed. The intervals used were time of diagnosis to six months post-diagnosis, six months post-diagnosis to time of first flare, and time of first flare to six months post-flare, and they were chosen under the assumption that SLE activity would change in the same direction for each patient. Because verification of this assumption was not done, it was possible that undetected variations in the change patterns may have biased the responsiveness estimates. For example, because of spontaneous variations in disease activity, it was possible that for some patients, disease activity was actually higher six months post-diagnosis, than at time of diagnosis.

Based on Cohen's "rule of thumb" (1988) for interpreting ES and SRM, in which 0.3 represents small, 0.6 moderate, and 1.0 large responsiveness, SLAM-R showed small to moderate responsiveness and SLEDAI showed small responsiveness to changes in activity relevant to physicians. In comparison, SLAM-R showed small responsiveness to patients, and SLEDAI showed almost no responsiveness at all. It is possible that rather modest values of the responsiveness indices are due to the fact that in this study, patients were seen at regular monthly visits rather than just before and just after administration of a single treatment, and so dramatic changes in disease activity were quite rare.

A similar trend was observed in the predictive value of both instruments. Change in SLAM-R scores was a consistently stronger predictor than change in SLEDAI of relevant change in both directions, as assessed by both physicians and patients, and indeed SLEDAI was borderline non-significant as a predictor for change reported by patients. Figures 1 and 2 suggest that physicians and patients were more likely to report relevant improvement than no change when SLAM-R decreased by more than two points, and more likely to report relevant deterioration than no change when SLAM-R increased by more than five points, for physicians, and ten points, for patients. These results support the findings of Fortin et al.

(2000). As Figures 3 and 4 show, physicians and patients appeared more likely to report improvement than no change when SLEDAI decreased by more than five points. However, only the probability of physicians reporting deterioration rather than no change ever exceeded 0.50, and this was when SLEDAI increased by at least 15 points. The probability of patients reporting deterioration, meanwhile, changed only slightly over the range of SLEDAI score changes. The difference in the results for patients and physicians emphasizes the need to study both perspectives when evaluating instrument responsiveness, especially if we wish to relate instrument score changes to patient assessments of change in future studies.

Dissimilarities in the content and scoring system of SLAM-R and SLEDAI have been proposed as a possible reason for the higher responsiveness of SLAM-R to change reported by patients (Ward et al., 2000; Fortin et al., 2000). SLAM-R includes a severity gradient for organ involvement, and also contains many items measuring subjective manifestations such as pain and fatigue (Liang et al., 1989). In contrast, SLEDAI measures presence versus absence of organ activity, and includes only manifestations that can be confirmed by the physician (Bombardier et al., 1992). For example, a patient with joint pain would need visible signs of inflammation in at least two joints for it to be recorded in SLEDAI, while the joint pain by itself would be reported in SLAM-R. Further study is needed to determine whether, in fact, the more subjective nature of SLAM-R accounts for the difference in responsiveness.

In addition, some manifestations, such as alopecia and proteinuria, were recorded in SLEDAI only if they had not been present at the previous visit (Bombardier et al., 1992). Therefore, if a patient had any of these signs for two or more consecutive visits, the corresponding item would be scored as zero after the first visit. If the manifestation then disappeared at the third visit or after, and the patient or the physician reported an improvement in the lupus activity, the score for this item would remain unchanged. This lack of sensitivity to changes in some organ-specific activity would lower the predictive power of SLEDAI, giving results similar to those reported here.

We evaluated two aspects of responsiveness, namely sensitivity to important change as assessed by either patients or physicians, and ability to predict relevant change. Measures of sensitivity, which are more commonly used (Brunner et al., 1999; Ward et al., 2000), can be problematic because their values depend on the population being studied and thus may not

truly reflect instrument responsiveness. Let us suppose, for example, that there are two groups and one demonstrates more variability in "true" change in disease activity than the other. If an instrument is able to detect these changes and also differentiate between small and large amounts of change, the means of the resulting score changes might be the same for each group, but the variance will be much larger in the more heterogeneous group. Consequently, a sensitivity index such as SRM will be lower in this group. These measures do not indicate to us how likely it is that changes in instrument score will correspond with true changes in disease activity. One way to describe this is by modelling the probability of meaningful change in disease activity, defined as change reported by an assessor, as a function of instrument score change.

In this study, we used repeated measures of the patients to assess responsiveness. This allowed us to increase the precision of our estimates without having to recruit a large group of participants, an advantage for research on a rare disease such as SLE. In addition, by employing bootstrap methods, we were able to compute 95% confidence intervals for the measures of sensitivity, which allowed us to evaluate statistical significance of the differences between responsiveness of the two instruments. Another advantage was that we were able to assess the responsiveness of SLAM-R and SLEDAI to patient perception of change, which turned out to be lower than the responsiveness of the instruments to physician-reported changes.

CONCLUSION

We found that both SLAM-R and SLEDAI were responsive to physician assessment of relevant change in SLE activity, and that of the two, SLAM-R was slightly more so, with the difference often reaching statistical significance. In addition, both instruments tended to be less responsive to changes in disease activity important to patients, and only SLAM-R showed a statistically significant ability to predict changes perceived by them. This difference may be due to the inclusion of more subjective elements in SLAM-R.

Table 1. Patient baseline characteristics

Characteristic	N	Median (IQR) [a]	%
Age (at time of analysis, in years)	81	42.5 (36.3, 50.2)	
Sex (% female)	81		91.4
Marital status (%)	81		
Single			24.7
Married			61.7
Separated, divorced, or widowed			13.5
Educational background (%)	75		
Completed high school (CEGEP 3 in Quebec)			20.3
Baseline scores	81		
SLAM-R		12 (9, 14)	
SLEDAI		10 (6, 14)	
SLICC (n = 80)		1 (0, 2)	
Score changes observed over the follow-up period for analyses using patient-reported change as external criteria[b]	755		
SLAM-R		0 (-2, 2)	
SLEDAI		0 (-2, 2)	
Score changes observed over the follow-up period for analyses using physician-reported change as the external criteria[c]	754		
SLAM-R		0 (-2, 2)	
SLEDAI		0 (-2, 2)	

(a) IQR = Inter-quartile range, i.e. the interval lying between the 25th and 75th percentiles of values for the variable of interest.

(b) Based on the distribution of 755 differences, pooled together from all relevant visits of all 81 patients.

(c) Based on the distribution of 754 differences, pooled together from all relevant visits of all 81 patients.

Table 2. Bootstrap estimates of effect size, standardized response mean, and control-standardized response mean, by patient- and physician-defined transition categories.

Transition category	Responsiveness statistic[†]	Assessor							
		Patients				Physicians			
		SLAM-R		SLEDAI		SLAM-R		SLEDAI	
		Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Decreased SLE activity	ES	-0.26	-0.41, -0.14	-0.06	-0.19, 0.10	-0.40	-0.56, -0.26	-0.24	-0.41, -0.13
	SRM	-0.31	-0.45, -0.19	-0.11	-0.27, 0.04	-0.52	-0.73, -0.36	-0.27	-0.46, -0.13
	CSRM	-0.32	-0.54, -0.16	-0.13	-0.35, 0.03	-0.60	-0.82, -0.43	-0.36	-0.55, -0.19
Unchanged SLE activity	ES	-0.04	-0.21, 0.06	0.03	-0.14, 0.18	0.04	-0.11, 0.18	0.00	-0.15, 0.13
	SRM	-0.09	-0.27, 0.06	-0.01	-0.13, 0.12	0.02	-0.13, 0.14	-0.01	-0.12, 0.12
	CSRM	-0.09	-0.27, 0.06	-0.01	-0.13, 0.12	0.02	-0.13, 0.14	-0.01	-0.12, 0.12
Increased SLE activity	ES	0.29	0.14, 0.54	0.11	-0.11, 0.34	0.47	0.28, 0.71	0.42	0.14, 0.82
	SRM	0.37	0.16, 0.74	0.03	-0.15, 0.21	0.65	0.38, 1.02	0.36	0.13, 0.67
	CSRM	0.39	0.19, 0.65	0.05	-0.15, 0.25	0.79	0.51, 1.12	0.39	0.13, 0.71

[†] ES = effect size; SRM = standardized response mean; CSRM = control-standardized response mean

Table 3. Comparisons of the responsiveness statistic estimates for SLAM-R and SLEDAI.[§]

Transition category	Responsiveness statistic[†]	Assessor			
		Patients		Physicians	
		Mean	95% CI	Mean	95% CI
Decreased activity (improvement)	ES	-0.20	-0.41, -0.04	-0.16	-0.34, 0.02
	SRM	-0.20	-0.35, -0.05	-0.25	-0.46, -0.04
	CSRM	-0.21	-0.44, 0.03	-0.25	-0.52, 0.00
Unchanged activity	ES	-0.06	-0.30, 0.13	0.04	-0.14, 0.23
	SRM	-0.08	-0.28, 0.08	0.03	-0.20, 0.21
	CSRM	-0.08	-0.28, 0.08	0.03	-0.20, 0.21
Increased activity (deterioration)	ES	0.18	-0.07, 0.48	0.05	-0.37, 0.36
	SRM	0.34	0.12, 0.67	0.30	-0.03, 0.68
	CSRM	0.35	0.1, 0.63	0.40	0.04, 0.74

[§] Bootstrap estimates of the mean and 95% CI for SLAM-R responsiveness statistic – SLEDAI responsiveness statistic.

[†] Abbreviations: ES = effect size, SRM = standardized response mean, CSRM = control-standardized response mean.

Table 4. Odds ratios for perceived relevant change associated with a 1-point change in instrument score from the previous visit, not adjusted for the score at the previous visit.

		SLAM			SLEDAI		
Outcome[§]	Observer	OR	95% CI	P-value	OR	95% CI	P-value
Better	Patient	1.07*	1.00, 1.13	0.035	1.02	0.98, 1.05	0.340
	Physician	1.21*	1.13, 1.30	0.000	1.05*	1.02, 1.08	0.004
Worse	Patient	1.08*	1.02, 1.14	0.012	1.00	0.97, 1.04	0.752
	Physician	1.19*	1.10, 1.29	0.000	1.05*	1.01, 1.09	0.008

[§]The OR when outcome = "Better" is for a 1-point instrument score decrease, and the OR when outcome = "Worse" is for a 1-point score increase.

*The point estimate is significant at the $\alpha = 0.05$ level.

Table 5. Odds ratios for perceived relevant change when instrument score changes by one point from previous visit, adjusted for score of previous visit.

			SLAM			SLEDAI		
Outcome[§]	**Observer**	**Variable**	**OR**	**95% CI**	**P-value**	**OR**	**95% CI**	**P-value**
Better	Patient	Change in score	1.14*	1.05, 1.23	0.001	1.03	0.99, 1.07	0.150
		Score at previous visit	1.13*	1.04, 1.22	0.003	1.02	0.99, 1.06	0.203
	Physician	Change in score	1.26*	1.16, 1.36	0.000	1.07*	1.03, 1.10	0.000
		Score at previous visit	1.07*	1.00, 1.15	0.043	1.03	0.99, 1.07	0.144
Worse	Patient	Change in score	1.10*	1.04, 1.17	0.002	1.03	0.99, 1.07	0.148
		Score at previous visit	1.04	0.98, 1.12	0.214	1.04	0.99, 1.10	0.116
	Physician	Change in score	1.28*	1.17, 1.41	0.000	1.08*	1.03, 1.13	0.001
		Score at previous visit	1.14*	1.05, 1.24	0.001	1.06*	1.00, 1.11	0.033

[§]The OR when outcome = "Better" is for a 1-point instrument score decrease, and the OR when outcome = "Worse" is for a 1-point score increase.

*The point estimate is significant at the $\alpha = 0.05$ level

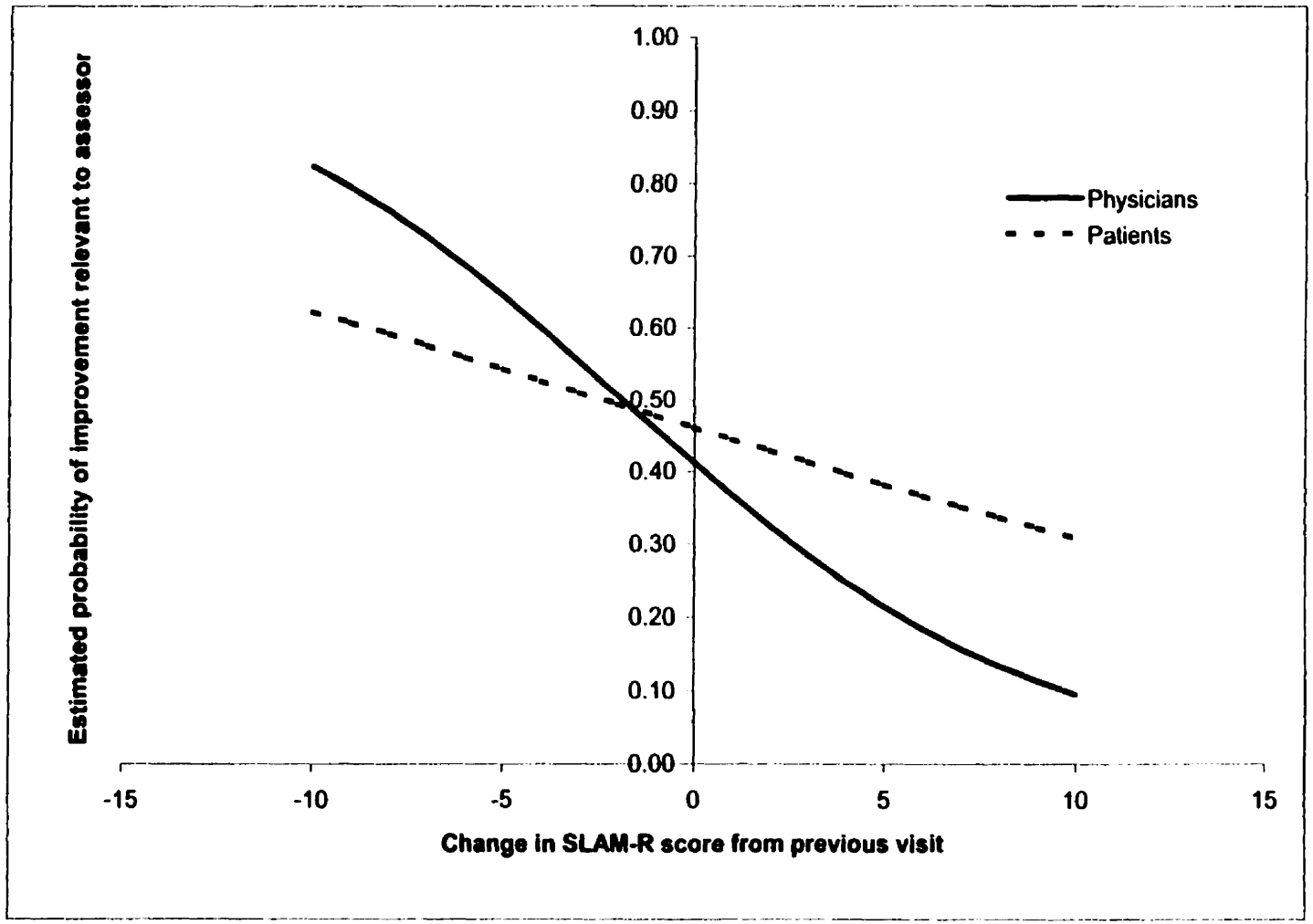

Figure 1. Estimated probability of improvement relevant to assessor, vs. no change, in disease activity at given changes in SLAM-R scores from the previous visit (unadjusted for score at previous visit)

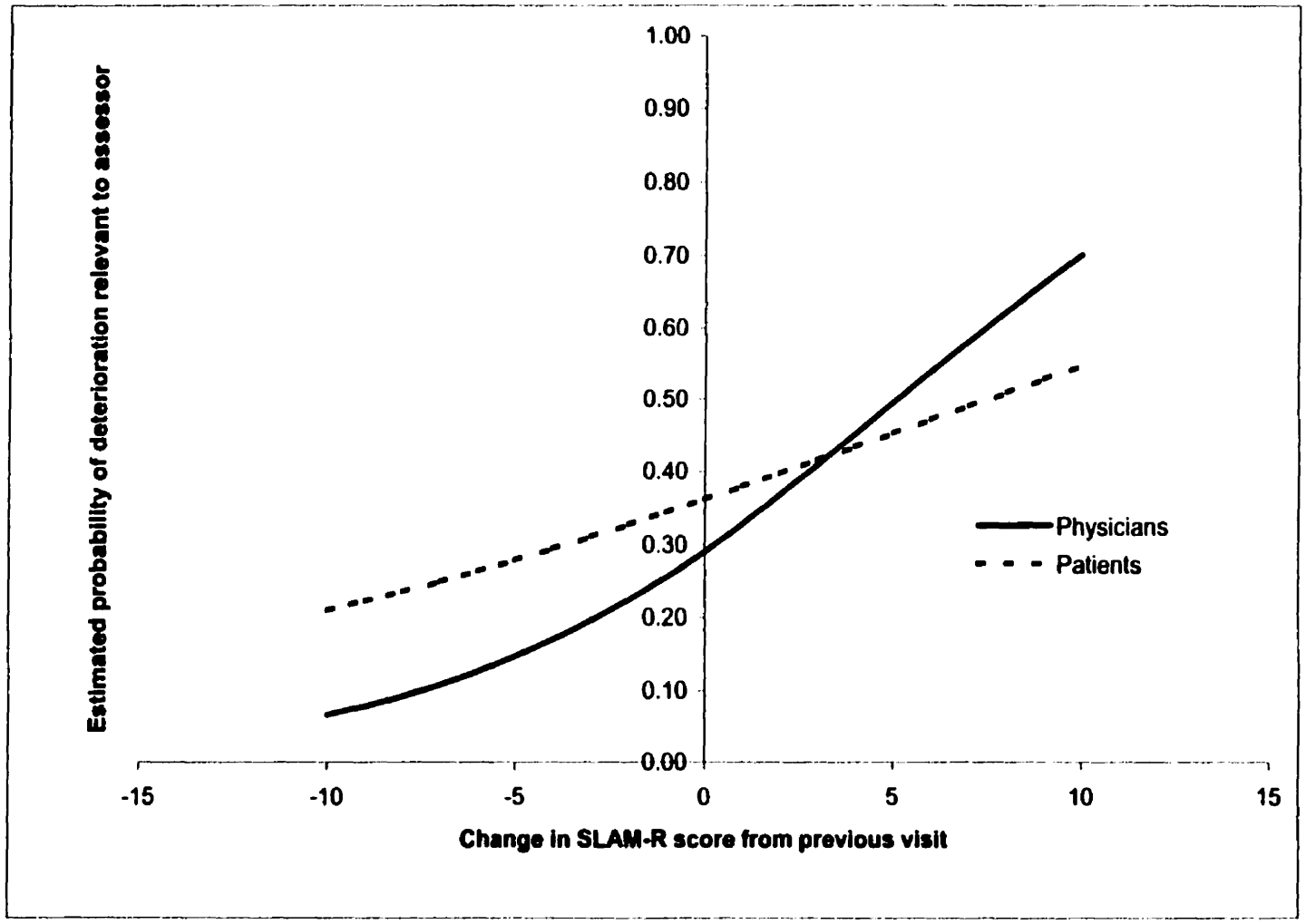

Figure 2. Estimated probability of perceived deterioration, vs. no relevant change, in disease activity at given changes in SLAM-R scores from the previous visit (unadjusted for score at previous visit)

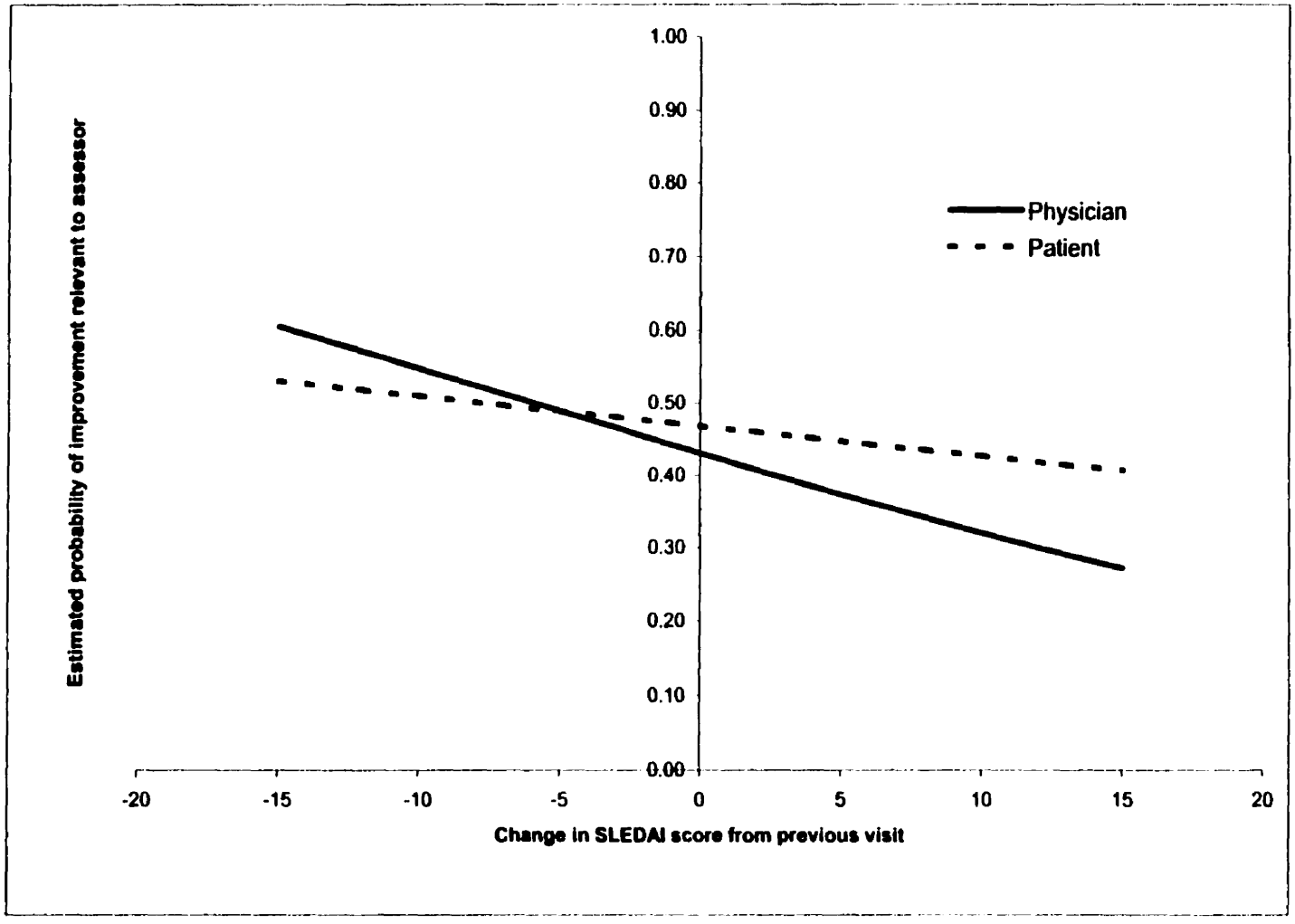

Figure 3. Estimated probability of improvement relevant to assessor, vs. no change, in disease activity at given changes in SLEDAI scores from the previous visit (unadjusted for score at previous visit)

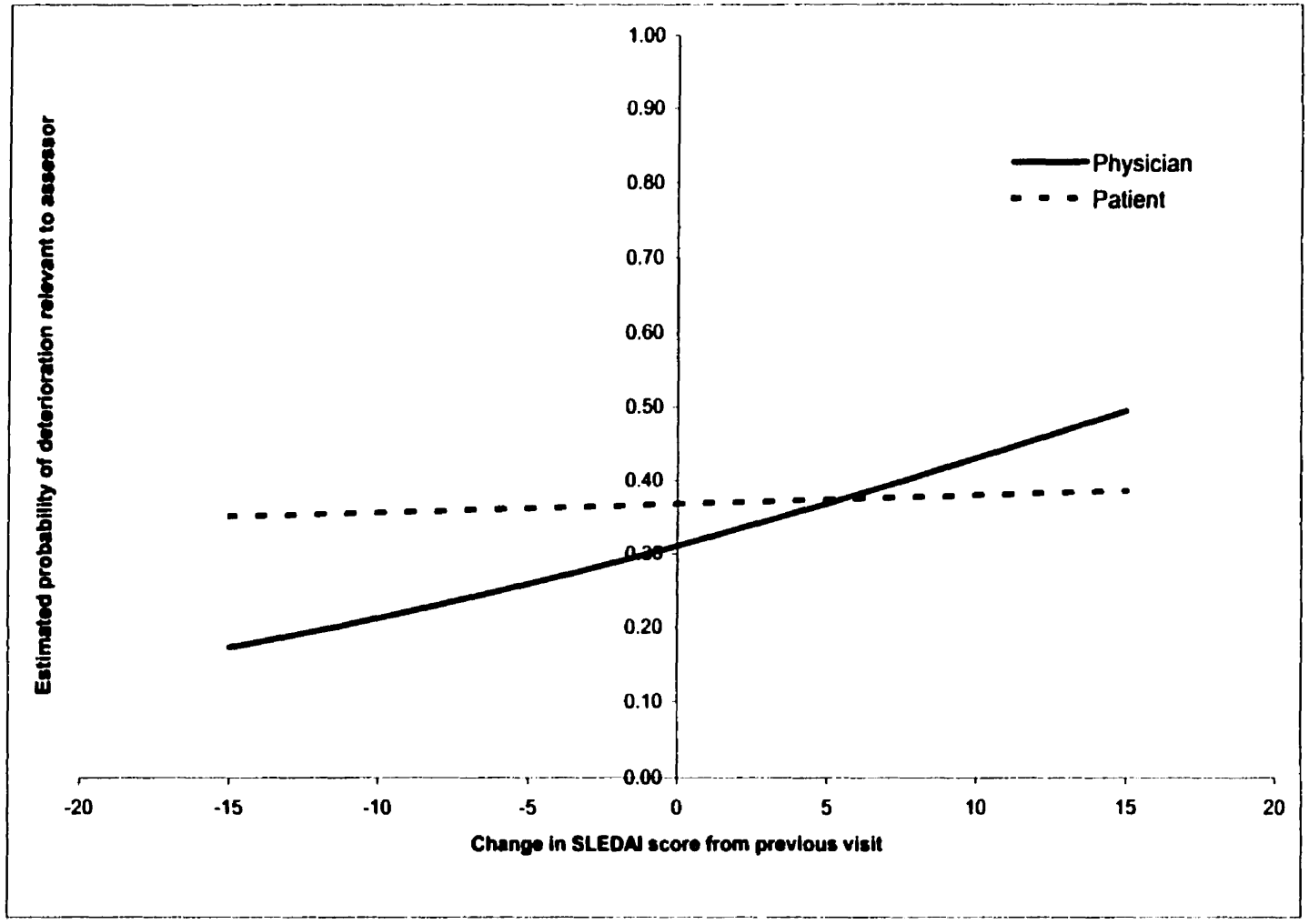

Figure 4. Estimated probability of perceived deterioration, vs. no relevant change, in disease activity at given changes in SLEDAI scores from the previous visit (unadjusted for score at previous visit)

CHAPTER 4: MANUSCRIPT OF SECOND ARTICLE TO BE SUBMITTED FOR PUBLICATION

Preface to Article 2

Article One demonstrated that SLAM-R was more responsive to relevant change than SLEDAI, in terms of both sensitivity to change and predictive value. Furthermore, both instruments appeared more responsive to physician assessments than to patient evaluations of global change in SLE activity. In Article Two, we investigate whether the responsiveness of each instrument is driven by particular organ systems, and, if so, whether differences in organ-specific responsiveness may account for the variations in the overall ability of SLAM-R and SLEDAI to predict changes in disease activity significant to patients and physicians.

To date, no studies have investigated the role of type of organ involvement in the responsiveness of SLE disease activity instruments. It has been suggested, however, that the greater responsiveness of SLAM-R to patient global assessment of change could be attributed to the inclusion of self-reported manifestations in SLAM-R but not SLEDAI (Ward et al., 2000). If this is correct, then score changes in SLAM-R organ system subgroups containing such items might show a stronger relationship to changes in overall SLE activity relevant to patients, than the changes in the corresponding subgroups in SLEDAI.

In addition, other investigators have observed variations in the pattern of organ involvement amongst separate groups of SLE patients (Symmons et al., 1988; Thumboo et al., 1998). It is thus possible that the type of organ activity that changes most frequently depends on the group being studied. If the relevance of change in disease activity depends on the organ system involved, this may explain in part why different studies have produced conflicting findings with respect to the responsiveness of SLAM-R and SLEDAI.

ARTICLE 2

Does the type of organ involvement differentially affect patient and physician assessments of the relevance of change in overall SLE activity?

Erika Chang[1,3], BSc, Michal Abrahamowicz[1,2], PhD, Diane Ferland[2,4], RN, BScN, Paul R. Fortin[3,4], MD, MPH, for CaNIOS investigators[4]

[1] Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada.

[2] Division of Clinical Epidemiology, the Montreal General Hospital, McGill University Health Centre, Montreal, Quebec, Canada.

[3] The Arthritis Center of Excellence, the Division of Rheumatology, The Toronto Western Hospital, University Health Network, University of Toronto, Toronto, Ontario, Canada.

[4] CaNIOS investigators are listed in Appendix E

Supported in part by grants from the Canadian Arthritis Network and from The Arthritis Society (Grant # 95072). Dr. Fortin is a Senior Research Scholar of The Arthritis Society and Director of Clinical Research, the Arthritis Center of Excellence of The University Health Network. Dr. Abrahamowicz is supported by a Health Scientist Award from the Medical Research Council of Canada.

Address reprint requests to Michal Abrahamowicz, Division of Clinical Epidemiology, Montreal General Hospital, 1650 Cedar Ave., Montreal, Quebec, Canada, H3G 1A4

Conflicts of interest declared: none

ABSTRACT

Background: The responsiveness to change in overall SLE activity of SLAM-R and SLEDAI differs. Moreover, both instruments seem more responsive to changes in disease activity relevant to physicians than to patients. Yet, the reasons for these differences remain unclear.

Objectives: To determine 1) if differences in the responsiveness of SLAM-R and SLEDAI are attributable to content; 2) which items contribute to the overall responsiveness of each instrument; and 3) whether responsive items in SLAM-R and SLEDAI differ for physicians and patients.

Methods: Blinded data were obtained from 76 patients participating in the Study of Methotrexate in Lupus Erythematosus (SMILE). At each visit, physicians and patients reported, independently of each other, whether improvement, no change, or deterioration had occurred. Also, physicians evaluated patient SLE activity with SLAM-R and SLEDAI. Individual items in SLAM-R and SLEDAI were grouped by organ system. The GEE approach, which extends logistic regression to handle repeated measurements, was used to determine which organ systems correlate with physician and patient responses with respect to change in disease activity. The outcomes assessed, in separate analyses, were improvement and deterioration from the previous visit, as perceived by patients or physicians.

Results: Seventy-six patients contributed a total of 591 observations. The strongest correlates of improvement relevant to physicians included decreases in constitutional, musculoskeletal, and renal involvement recorded by SLAM-R, and decreases in arthritis, CNS, integument, and renal involvement recorded by SLEDAI. Improvement reported by patients was most strongly associated with decreases in constitutional and musculoskeletal activity recorded in SLAM-R, and in arthritis, CNS, and hematological activity recorded by SLEDAI. Increases in SLAM-R constitutional, integument, musculoskeletal, and reticuloendothelial subscores, and in SLEDAI arthritis, CNS, integument, and fever subscores, appeared relevant to physicians who reported worsening of overall activity in their patients. Patients, on the other hand, found increased SLAM-R constitutional,

gastrointestinal, and musculoskeletal subscores important, but only the integument subscore in SLEDAI appeared to reflect perceived deterioration.

Conclusion: Differences in the relevance of organ-specific changes for patients vs. physicians may account for the differing responsiveness to change relevant to each of SLAM-R and SLEDAI, and may suggest the need for improved communication between lupus patients and their physicians.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease characterized by unpredictable flares and remissions over time. It appears in people of all ages and both sexes, but primarily affects women of childbearing age (Hochberg, 1993). The manifestations vary over time and between individuals.

Improved survival rates of SLE patients have led to increasing interest in studying other outcomes, such as disease activity. Currently, over 60 different measures of SLE activity with varying psychometric properties exist (Liang et al., 1988a). Two measures commonly used in North America, SLAM-R (Liang et al., 1989) and SLEDAI (Bombardier et al., 1992), are valid and reliable (Petri et al., 1992; Liang et al., 1989; Ward et al., 2000), but less is known about their ability to detect relevant change. Recent studies have indicated that, although both are responsive to clinically meaningful change, SLAM-R is slightly more so (Ward et al., 2000; Fortin et al., 2000). Ward et al. (2000) have also demonstrated that of the two, SLAM-R is more sensitive to patient assessments of important change. Our results in the companion paper to this one (Chang, Abrahamowicz, Ferland, & Fortin, 2000) support the findings in those studies. In particular, score changes in SLAM-R but not in SLEDAI were shown to reflect both improvement and deterioration reported by patients. However, the responsiveness of both SLAM-R and SLEDAI was stronger for changes relevant to physicians than to patients.

The inclusion of self-reported symptoms in SLAM-R but not SLEDAI has been suggested as a possible reason for this difference in responsiveness (Ward et al., 2000; Fortin et al., 2000). Because SLE manifestations such as pain and fatigue may be difficult to confirm objectively, their presence may not always be reflected in SLEDAI scores even if it affects patients. Another possible related reason for the disparity in responsiveness of SLAM-R and SLEDAI to patient assessment of change is that the factors governing the relevance of organ involvement differ between patients and physicians. Neville et al. (2000) found discordance in patient and physician global assessment of SLE disease activity at a single point in time, and suggested that this disagreement might have arisen because of differences in the manner in which patients and physicians judged activity levels. If this is also true for evaluations of relevant improvement and deterioration in SLE activity, then an investigation of the importance of change in organ systems included in SLAM-R and

SLEDAI to patients and physicians might reveal how this differs for each, and might provide more insight into why SLAM-R is more able to reflect change relevant to patients.

Our objectives in this study are to determine whether differences in the overall responsiveness of SLAM-R and SLEDAI can be attributed to changes in manifestations of specific organ systems, and, if so, whether the pattern of relevant organ systems differs for physicians and patients.

PATIENTS AND METHODS

Population. A secondary analysis was performed on blinded data obtained from an ongoing Canadian multi-centre randomized controlled trial involving SLE patients. Approval for the RCT was granted by the research ethics boards of all participating centres, informed consent was provided by all recruited patients. Due to the inclusion and exclusion criteria used in this study, at the time of entry, all participants had a SLAM-R score of 8 or more. After allocation to a treatment arm, the study participants were followed up through monthly visits for up to 18 months.

Data collection. At each visit, the physician and the patient recorded, independently of each other, their response to the question "Over the past month, has the lupus been...," on a five-point transition scale that ranged from "Much worse" to "Much better". This physician then completed the non-laboratory components of SLAM-R and SLEDAI. To ensure that the physician attending the patient was blinded with respect to the patient's treatment arm, another physician reviewed all laboratory measures of SLAM-R and SLEDAI. Data were entered in Medlog (Medlog, 1995) and analyzed with SAS version 6.12 (SAS, 1996). The visits for which all instrument scores had been obtained were used in the following analyses.

Definitions. "Clinically relevant change" was defined as any change reported by the physician or the patient on the transition scale. Because few respondents reported the lupus being "much better" or "much worse", these categories were combined with "better" and "worse", respectively, resulting in a total of three transition categories: "better", "same", and "worse".

Classification of SLAM-R and SLEDAI items. With a few exceptions, the grouping of items by organ system, shown in Table 1, was based mostly on the subgroups defined in SLAM-R (Liang et al., 1989), and on the groupings used in the development of SLEDAI

(Bombardier et al., 1992). In SLEDAI, myositis was categorized with central nervous system (CNS) manifestations, since a similar grouping was used in SLAM-R and it was felt that this manifestation reflected CNS involvement. In SLAM-R, erythrocyte sedimentation rate (ESR) was classified separately from the other items, and the other laboratory scores were divided into renal and hematological signs. Although vasculitis appears in both instruments, it was grouped differently in each since it involves skin blood vessels in SLAM-R and organ vessels in SLEDAI. Visual manifestations were classified as a CNS sign in SLEDAI but not in SLAM-R, as they include both anterior and posterior eye activity in SLEDAI but only anterior activity in SLAM-R.

-INSERT TABLE 1 HERE-

Measurement of association between changes in SLAM-R or SLEDAI organ system subscores, and change in overall SLE activity perceived by patients or physicians. We modelled the probabilities of perceived improvement and deterioration as functions of changes in SLAM-R and SLEDAI organ-specific scores. Because repeated measurements from each patient were included in our study sample, the regression parameters were estimated with the generalized estimating equations (GEE) approach to logistic regression (Liang et al., 1986), using an autoregressive order one (AR(1)) correlation structure (Jennrich et al., 1986). SLAM-R and SLEDAI were assessed in different models. Separate analyses were carried out to compare (i) decreased vs. unchanged activity, and (ii) increased vs. unchanged activity. In each model, the binary dependent variable was the relevant change of interest, and the independent variables were the changes in each organ system score of one of the instruments, and the total instrument score from the previous visit:

$$\text{logit}\, p = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \beta_{\text{previous score}} X_{\text{previous score}}$$

where p = the probability of the observer perceiving relevant change, X = change in score for a specific organ subgroup, β = expected change in logit p associated with a 1-point difference in X when all other X remain the same, and k is the total number of organ subgroups included in the model. The instrument score from the previous visit was included as a covariate because it had been shown to be a significant predictor of relevant change in the companion paper. Since all subscore changes were modelled simultaneously, change in one subscore was automatically adjusted for changes in all other subscores.

Because item weights in SLEDAI are pre-set and vary between organ groups according to perceived seriousness of the organ involvement, subscores for SLEDAI organ groups changed only by very specific increments. For example, arthritis, renal, and CNS scores could increase or decrease only by multiples of 4. Although most of the items in the CNS subgroup were assigned a weight of 8, smaller magnitudes of subscore change were made possible by the inclusion of the myositis item. Therefore, when odds ratios and 95% confidence intervals were calculated for SLEDAI item subgroups, the regression parameters and corresponding standard errors were first multiplied by the smallest possible change in the respective organ system scores. Since it was possible for SLAM-R subscores to change by 1-point intervals, the ORs and CIs for their change were calculated directly from the regression parameters.

RESULTS

Study population. Eighty-six patients enrolled in SMILE at the time the data were obtained contributed 1125 visits. Therefore, had instrument and transition scores been recorded at all visits, we would have expected 1039 score changes in our dataset. However, SLAM-R and SLEDAI scores were not collected at follow-up visits after adverse events (n = 13), visits occurring 2 weeks post-recruitment in which drug toxicity was assessed (n = 86), nor when patients were assessed in their homes (n = 25). In addition, scores for some SLAM-R and SLEDAI items not recorded for several other visits. The complete data consisted of 591 score changes contributed by 76 patients. Baseline characteristics of the patients included in the analysis, and change in disease activity scores over follow-up, are shown in Table 2. Patients reported improvement at 212 visits (35.9%), no change in 243 visits (41.1%), and deterioration in 136 visits (23.0%). In comparison, physicians reported improvement, no change, and deterioration at 193 (32.7%), 276 (46.7%), and 122 visits (20.6%), respectively.

-INSERT TABLE 2 HERE-

Tables 3 and 4 show the percent of total visits in which changes in organ subscores were recorded in SLAM-R and SLEDAI, respectively. The tables also include the median changes in each direction for all subscores. SLAM-R and SLEDAI subscore changes were generally small, with the median change in either direction being the smallest possible for each organ system. The proportion of visits in which change in either direction was recorded ranged

from 0.8%, for the eye item in SLAM-R, to 47.9%, for the integument item group in SLAM-R.

-INSERT TABLES 3 AND 4 HERE-

Associations between changes in organ system subscores, and changes relevant to patients and physicians. In Tables 5 to 8, the strength of the association between score change and meaningful change in disease activity is represented by odds ratios (OR). Table 5 shows the ORs for improvement relevant to patients and physicians corresponding to a 1-point decrease in subscores of SLAM-R, after adjustment for changes in other organ systems. The effect of each organ-specific subscore change is adjusted for simultaneous changes in all other organ systems as well as for the previous visit total SLAM-R score. As expected, the point estimates of most of these ORs were greater than one, indicating that the greater the decrease in a given subscore, the more likely it was that physicians and patients would report a relevant lessening of total disease activity. For both patients and physicians, improvements in constitutional and musculoskeletal activity were the factors most strongly associated with important overall improvement, although the associations were statistically significant only for physicians' assessments. For physicians alone, improvement in renal activity was also relevant. To give an example of the clinical implications of these changes, a one-point decrease in the musculoskeletal activity scale score could mean presence of objective inflammation where there had been limited joint function at the previous visit, or of no activity when arthralgia had been reported previously. The interpretation of the same change in an organ system containing two or more items was more complicated, since either a manifestation could disappear and be replaced by another with a lower severity score, or its severity might decrease while the severity of other manifestations remained stable. A comparison between patient and physician point estimates reveals that the ORs for physician-reported improvement tended to be greater, indicating that physicians' overall assessments were slightly more affected by changes in specific organ involvement. Organ-specific activity in which recorded changes had a borderline statistically significant association with perceived change included reticuloendothelial activity, for physicians, and cardiovascular activity, for patients.

-INSERT TABLE 5 HERE-

Table 6 shows the ORs for association between reported improvement in overall disease activity as assessed by either patients or physicians, and decreases in individual SLEDAI organ system subscores, after adjustment for change in other subscores. A decrease in one of the subscores could signify disappearance of manifestations from the previous visit, or also, in the case of integument involvement, the continuing presence of activity in that organ system. Because it was only possible for SLEDAI subscores to change by intervals corresponding to pre-specified weights of individual items, the ORs in Table 6 correspond to the smallest subscore decrease possible for that group of items. These ORs were of roughly the same magnitude as those for one-point decreases in the corresponding SLAM-R subscores. Organ involvement in which change was relevant to patients, in terms of statistical significance, included arthritis, the central nervous system (CNS), and hematological activity. In comparison, changes important to physicians included those in arthritis, the CNS, the integument, and the renal system.

-INSERT TABLE 6 HERE-

Tables 7 and 8 focus on increases in SLE activity. Table 7 shows the ORs for the association between reported overall increases in SLE activity, and one-point increases in the score of individual SLAM-R organ systems after adjustment for change in all other organ systems. The one-point score increase indicated either onset of a manifestation, or increase in its severity. Here, an OR higher than 1 indicated that as an organ system subscore increased, the assessor was more likely to find relevant worsening of activity. For both patients and physicians, there were statistically significant and borderline significant associations. The changes most strongly associated with patient-reported increase in activity were related to greater gastrointestinal activity, while constitutional, musculoskeletal, and reticuloendothelial activity were borderline significant. The OR for change in hematological activity was statistically significant, but in an unexpected manner, since it suggests that a one-point increase in activity in this system would be associated with a lower, not higher, probability of patient-reported deterioration. Physicians found most important the onset or worsening of constitutional, integument, musculoskeletal, renal, and reticuloendothelial involvement.

-INSERT TABLE 7 HERE-

In comparison to the findings for SLAM-R, increases in only very few of the SLEDAI organ system subscores were systematically associated with worsening of overall SLE activity reported by patients (Table 8). Only increased integument activity, defined as onset of a rash, alopecia, or mucous membrane ulceration, appeared relevant to patients. Physicians, on the other hand, were more likely to report deterioration if the scores increased in the arthritis, central nervous system, integument, or fever subscores. These results suggest that different factors may have influenced the patients and physicians in their assessments of overall worsening of SLE activity.

-INSERT TABLE 8 HERE-

DISCUSSION

This study evaluated the association between organ-specific changes in SLAM-R and SLEDAI scores, and improvement and deterioration in overall SLE activity that were relevant to physicians and patients. We found that the organ-specific responsiveness of SLAM-R and SLEDAI to important changes differed for each instrument. In addition, the relationships between changes in organ involvement and relevant changes in overall disease activity depended on whether the assessor was the patient or the physician, and on whether improvement or deterioration was reported. In SLAM-R, the only significant or borderline significant predictors of relevant change were alterations in constitutional symptoms, which included fever, fatigue, and weight loss, and changes in musculoskeletal involvement. Other organ changes were found to have statistically significant or borderline significant effects only in a subset of analyses. They included reticuloendothelial changes when evaluated by clinicians; hematological changes, either when the patient perceived deterioration or when the physician perceived improvement; and gastrointestinal changes when the patient felt that relevant deterioration had occurred. None of the organ system changes recorded in SLEDAI were consistent predictors of change. Those which were sometimes significant included musculoskeletal or central nervous system changes, except when deterioration was perceived by patients; hematological changes when improvement was reported by patients; changes in integumentary involvement, except when improvement was observed by patients; changes in renal activity, when improvement was reported by clinicians; and changes in temperature, when deterioration was observed by physicians. Overall, there did not seem to be any type of

change in organ activity that was consistently relevant regardless of assessor, SLE activity instrument, and reported direction of change. However, it did appear that, compared to patients, physicians' overall assessments of the changes in SLE activity showed more frequent and somewhat stronger associations with organ-specific changes.

Although some organ systems showed statistically significant associations with changes relevant to both physicians and patients, others showed associations with change important to one only, or in one direction of change but not the other. It is possible that the factors influencing the relevance of a given change in organ involvement may depend partly on the assessor and the direction of change. For example, whereas changes in constitutional and musculoskeletal activity were relevant to both patients and physicians, increased gastrointestinal involvement recorded in SLAM-R exhibited a strong and statistically significant association with patient-reported deterioration only (OR [95% CI] = 2.10 [1.32-3.24]). Since the SLAM-R definition of GI tract activity includes subjective elements such as the amount and limiting effect of pain experienced by the patient, and includes an objective component for severe activity only, this association suggests that the importance to patients of changes in activity in certain organs is attributable in part to the immediate impact of these changes on their lives. Changes in constitutional and musculoskeletal involvement would be relevant for the same reason, since joint pain and constitutional symptoms such as fatigue would limit mobility and independence. Conversely, relatively asymptomatic manifestations, such as those involving the renal and cardiovascular systems, would be expected to show less effect on patient assessment of change in overall disease activity. In comparison to patients, physicians seemed less influenced in their assessments by subjective factors such as pain. However, they may have used constitutional symptoms, which are partly subjective, as a gauge of overall disease activity, thereby artificially increasing the relevance of changes in such manifestations. The findings in this study support the argument by Ward, Marx, and Barry (2000) that the comparatively poor responsiveness of SLEDAI to patient assessment of change in disease activity may have been caused by the absence of a subjective component from that instrument. It is also interesting to note that a recent study reported variation in the relevance to patients and physicians of activity in particular organs at a single point in time (Neville et al., 2000).

Although renal involvement has been reported previously as a predictor of discordance between physician and patient assessment of disease activity at a single point in time (Neville et al., 2000), it was not possible in this analysis to evaluate its impact on the relevance of changes in disease activity. As part of the original blinded SMILE protocol, the physicians evaluating global change in the patients did not see laboratory data, which might have revealed the treatment arm of the patients. Thus, the direct relevance of changes in renal and hematological involvement to physicians and patients could not be assessed in this study. It did appear though that change in both renal and hematological scores did sometimes correspond with perceived overall changes. It is possible that activity in these systems produced observable effects on the patient.

Differences in the structure and content of SLAM-R and SLEDAI accounted for some of the disparities in organ-specific responsiveness. As was discussed earlier, manifestations included in SLEDAI are assigned pre-specified weights based on their seriousness, regardless of their actual severity, and when evaluating a given patient, only their presence or absence is scored. In contrast, in SLAM-R, the score of an item depends on the actual severity of the corresponding manifestation in the patient. Because of this gradient, changes in organ-specific severity may have been reflected in SLAM-R but not in SLEDAI. In some organ systems, it appeared that the amount of improvement or deterioration perceived as relevant to either patients or physicians was less than that detectable by SLEDAI. For example, a transition from lack of joint pain to arthralgia, or from objective inflammation to "limited function", would not have qualified as a change in activity on SLEDAI, yet, based on the odds ratio for a one-point change in this item in SLAM-R (OR = 1.47 and OR = 1.80, respectively, for improvement and deterioration reported by physicians), both were important to observers.

The ability to measure clinically relevant changes in disease activity is important when we study patients with SLE, since remissions and flares are an integral part of the disease. Although an instrument may be reliable and valid, this does not guarantee that it can detect important changes over time. If items in the instrument use scales with too few levels, the score may remain the same even when relevant improvement or deterioration occurs. On the other hand, it is possible for a valid and reliable instrument to be a poor predictor of relevant improvement or deterioration, if it uses so many levels that it often records changes

that are irrelevant to the observer. Finally, if changes intended to be measured by a given item are too rare, the item will appear to have poor responsiveness (Fortin et al., 1995).

This study differed from others in this area, in that it modelled how the probability of an observer reporting relevant changes depends on a given change in score. The use of repeated measures increased the precision of estimates, and reduced concerns about low statistical power, typical of SLE studies, which often suffer from small sample sizes. This allowed us to model the effects of changes in organ-specific scores while simultaneously adjusting for changes in all other organs. Some limitations to this analysis should be discussed. First, because the relationship between changes in organ activity scores and perceived changes in disease activity was modelled with a repeated-measures generalization of logistic regression, only two outcomes could be compared at a time. As a result, it was not possible to estimate directly the probabilities of all three relevant outcomes, at a given change in score. The problem could have been surmounted by modelling the association between absolute value of score change and the probability of a binary outcome contrasting any relevant change at all with no change. However, this approach would have involved assuming that the associations between organ-specific changes and change reported by physicians and patients were the same regardless of direction of perceived change, and as the results of this analysis indicate, such an assumption would have been erroneous.

Second, it has been suggested that GEE modelling may result in biased estimates of effect in longitudinal data, if the exposure variable is affected by the outcome variable (Greenland, 1998). In this study, it was possible that increasing disease activity perceived by the physician or the patient might have resulted in more vigilance before and at the following visit. This could have caused either observer to notice and report changes in organ involvement at the next visit that they would not have scored otherwise. It would be difficult, though, to detect and adjust for such a phenomenon.

Third, we should also consider the possibility that participation in the trial might have altered the patients' or the physicians' perception of the relevance of changes in disease activity. For example, the patients might have been more likely than usual to feel they had improved. If this were the case, then we might observe an asymmetry in the responsiveness of SLAM-R and SLEDAI to perceived decreases and increases in disease activity.

Finally, the set of useable observations was greatly reduced by the presence of missing values for some SLAM-R and SLEDAI items. It was possible that visits for which some item scores were missing differed systematically from the rest of the visits, with respect to reported change in overall disease activity or activity in the corresponding organ system. Therefore, additional study on the relevance of changes in organ activity in other patient populations is necessary.

Generalization of these results to all SLE patients and their physicians should be done cautiously. Because we performed a secondary data analysis, our study population was defined by the inclusion and exclusion criteria of the original RCT (Appendix D). This meant that for all patients, the initial SLICC damage score was 15 or less, and the SLAM-R score was 8 or more. Also, patients were excluded if they were unable to comply with the treatment regimen. It is possible that there were systematic differences in perception of the relevance of disease activity by those with more organ damage, or by non-compliers.

The imprecision of some of the odds ratios was due to the lack of activity or change in activity of organ systems such as the eye. Further work is needed, therefore, in other groups of patients with different patterns of change in organ involvement, to increase the precision of some of the estimates of association. It may also be useful to investigate the effects of level of organ damage on the relevance of changes in manifestations, and to determine whether the pattern of change in organ-specific activity is related to patient compliance with treatment.

CONCLUSION

We have found that changes in organ systems thought to be important vary depending on the assessor, the instrument, and the direction of perceived overall change in disease activity. Changes in some of the self-reported manifestations appear to be more meaningful to patients than to physicians. The differences between patient and physician evaluation, and the differences in the relevance to each of change in activity recorded by SLAM-R and SLEDAI, highlight the need for better communication between patients and physicians, and for inclusion of more patient-reported manifestations in instruments if they are to respond better to patient assessments of change in disease activity.

Table 1. Categorization of items into organ system subgroups.

Organ system	Items (SLAM-R)	Maximum subscore	Items (SLEDAI)	Maximum subscore
CNS/ neuromotor	Stroke; seizure; cortical dysfunction; headache; myalgia/myositis	14	Seizure; psychosis; organic brain syndrome; visual disturbances; cranial nerve; lupus headaches; CVA; myositis	60
Cardiovascular	Raynaud's; hypertension; carditis	7	N/a	N/a
Constitutional	Weight loss; fatigue; fever	8	Fever	1
ESR	ESR	3	N/a	N/a
Eye	Cytoid bodies; hemorrhages or episcleritis; papillitis or pseudotumor cerebri	9	N/a (grouped with CNS)	N/a
Gastrointestinal tract	Abdominal pain	3	N/a	N/a
Hematological	Hematocrit; white blood cells; lymphocyte count; platelet count	12	Thrombocytopenia; leucopenia	2
Immunological	N/a	N/a	Increased DNA binding; low complement	4
Integument	Oral/nasal, or periungal erythema, or malar rash, or photosensitive rash or nail fold infarct; alopecia; erythematous, maculopapular rash, or discoid lupus, or lupus profundus, orbilous lesions; vasculitis (leucocytoclastic vasculitis, urticaria, palpable purpura, livedo reticularis, ulcer or panniculitis)	9	New rash; alopecia; mucous membrane ulcers	6
Musculoskeletal	Joint pain	3	Arthritis	4
Other	*Ad hoc* subscale	3	N/a	N/a
Pulmonary	Shortness of breath or pain	3	N/a	N/a

Organ system	Items (SLAM-R)	Maximum subscore	Items (SLEDAI)	Maximum subscore
Renal	Serum creatinine; urine sediment	6	Urinary casts; haematuria, proteinuria; pyuria	16
Reticuloendothelial	Lymphadenopathy; hepato- or splenomegaly	4	N/a	N/a
Serosal	N/a	N/a	Pleurisy; pericarditis	4
Vascular	N/a	N/a	Vasculitis	8
Total		84		105

Table 2. Patient baseline characteristics and change in disease activity score over follow-up

Characteristic	Median (IQR)[§]	%
Age (years)	42.4 (36.2, 50.5)	
Sex (% female)		90.8
Marital status (%)		
Single		26.3
Married		61.8
Separated, Divorced, or Widowed		11.8
Education (n = 70)		
> High school (%)		20.3
Disease activity and damage scores at time of enrollment in study		
SLAM	12 (9, 14)	
SLEDAI	10 (6, 14)	
SLICC	1 (0, 2)	
Change in disease activity scores over follow-up period (N = 591)[a]		
SLAM-R	0 (-2, 2)	
SLEDAI	0 (-3, 2)	

[§] IQR = inter-quartile range, i.e. the range between the 25th and 75th percentiles of the variable values.

[a]These changes in SLAM-R and SLEDAI scores are pooled over the entire set of observations used in the analyses reported in this paper.

Table 3. Frequencies of changes recorded in SLAM-R organ-specific subscores (N = 591).

Organ system	% visits with score decrease (median score change)	% of visits with no score change	% of visits with score increase (median score change)
Constitutional	24.0 (-1)	53.0	23.0 (1)
Cardiovascular	13.5 (-1)	71.6	14.9 (1)
ESR	11.2 (-1)	77.0	11.8 (1)
Eye	0.5 (-1)	99.2	0.3 (1)
Gastrointestinal tract	6.6 (-1)	86.6	6.8 (1)
Hematological	19.3 (-1)	55.7	25.0 (1)
Integument	24.5 (-1)	52.1	23.4 (1)
Musculoskeletal	22.0 (-1)	56.9	21.2 (1)
Neurological	21.7 (-1)	60.7	17.6 (1)
Ad hoc	6.1 (-1)	88.8	5.1 (1)
Pulmonary	13.0 (-1)	76.0	11.0 (1)
Renal	17.6 (-1)	65.1	17.3 (1)
Reticuloendothelial	6.3 (-1)	87.7	6.1 (1)

Table 4. Frequencies of changes recorded in SLEDAI organ-specific subscores (N = 591).

Organ system	% of visits showing score decrease (median score change)	% of visits showing no score change	% of visits showing score increase (median score change)
Arthritis	12.0 (-4)	76.8	11.2 (4)
CNS	5.8 (-8)	88.8	5.4 (8)
Hematological	2.5 (-1)	93.7	3.7 (1)
Immunological	12.7 (-2)	75.3	12.0 (2)
Integument	19.0 (-2)	63.5	17.6 (2)
Renal	18.4 (-4)	65.0	16.6 (4)
Serositis	2.0 (-2)	96.5	1.5 (2)
Fever	2.5 (-1)	95.4	2.0 (1)
Vasculitis	1.7 (-8)	96.1	2.2 (8)

Table 5. ORs for association between reported improvement in overall disease activity and 1-point decreases in SLAM-R subscores. [a]

	Patients		Physicians	
Organ system	**OR**	**95% CI**	**OR**	**95% CI**
Constitutional	1.30*	1.00, 1.67	1.56**	1.20, 2.01
Cardiovascular	1.20*	0.96, 1.59	0.84	0.64, 1.10
ESR	1.20	0.92, 1.61	1.26	0.89, 1.81
Eye	0.40	0.05, 3.13	0.43	0.07, 2.47
Gastrointestinal	1.40	0.90, 2.06	1.39	0.84, 2.30
Hematological	1.20	0.94, 1.39	1.18	0.97, 1.43
Integument	1.00	0.87, 1.21	1.18	0.91, 1.52
Musculoskeletal	1.30*	0.99, 1.59	1.47**	1.11, 1.95
Neuromotor	1.00	0.76, 1.30	1.23	0.96, 1.58
Ad hoc	1.10	0.60, 1.89	1.54	0.88, 2.72
Pulmonary	1.10	0.75, 1.74	1.42	0.93, 2.18
Renal	1.10	0.91, 1.43	1.30**	1.01, 1.66
Reticuloendothelial	0.90	0.61, 1.24	1.44*	0.99, 2.10

* denotes associations that are statistically significant at $\alpha = 0.10$

** denotes associations that are statistically significant at $\alpha = 0.05$

[a] Results of the GEE approach to logistic regression for repeated measurements. The OR for a 1-point score decrease in each organ system is adjusted for the effects of all other organs and for the total SLAM-R score from the previous visit.

Table 6. ORs for association between reported improvement in overall disease activity and decreases in SLEDAI organ system subscores. [b]

Organ system	Smallest possible decrease in subscore	Patient		Physician	
		OR	95% CI	OR	95% CI
Arthritis	4	1.60**	1.06, 2.32	1.70**	1.10, 2.56
CNS	4	1.30**	1.04, 1.67	1.50**	1.14, 1.84
Hematological	1	1.90**	1.27, 2.83	2.10	0.75, 5.86
Immunological	2	1.00	0.70, 1.29	1.00	0.71, 1.39
Integument	2	1.00	0.86, 1.23	1.30**	1.01, 1.79
Renal	4	1.00	0.78, 1.35	1.30**	1.03, 1.71
Serositis	2	0.80	0.43, 1.47	1.20	0.39, 3.85
Fever	1	1.70	0.82, 3.34	1.80	0.67, 4.90
Vasculitis	8	1.10	0.18, 6.81	1.10	0.23, 5.27

* denotes associations that are statistically significant at $\alpha = 0.10$

** denotes associations that are statistically significant at $\alpha = 0.05$

[b] Results of the GEE approach to logistic regression for repeated measurements. The OR for the smallest possible score decrease in each organ system is adjusted for the effects of all other organs and for the total SLEDAI score from the previous visit.

Table 7. ORs for association between reported deterioration in overall disease activity and 1-point increases in SLAM-R organ system subscores. [c]

	Patient		Physician	
Organ system	**OR**	**95% CI**	**OR**	**95% CI**
Constitutional	1.20*	0.98, 1.51	1.50**	1.12, 1.89
Cardiovascular	1.00	0.75, 1.21	1.40	0.92, 2.04
ESR	0.80	0.56, 1.10	1.20	0.87, 1.74
Eye	1.00	0.13, 8.28	1.30	0.25, 6.65
Gastrointestinal	2.10**	1.32, 3.24	1.20	0.64, 2.16
Hematological	0.80**	0.70, 0.98	0.90	0.74, 1.18
Integument	1.20	0.94, 1.56	1.70**	1.31, 2.23
Musculoskeletal	1.30*	0.97, 1.61	1.80**	1.26, 2.55
Neuromotor	1.00	0.78, 1.28	0.90	0.74, 1.21
Ad hoc	1.00	0.64, 1.56	1.30	0.86, 1.98
Pulmonary	1.40	0.84, 2.30	1.50*	0.95, 2.19
Renal	1.10	0.78, 1.62	1.30*	0.96, 1.67
Reticuloendothelial	1.40*	0.96, 1.99	1.50**	1.03, 2.21

* denotes associations that are statistically significant at $\alpha = 0.10$

** denotes associations that are statistically significant at $\alpha = 0.05$

[c] Results of the GEE approach to logistic regression for repeated measurements. The OR for a 1-point score increase in each organ system is adjusted for the effects of all other organs and for the total SLAM-R score from the previous visit.

Table 8. ORs for association between reported deterioration and increases in SLEDAI organ system subscores. [d]

Organ system	Smallest possible increase in subscore	Patient		Physician	
		OR	95% CI	OR	95% CI
Arthritis	4	1.00	0.67, 1.63	1.90**	1.10, 3.21
CNS	4	1.00	0.77, 1.40	1.50**	1.02, 2.35
Hematological	1	0.80	0.50, 1.30	1.60	0.78, 3.07
Immunological	2	1.10	0.73, 1.56	1.00	0.69, 1.48
Integument	2	1.30*	0.98, 1.69	1.60**	1.20, 2.26
Renal	4	1.20	0.83, 1.84	1.00	0.71, 1.28
Serositis	2	2.30*	0.90, 5.94	2.70	0.78, 9.13
Fever	1	1.20	0.44, 3.33	5.40**	1.97, 14.85
Vasculitis	8	0.90	0.40, 1.90	2.00	0.85, 4.49

* denotes associations that are statistically significant at $\alpha = 0.10$

** denotes associations that are statistically significant at $\alpha = 0.05$

[d] Results of the GEE approach to logistic regression for repeated measurements. The OR for the smallest possible score increase in each organ system is adjusted for the effects of all other organs and for the total SLEDAI score from the previous visit.

CHAPTER 5: DISCUSSION

The development of a reliable, valid, and responsive standardized measure of SLE disease activity is necessary for comparisons of patients seen by different physicians or participating in different studies (Fortin et al., 1995). However, because of the heterogeneity of SLE manifestations, experts disagree both on what constitutes disease activity, and on how to express level of disease activity in a manner that is both global and numerical (Liang et al., 1988a). Two widely-used indices that operationalize SLE activity in different ways are SLAM-R (Liang et al., 1989) and SLEDAI (Bombardier et al., 1992). Because previous studies on the responsiveness of these two instruments to relevant change in SLE activity have produced conflicting results, more investigation of this psychometric property is required.

One of the difficulties in measuring instrument responsiveness to change is that the operational definition of responsiveness itself is unclear, and thus many different methods of assessing responsiveness have been proposed. Several involve a "signal-to-noise" ratio (Kazis et al., 1989; Liang et al., 1990; Guyatt et al., 1987), two are graphical (Stucki et al., 1999; Deyo et al., 1986), and one uses a non-parametric polytomous regression technique to evaluate, in one model, the ability of the instruments to predict improvement and deterioration (Fortin et al., 2000). Strengths and weaknesses of each method vary. Effect size, for example, measures the ratio of mean change to baseline score, and thus may underestimate responsiveness if estimated in a population displaying a wide spectrum of true baseline disease activity that would be reflected in a large variation of scores on a truly discriminative instrument. To ensure that my results did not simply reflect my choice of responsiveness statistic, I used several different methods found in the literature plus a novel technique, focusing on the predictive ability of changes in instrument scores.

The absence of a gold standard for SLE disease activity, a result of the protean nature of the disease, presents another challenge in measuring responsiveness, and lupus investigators have dealt with this issue in a variety of ways. Brunner et al. (1999) measured the change in instrument scores between two points in time and assumed *a priori* that true change in disease activity would be homogeneous over this time period. However, they did not verify the assumption. Fortin et al. (1995a) and Ward et al. (2000) used physician and

patient global assessments as the external criteria. In another study (Fortin et al., 2000), instrument score changes were compared to physician responses to a transition question, which asks explicitly whether disease activity was felt to have improved or worsened; I used the same approach in this analysis.

The literature review of measurement of the responsiveness of SLAM-R and SLEDAI to change revealed some common methodological limitations. First, sample sizes in many of the studies were small, possibly as a result of the rareness of SLE. Second, although some studies used multiple measures from each patient, they did not pool estimates of responsiveness across all visits (Brunner et al., 1999; Fortin et al., 2000). Third, few studies investigated the relevance of instrument score changes of given magnitudes (Fortin et al., 2000). Fourth, precision of the estimates was quantified in only one study, and this was done based on a strong *a priori* assumption of normality of the underlying sampling distribution (Brunner et al., 1999). Moreover, none of the studies quantified the precision of the reported differences in the responsiveness of SLAM-R compared to SLEDAI, leaving it uncertain to what extent these differences may be due to sampling error. Finally, the responsiveness of SLAM-R and SLEDAI to patient assessment of relevant change was rarely investigated (Ward et al., 2000).

In the study described in Chapter Three, I attempted to address these methodological issues. The data I used were acquired as a part of a randomized clinical trial in which SLE patients were evaluated monthly by physicians, for up to eighteen months. By using weighted averages, I was able to obtain single estimates for the effect size, standardized response mean, and control-standardized response mean of SLAM-R and SLEDAI for visits in which relevant change was reported. Furthermore, instead of assuming *a priori* that these responsiveness statistics were parametrically distributed, I employed bootstrapping techniques to quantify the precision of my estimates. I also used the GEE approach to logistic regression to measure the strength of the association between SLAM-R or SLEDAI score changes, and perceived improvement or deterioration in the patients, and to estimate the probability of relevant change being reported at a given change in score. The GEE approach allowed me to enhance the precision of the regression parameters through the use of the repeated measures, while accounting for intra-patient dependence of variable values. The results indicated that SLAM-R was better able than SLEDAI to detect changes in overall

disease activity important to both patients and physicians, and that, in fact, SLEDAI was unable to detect deterioration reported by patients. My findings supported the conclusions of two recent studies (Fortin et al., 2000; Ward et al., 2000), but were not entirely consistent with findings from some other investigations (Gladman et al., 1994; Brunner et al., 1999). This disagreement may have been caused by the use of different study designs and techniques for assessing responsiveness. For example, Brunner et al. (1999) used as their "gold standard" changes in SLE activity occurring over pairs of pre-specified points in time. Lupus disease activity was assumed to change in the same direction over the same time intervals for all patients, but this assumption was not verified.

It has been suggested that the greater responsiveness of SLAM-R to changes important to patients, compared to that of SLEDAI, might result from the inclusion of more patient-reported manifestations in SLAM-R than in SLEDAI (Ward et al., 2000). Also, certain types of organ involvement have been shown to be associated with discordance in patient and physician assessments of disease activity at a single point in time (Neville et al., 2000). To determine whether the importance of change in organ involvement depended on the instrument used and/or the assessor, I investigated the association between changes in the activity of particular organs, defined as score changes in the corresponding subset of SLAM-R and SLEDAI items, and the relevance to physicians and patients of perceived change in overall SLE activity. I found that the pattern of organ-specific changes that were systematically associated with overall assessments varied for physicians and patients, for SLAM-R and SLEDAI, and for direction of relevant change. In particular, changes in most items in SLEDAI did not reflect patient-reported deterioration. This suggests that item definitions used in SLEDAI may not always correspond to aspects of disease activity that patients deem relevant. Examples of organ systems in which changes seemed relevant in at least a subset of the outcomes included the musculoskeletal system, constitutional systems, the central nervous system, and the integument.

Although SLAM-R was shown to be more responsive than SLEDAI to both physician and patient assessments of relevant overall change in SLE activity, this does not imply that one instrument should be used exclusively, as the organ systems covered by each do not overlap completely, and differences were found between the two instruments with respect to the organ systems that were found relevant. For example, changes in CNS activity recorded

in SLEDAI showed a statistically significant association with improvement relevant to both physicians and patients, and with deterioration relevant to physicians. In contrast, changes in neuromotor activity recorded in SLAM-R showed very little association with reported changes in overall activity. It is therefore possible that SLEDAI would be especially useful for evaluating patients with a great deal of CNS activity.

A few limitations of this study should be mentioned. First, some items were not completed at all visits, resulting in a loss of observations in the organ-specific analysis. Attempts were made initially to maximize the study sample used, but because this would have resulted in only partial overlap of the observations and possible incomparability of results from different analyses, it was decided that the sample would be restricted to those visits with complete data. This may have resulted in selection bias, especially if the patterns of changes in disease activity were different for patients eliminated from the study population, i.e. those with less complete data. As a point of interest though, the regression parameter estimates did not differ materially for the two approaches (data not shown).

Second, the use of pre-collected data meant that characteristics of the study population were defined by the inclusion and exclusion criteria of the original trial (Appendix D). Generally, this meant that patients with extensive organ damage and patients with a history of poor treatment compliance were not included in the study. It is possible that assessments of changes in disease activity by these patients and their physicians differ systematically. Therefore, it may be useful to investigate these two populations in greater depth.

Third, it was possible that some bias was introduced by the use of GEE modelling. Greenland (1998) mentions that this may occur if outcome variables affect endogenous variables. An example in the context of this study would be if awareness of changes in activity prompted physicians and patients to be more vigilant than usual and thus report manifestations that wouldn't have otherwise been reported. However, it would be difficult to detect and adjust for this problem.

Variations in the relevance to patients and physicians of organ-specific changes in activity may explain partly why results of previous studies on the responsiveness of SLAM-R and SLEDAI have been inconsistent. It might be helpful in future responsiveness studies to explore the patterns of organ-specific changes observed, to determine if there are any

systematic differences between study populations. In addition, the effects of organ damage on the relevance of changes in activity recorded in SLAM-R and SLEDAI, and the relationship between treatment compliance and changes relevant to patients and physicians should be studied further.

In conclusion, SLAM-R was shown to be more responsive than SLEDAI to changes in disease activity relevant to patients and physicians, and both instruments were more responsive to physician-reported changes than to those reported by patients. Differences in the relevance to patients and physicians of organ-specific changes in activity, and differences in item definitions of SLAM-R and SLEDAI, may account for the disparities. The contrast between types of organ involvement relevant to patients and physicians may reflect a need for greater communication between physicians and patients regarding the implications of specific organ activity.

REFERENCES

Abrahamowicz, M., Fortin, P. R., du Berger, R., Nayak, V., Neville, C., & Liang, M. H. (1998). The relationship between disease activity and expert physician's decision to start major treatment in active systemic lupus erythematosus: a decision aid for development of entry criteria for clinical trials. J.Rheumatol., 25, 277-284.

Abrahamowicz, M. & Ramsay, J. (1992). Multicategorical spline model for item response theory. Psychometrika, 57, 5-27.

Bombardier, C., Gladman, D. D., Urowitz, M. B., Caron, D., & Chang, C. H. (1992). Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum., 35, 630-640.

Brunner, H. I., Feldman, B. M., Bombardier, C., & Silverman, E. D. (1999). Sensitivity of the Systemic Lupus Erythematosus Disease Activity Index, British Isles Lupus Assessment Group Index, and Systemic Lupus Activity Measure in the evaluation of clinical change in childhood-onset systemic lupus erythematosus. Arthritis Rheum., 42, 1354-1360.

Chang, D. M., Chang, C. C., Kuo, S. Y., Chu, S. J., & Chang, M. L. (1998). The clinical features and prognosis of male lupus in Taiwan. Lupus, 7, 462-468.

Chang, E., Abrahamowicz, M., Ferland, D., & Fortin, P. Comparison of the responsiveness of two lupus disease activity measures to changes relevant to patients and physicians. 2000. Ref Type: Unpublished Work

Cohen, J. (1988). The t test for means. In L.Erlbaum (Ed.), Statistical power analysis for the behavioral sciences (pp. 19-74). Hillsdale: Lawrence Erlbaum Associates.

Deyo, R. A. & Centor, R. M. (1986). Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J.Chronic.Dis., 39, 897-906.

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin.Trials, 12, 142S-158S.

Efron, B. & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician, 37, 36-48.

Fortin, P. R., Abrahamowicz, M., Clarke, A. E., Neville, C., du Berger, R., Fraenkel, L., & Liang, M. I. H. (2000). Do lupus disease activity measures detect clinically important change? Journal of Rheumatology, 27, 1421-1428.

Fortin, P. R., Abrahamowicz, M., & Danoff, D. (1995). Small changes in outpatients lupus activity are better detected by clinical instruments than by laboratory tests. J.Rheumatol., 22, 2078-2083.

Fortin, P. R., Stucki, G., & Katz, J. N. (1995). Measuring relevant change: an emerging challenge in rheumatologic clinical trials [editorial]. Arthritis Rheum., 38, 1027-1030.

Gladman, D., Ginzler, E., Goldsmith, C., Fortin, P., Liang, M., Urowitz, M., Bacon, P., Bombardieri, S., Hanly, J., Hay, E., Isenberg, D., Jones, J., Kalunian, K., Maddison, P., Nived, O., Petri, M., Richter, M., Sanchez-Guerrero, J., Snaith, M., Sturfelt, G., Symmons, D., & Zoma, A. (1996). The development and initial validation of the Systemic Lupus

International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus [see comments]. Arthritis Rheum., 39, 363-369.

Gladman, D. D., Goldsmith, C. H., Urowitz, M. B., Bacon, P., Bombardier, C., Isenberg, D., Kalunian, K., Liang, M. H., Maddison, P., & Nived, O. (1994). Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation. J.Rheumatol., 21, 1468-1471.

Gladman, D. D., Urowitz, M. B., Kagal, A., & Hallett, D. (2000). Accurately describing changes in disease activity in Systemic Lupus Erythematosus. Journal of Rheumatology, 27, 377-379.

Greenland, S. (1998). Introduction to regression modelling. In K.J.Rothman & S. Greenland (Eds.), Modern Epidemiology (2 ed., pp. 401-434). Philadelphia: Lippincott Williams & Wilkins.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. J.Chronic.Dis., 40, 171-178.

Guyatt, G. H., Deyo, R. A., Charlson, M., Levine, M. N., & Mitchell, A. (1989). Responsiveness and validity in health status measurement: a clarification. J.Clin.Epidemiol., 42, 403-408.

Hochberg, M. (1993). The epidemiology of systemic lupus erythematosus. In D.J.Wallace & B. H. Hann (Eds.), Dubois' Lupus Erythematosus (pp. 49-57). Pennsylvania: Lea & Febiger.

Jacobsen, S., Petersen, J., Ullman, S., Junker, P., Voss, A., Rasmussen, J. M., Tarp, U., Poulsen, L. H., van Overeem, H. G., Skaarup, B., Hansen, T. M., Podenphant, J., & Halberg, P. (1998). A multicentre study of 513 Danish patients with systemic lupus erythematosus. II. Disease mortality and clinical factors of prognostic value. Clinical Rheumatology, 17, 478-484.

Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin.Trials, 10, 407-415.

Jennrich, R. I. & Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. Biometrics, 42, 805-820.

Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. Med.Care, 27, S178-S189.

Kiss, E., Regeczy, N., & Szegedi, G. (1999). Systemic lupus erythematosus survival in Hungary. Results from a single centre. Clinical & Experimental Rheumatology, 17, 171-177.

Klippel, J. H. (1997). Systemic lupus erythematosus: demographics, prognosis, and outcome. [Review] [44 refs]. Journal of Rheumatology - Supplement, 48, 67-71.

Kwoh, C. K., O'Connor, G. T., Regan-Smith, M. G., Olmstead, E. M., Brown, L. A., Burnett, J. B., Hochman, R. F., King, K., & Morgan, G. J. (1992). Concordance between clinician and patient assessment of physical and mental health status. Journal of Rheumatology, 19, 1031-1037.

Liang, K. Y. & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73, 13-22.

Liang, M. H. (1995). Evaluating measurement responsiveness. J.Rheumatol., 22, 1191-1192.

Liang, M. H., Fossel, A. H., & Larson, M. G. (1990). Comparisons of five health status instruments for orthopedic evaluation. Med.Care, 28, 632-642.

Liang, M. H., Larson, M. G., Cullen, K. E., & Schwartz, J. A. (1985). Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum., 28, 542-547.

Liang, M. H., Socher, S. A., Larson, M. G., & Schur, P. H. (1989). Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum., 32, 1107-1118.

Liang, M. H., Socher, S. A., Roberts, W. N., & Esdaile, J. M. (1988a). Measurement of systemic lupus erythematosus activity in clinical research. Arthritis Rheum., 31, 817-825.

Liang, M. H., Stern, S., & Esdaile, J. M. (1988b). Systemic lupus erythematosus activity. An operational definition. [Review] [35 refs]. Rheumatic Diseases Clinics of North America, 14, 57-66.

Matthews, D. E. & Farewell, V. T. (1988). Linear regression models for medical data. In Using and understanding medical statistics (2nd ed., pp. 124-140). Basel, Switzerland: Karger.

Medlog (1995). Medlog Systems (Version 95.6A) [Computer software]. Incline Village, NV.

Meenan, R. F., Anderson, J. J., Kazis, L. E., Egger, M. J., Altz-Smith, M., Samuelson, C. O., Jr., Willkens, R. F., Solsky, M. A., Hayes, S. P., & Blocka, K. L. (1984). Outcome assessment in clinical trials. Evidence for the sensitivity of a health status measure. Arthritis Rheum., 27, 1344-1352.

Mills, J. A. (1994). Systemic lupus erythematosus [see comments]. N.Engl.J.Med., 330, 1871-1879.

Naylor, C. D. & Llewellyn-Thomas, H. A. (1994). Can there be a more patient-centred approach to determining clinically important effect sizes for randomized treatment trials? Journal of Clinical Epidemiology, 47, 787-795.

Neville, C., Clarke, A. E., Joseph, L., Belisle, P., Ferland, D., & Fortin, P. R. (2000). Learning from discordance in patient and physician global assessments of systemic lupus erythematosus disease activity. Journal of Rheumatology, 27, 675-679.

Panush, R. S., Levine, M. L., & Reichlin, M. (2000). Do I need an ANA? Some thoughts about man's best friend and the transmissibility of lupus [editorial]. Journal of Rheumatology, 27, 287-291.

Petri, M., Genovese, M., Engle, E., & Hochberg, M. (1991). Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. Arthritis Rheum., 34, 937-944.

Petri, M., Hellmann, D., & Hochberg, M. (1992). Validity and reliability of lupus activity measures in the routine clinic setting. J.Rheumatol., 19, 53-59.

Redelmeier, D. A. & Lorig, K. (1993). Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Archives of Internal Medicine, 153, 1337-1342.

SAS (1996). SAS Institute Inc. (Version 6.12) [Computer software]. SAS Institute Inc.

Schroeder, J. O. & Euler, H. H. (1997). Recognition and management of systemic lupus erythematosus. Drugs, 54, 422-434.

Senécal, J. L. (1991). Lupus: the disease with 1000 faces. (2 ed.) Calgary: Lupus Canada.

Stoll, T., Stucki, G., Malik, J., Pyke, S., & Isenberg, D. A. (1996). Further validation of the BILAG disease activity index in patients with systemic lupus erythematosus. Annals of the Rheumatic Diseases, 55, 756-760.

Stucki, G., Liang, M. H., Stucki, S., Katz, J. N., & Lew, R. A. (1999). Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. Clin.Rheumatol., 18, 101-105.

Symmons, D. P., Coppock, J. S., Bacon, P. A., Bresnihan, B., Isenberg, D. A., Maddison, P., McHugh, N., Snaith, M. L., & Zoma, A. S. (1988). Development and assessment of a computerized index of clinical disease activity in systemic lupus

erythematosus. Members of the British Isles Lupus Assessment Group (BILAG). Q.J.Med., 69, 927-937.

Thumboo, J., Fong, K. Y., Chng, H. H., Koh, E. T., Chia, H. P., Leong, K. H., Koh, W. H., Howe, H. S., Leong, K. P., Wong, M. H., Chew, S. M., Chai, P., Goh, L. H., Goon, T. J., Lau, T. C., Lim, W. S., Pek, W. Y., Tong, K. L., Yang, W. L., Feng, P. H., & Boey, M. L. (1998). The effects of ethnicity on disease patterns in 472 Orientals with systemic lupus erythematosus. J.Rheumatol., 25, 1299-1304.

Uramoto, K. M., Michet, C. J., Jr., Thumboo, J., Sunku, J., O'Fallon, W. M., & Gabriel, S. E. (1999). Trends in the incidence and mortality of systemic lupus erythematosus, 1950-1992 [see comments]. Arthritis Rheum., 42, 46-50.

Vliet Vlieland, T. P., Zwinderman, A. H., Breedveld, F. C., & Hazes, J. M. (1997). Measurement of morning stiffness in rheumatoid arthritis clinical trials. J.Clin.Epidemiol., 50, 757-763.

Ward, M. M., Marx, A. S., & Barry, N. N. (2000). Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus. Journal of Rheumatology, 27, 664-670.

Wells, G. A., Tugwell, P., Kraag, G. R., Baker, P. R., Groh, J., & Redelmeier, D. A. (1993). Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. Journal of Rheumatology, 20, 557-560.

Wright, J. G. (1996). The minimal important difference: who's to say what is important? [comment]. Journal of Clinical Epidemiology, 49, 1221-1222.

Xie, S. K., Feng, S. F., & Fu, H. (1998). Long term follow-up of patients with systemic lupus erythematosus. Journal of Dermatology, 25, 367-373.

Appendix A

The SLAM-R and SLEDAI forms used in SMILE. SLAM-R is 3 pages long, and SLEDAI is 1 page long.

© S.M.I.L.E. PROJECT

SLAM-R Form
Date:___/____/___

Study Number:__ __ - __ __ __
Visit:____ Week:____

		ABSENT NORMAL	MILD	MODERATE	SEVERE	UNKNOWN
Score.....		(0)	(1)	(2)	(3)	(-)
CONSTITUTIONAL						
1.	Weight loss	☐	☐ < 10% body weight		☐ > 10%	☐
2.	Fatigue	☐	☐ Little or no limit on activity	☐ Limits ADL		☐
3.	Fever	☐	☐ 37.5-38.5°C		☐ > 38.5°C	☐
INTEGUMENT						
4.	Oral/nasal, or periungal erythema, or malar rash, or photosensitive rash or nail fold infarct	☐	☐ present			☐
5.	Alopecia	☐	☐ Hair loss with trauma	☐ Alopecia observed		☐
6.	Erythematous, maculopapular rash, or discord lupus, or lupus profundus, or bilous lesions	☐	☐ < 20% Total Body surface (TBA)	☐ 20-50% TBA	☐ > 50% TBA	☐
7.	Vasculitis, (leucocytoclastic vasculitis, urticaria, palpable purpura, livedo reticularis, ulcer or panniculitis)	☐	☐ < 20% TBA	☐ 20-50% TBA	☐ > 50% TBA or necrosis	☐
EYE						
8.	Cytoid bodies	☐	☐ Present		☐ Visual acuity < 20/200	☐
9.	Hemorrhages (retinal or choroidal) or episcleritis	☐	☐ Present		☐ Visual acuity < 20/200	☐
10.	Papillitis or pseudotumor cerebri	☐	☐ Present		☐ Visual acuity < 20/200 or field cut	☐
RETICULOENDOTHELIAL						
11.	Lymphadenopathy	☐	☐ Shotty	☐ Diffuse or nodes > 1 cm x 1.5 cm		☐
12.	Hepato- or splenomegaly	☐	☐ Palpable only with inspiration	☐ Palpable without inspiration		☐

SLAM-R Form (con't)

Study Number:___ ___ - ___ ___ ___
Visit:____ Week:____

	ABSENT NORMAL (0)	MILD (1)	MODERATE (2)	SEVERE (3)	UNKNOWN (-)
Score.....					
PULMONARY					
13.	☐	☐ Shortness of breath Or pain. Exam normal.	☐ Shortness of breath or pain with exercise or abnormal lung exam.	☐ Shortness of breath or pain at rest or abnormal lung exam.	☐
CARDIOVASCULAR					
14. Raynaud's	☐	☐ Present			☐
15. Hypertension	☐	☐ Diast. 90-105	☐ Diast. 105-115	☐ Diast. > 115	☐
16. Carditis	☐		☐ Chest pain or arrythmia	☐ Myocarditis with hemodynamic compromise &/or arrythmia	☐
GASTROINTESTINAL					
17. Abdominal pain (Serositis, pancreatitis, ischemic bowel, etc)	☐	☐ Complaint	☐ Limiting pain	☐ Peritoneal signs/ ascites	☐
NEUROMOTOR					
18. Stroke (includes mononeuritis multiplex, reversible neurologic deficit (RND), cerebrovascular accident (CVA), retinal vascular thrombosis)	☐		☐ RND, or mononeuritis multiplex or cranial Neuropathy or chorea	☐ CVA/myelitis, retinal vascular occlusion	☐
19. Seizure	☐		☐ 1 or more/month	☐ Status epilepticus	☐
20. Cortical dysfunction	☐	☐ Mild depression/ personality disorder	☐ ▲ in sensorium or severe depression or limiting cognitive impairment	☐ or coma	☐ Psychosis or dementia
21. Headache (including migraine equivalents and aseptic meningitis)	☐	☐ Symptoms or transient neuro deficit	☐ Interferes with normal activities/ Aseptic meningitis		☐
22. Myalgia/myositis	☐	☐ Complaint	☐ Limits some activity	☐ Incapacitating	☐

SLAM-R Form (con't)

Study Number:__ __ - __ __ __
Visit:____ Week:____

		ABSENT NORMAL	MILD	MODERATE	SEVERE	NOT RECORDED
Score.....		(0)	(1)	(2)	(3)	(-)
JOINTS						
23.	Joint pain	☐	☐ Arthralgia only	☐ Objective inflammation	☐ Limited function	☐
OTHER						
24.	(Write rules for ascertainment and ad hoc scale)	☐	☐	☐	☐	☐
LABORATORY		NORMAL	MILD	MODERATE	SEVERE	UNKNOWN NOT RECORDED
25.	Hematocrit	☐ > 0.35	☐ 0.30-0.35	☐ 0.250-0.299	☐ < 0.25	☐
26.	WBC	☐ > 3.5	☐ 3.5-2.0	☐ 2.0-1.0	☐ <1.0	☐
27.	Lymphocyte count	☐ 1.5-4.0	☐ 1.49-1.0	☐ 0.99-0.50	☐ <0.49	☐
28.	Platelet count	☐ > 150 T	☐ 100-150 T	☐ 99-50 T	☐ < 50 T	☐
29.	ESR (Westergren)	☐ < 25	☐ 25-50	☐ 51-75	☐ > 75	☐
30.	Serum creatinine or creatinine clearance	☐ 50 -130 mEq/L or 80-100 % CrCl	☐ 140-200 Eq/L or 79-60 % CrCl	☐ 210-400 mEq/L or 30-60 % CrCl	☐ > 400 mEq/L or < 30 % CrCl	☐
31.	Urine sediment	☐	☐ > 5 RBC &/or WBC/hpf &/or 0 to 1-3 granular &/or non RBC cellular casts/hpf &/or tr-1+ proteinuria &/or < 500 mg/L 24° urine protein.	☐ > 10 RBC &/or WBC/hpf or or > 3 granular &/or non RBC cellular casts/hpf &/or 2+-3+ proteinurea &/or ≥ 500 mg/L - 3.5 g/L 24° urine protein.	☐ > 25 RBC &/or WBC/hpf &/or Red cell cast &/or 4+ proteinuria &/or > 3.5 g/L 24° urine protein.	☐

MD GLOBAL RATING OF SLE ACTIVITY *

0 None — 10 Most

PATIENT GLOBAL RATING OF THEIR DISEASE ACTIVITY

0 None — 10 Most

* Over last month. Use most active rating during the month.
Assume that **"NONE"** and **"MOST ACTIVE"** refer to what patient has experienced.

Total SLAM Score

☺ S.M.I.L.E. PROJECT

SLEDAI Form

Study Number: __ __ - __ __ __

Today's Date: __ / ____ / __

Visit:____ Week:____

(Enter weight in SLEDAI score column if descriptor present at the time of the visit or in the preceding 10 days)

Weight	SLEDAI score	Descriptor	Definition
8	_____	Seizure	Recent onset. Exclude metabolic, infectious or drug causes.
8	_____	Psychosis	Altered ability to function in normal activity due to severe disturbance in the perception of reality. Include hallucinations, incoherence, marked loose associations, improvised thought content, marked illogical thinking, bizarre, disorganized or catatonic behaviour. Exclude presence of uraemia and offending drugs.
8	_____	Organic Brain Syndrome	Altered mental function with impaired orientation, memory or other intellectual function with rapid onset, fluctuating clinical features. Such as any of the following: a) clouding of consciousness with reduced capacity to focus and inability to sustain attention to environment. Plus at least 2 of b) of perceptual disturbance; incoherent speech; insomnia or daytime drowsiness; increased or decreased psychomotor activity. (Exclude metabolic, infectious, drug causes).
8	_____	Visual	Retinal changes of SLE; any of cytoid bodies, retinal haemorrhages, serous exudate or haemorrhages in the choroid, optic neuritis. (Not due to hypertension or drugs or infection).
8	_____	Cranial Nerve	New onset of sensory or motor neuropathy involving cranial nerves.
8	_____	Lupus Headaches	Severe, persistent headache, may be migraines, but must be non-responsive to narcotic analgesia.
8	_____	CVA	New syndrome. Exclude arteriosclerosis.
8	_____	Vasculitis	Ulcerations, gangrene, tender finger nodules, periungual infarction, splinter haemorrhages, biopsy or angiogram proof of vasculitis.
4	_____	Arthritis	More than 2 joints with pain and signs of inflammation (ie. Tenderness, swelling, or effusion).
4	_____	Myositis	Proximal muscle aching/weakness, associated with elevated CPK/aldolase or EMG changes or a biopsy showing myositis.
4	_____	Casts	Heme granular or RBC.
4	_____	Haematuria	> 5 RBC/HPF. Excluding other causes (stone, infection).
4	_____	Proteinuria	> 0.5 g/24 hours. New onset or recent increase of more than 0.5 g/24 hrs.
4	_____	Pyuria	> 5 WBC/HPF. Exclude infection.
2	_____	New Rash	New onset or recurrence of inflammatory type rash.
2	_____	Alopecia	New or recurrent. An abnormal patch of diffuse loss of hair.
2	_____	Mucous membrane	New onset or recurrence of oral or nasal ulcerations.
2	_____	Pleurisy	Pleuritic chest pain with pleural rub or effusion, or pleural thickening.
2	_____	Pericarditis	Pericardial pain with at least one of the following: rub, effusion, ECG, echo confirmation.
2	_____	Low Complement	Decreased any of CH5O, C3, C4. Below the lower limit of normal for lab.
2	_____	Increased DNA binding	> 25% binding by Farr assay. Above normal range of lab value (eg.25%).
1	_____	Fever	> 38℃. After exclusion of infection.
1	_____	Thrombocytopenia	< 100,000 platelets.
1	_____	Leucopenia	WBC < 3000 (not due to drugs).

Appendix B

Documentation showing approval of SMILE by the research ethics boards of each study centre. No new data were gathered for the analyses reported in this thesis. There are 25 pages in total.

Appendix C

English and French SMILE consent forms from Montreal General Hospital. Each consent form is 8 pages long.

MONTREAL GENERAL HOSPITAL

CONSENT FORM

A MULTICENTER, RANDOMIZED, DOUBLE-BLIND PLACEBO-CONTROLLED STUDY OF METHOTREXATE AND FOLIC ACID IN SYSTEMIC LUPUS ERYTHEMATOSUS A PHASE III TRIAL

Sponsor: Canadian Network for Improved Outcome In Systemic Lupus Erythematosus (CANIOS). This study is done with no pharmaceutical support. Faulding (CANADA) Inc. will be providing the methotrexate and placebo at no extra cost and Novopharm Quebec will provide us with the folic acid at no cost.

Principal Investigator: Paul R Fortin, MD, MPH, FRCP(C)

Institution: The Montreal General Hospital
1650 Cedar
Montreal, Qc, H3G 1A4
Tel.: (514) 937-6011 (2437)
FAX: (514) 934-8292

I understand that I am invited to take part in a research study at the Montreal General Hospital which is one of the many centers throughout Canada seeking to identify a more effective means of treating my illness: systemic lupus erythematosus (SLE). Taking part in this study is entirely voluntary. Personal benefit beyond that from ordinary treatment may not result from taking part, but knowledge may be gained that may benefit others. I may withdraw from the study at any time without penalty or compromise in my future medical care. The nature of the study, the risks, inconveniences, discomforts and potential benefits are discussed below. I am urged to discuss any questions I have about this study with one of the investigators.

1) STUDY PURPOSE, PROCEDURES AND DURATION:

PURPOSE:

This research is designed to evaluate the response to methotrexate (MTX) with folic acid of persons suffering from active systemic lupus erythematosus (SLE). MTX resembles the vitamin called folate and it inhibits an enzyme that is thought to be important in the control of immune reactions. Since SLE is a chronic autoimmune and inflammatory disease, MTX may be able to cause and sustain remissions (or control of the lupus) in patients with active lupus. Furthermore, it may have a steroid-sparing effect, allowing for the diminution of prednisone doses required. MTX is taken orally once a week. It is already used and approved in another autoimmune/inflammatory disease: rheumatoid arthritis. In randomized controlled trial, the use of MTX has shown significant effect in controlling rheumatoid arthritis.

PROCEDURE:

I have been approached to participate in this study because my physician has judged that my lupus is more active and/or that I have been requiring high doses of steroids (prednisone) to keep it under control. Because prolonged use of steroids is associated with serious side-effects such as osteoporosis with higher risk for fractures, glucose-intolerance similar to diabetes mellitus, changes in my appearance, possibility of higher blood pressure and higher risk of atherosclerotic disease in the blood vessels of my heart or brain, it is thought that MTX may be a better alternative than standard therapy.

The purpose of this study is to determine whether treatment with MTX and folic acid can result in better control of my illness and enable my doctors to successfully decrease and possibly discontinue steroid treatment sooner than would be possible with standard therapy. Although it is not known whether this will be the case, MTX has enabled people with other disease that cause inflammation to sustain better control and use less prednisone.

To find out if MTX will be helpful, patients such as myself treated with standard therapy [use of non steroidal anti-inflammatory drugs (NSAIDs), steroids and/or antimalarial drugs such as Plaquenil™ (hydroxychloroquine)] will be compared to those treated with MTX in addition to the standard therapy. In this way, I am sure to receive what is considered normal treatment for my lupus.

If I agree to participate, I will be assigned, on the basis of "chance" (like flipping a coin), to either the placebo group or the MTX group. The placebo is a pill that is identical to methotrexate but that has no medication in it (it is like a pill of corn starch or sugar). In that way, I will not be biased in reporting the effects of this treatment on me since I will not know whether I am receiving MTX or placebo.

I will be seeing two doctors, the first will assess how I am progressing only in regard to treatment for SLE. The other doctor will oversee my health care in general. Neither physician will know whether I am taking MTX or placebo, although the second physician will know about all of my other medicines and laboratory tests. That individual will also make necessary adjustments in my medications based on a schedule designed for all patients. A third person, the study coordinator, will dispense the MTX/placebo pill, and will hold the randomization list that indicates which substance is being provided.

Initially, all patients will receive on one day of the week a dose of MTX or placebo and the vitamin folic acid (2.5 mg/day) for the other six days. Once I have clearly improved, prednisone will be reduced according to a written schedule that will be given to me. For some people MTX doses (or placebo) will be changed to suit their tolerance of that medication, based on symptoms and laboratory tests. It will be determined whether MTX is helpful by comparing:

1) the decrease in the mean disease activity in the MTX group compared with the standard therapy group,
2) the time necessary to control the lupus between the two groups,
3) the decrease in prednisone between the two groups,

4) the frequency of lupus flares,
5) the frequency of MTX side-effects.

If someone relapses (increase in their lupus activity or flare-ups) twice in the course of treatment (with or without MTX), he/she would be removed from the study and alternative therapies would be considered based on the judgment of his/her regular physician.

These are certain conditions that would increase a patient's risk of MTX side-effects. If any of these conditions existed for me, I would not be eligible for this study. These conditions include:

1) Kidney failure such that a blood test showed an increased serum creatinine value of $\geq$ 175 µmol/l (SI units) or 2.0 mg/dl.

2) Low white blood cell or platelet counts.

3) The presence of liver abnormalities known to me or detected by a blood test.

4) Alcohol ingestion of more than 2 ounces of 100 proof liquor or beer or its equivalent per week. We can not police alcohol intake, but we strongly discourage its ingestion during this study because alcohol increases the risk of liver toxicity.

5) The combination of marked obesity (more than 1/3 your ideal body weight) plus the presence of diabetes treated with insulin.

6) Use of sulfa-containing drugs. Tylenol will be preferred to nonsteroidal anti-inflammatory drugs whenever possible.

7) A previous history of allergy or intolerance to either methotrexate (MTX) or folic acid (FA).

8) Interstitial lung disease as defined by an abnormal chest x-ray or decrease diffusion capacity (DLCO < 70% of predicted) without evidence of pulmonary hypertension.

DURATION:

All patients will receive study medication (placebo or methotrexate) for twelve months. If I am on steroid, every effort will be made to taper prednisone according to a schedule. There will be an open phase at the end of the trial during which those responding favorably to methotrexate and those on placebo will be offered to either continue or start methotrexate for another six months.

Each visit will include questionnaires to fill in, a brief physical examination and a blood test. Each visit will take between half and hour and one hour of my time.

2) MONITORING RESPONSE AND PROGRESS:

I will visit the Montreal General Hospital for Screening evaluations to see if I am eligible to be in the study. My physician and I will discuss medical history and past medication use. I will have a physical examination and laboratory testing of both blood (no more than eight tablespoons will be drawn) and urine. A chest x-ray will be taken if none has been done in the previous 3 months.

Because of the potential for severe malformations to the embryo/fetus during pregnancy, women must take precautions to avoid becoming pregnant during the course of the study by using a reliable form of birth control. Birth control pills may be associated with an increase risk of having lupus flares and should be avoided, but the decision as to what form of birth control you may wish to use will be left for you to decide with your physician. A pregnancy test will be done in all pre-menopausal women to make sure that they are not pregnant before starting the study.

I will return to the Montreal General Hospital within 10 days of the Screening Visit for the Qualifying Visit, and will finish my eligibility evaluations. My physician will fill out several evaluations of my health, and I will be asked to complete a self-assessment questionnaire on my general health and on the value I attribute to my present state of health. If I am eligible to enter the study and wish to do so, I will then be randomly assigned (or in other word assigned by chance) to one of the two treatment groups as described above. At this time, I will receive my first set of study medication (pills in labeled packs) and instructions on how to take the pills.

The first dosing visit will be two weeks later and then monthly for a total of fourteen visits. I will have laboratory tests of blood (no more than four tablespoons) and urine at each clinic visit. At the actual dosing visit, I will undergo physical examinations, be questioned about my health, and complete the self-assessment questionnaires. Study medication will be dispensed at the 4 week intervals, and I am to return all unused medication and all of the packaging at subsequent visits.

At each of the 4 weekly dosing visits, my health will be assessed for progression or improvement of my lupus. I may be asked to lower the dose of some of my medications. If I experience any severe events, I may discontinue treatment and will undergo the Completion Visit. I will also be asked to discontinue by my physician if I am unable to comply with the study procedures.

If my study medication (methotrexate or placebo) dosage needs to be adjusted, I will be asked to come for an interim visit two weeks after the dosing visit for another blood test and verification of the dose of the study medication I should be on.

At the end of the study or in the event that I must with draw from the study early, I will undergo a study Completion Visit. I will have a physical exam, be questioned about my health, and undergo laboratory tests of my blood (no more than eight tablespoons) and urine. The self assessment questionnaire will be completed, and I will return my last set of study medication packaging.

After the planned treatment period is completed, follow-up will depend on whether I have completed the study or not and whether I am interested in receiving MTX treatment in an open fashion for an additional six months. In that way, those who were on MTX in the blinded study, benefitted from

it and want to continue it can do so along with those who were on placebo and wish to try it under supervision. If I haven't completed the study, I will be asked to come for one follow-up visit 12 months after the Qualifying Visit for a physical exam, be questioned about my health, and undergo laboratory tests of my blood (no more than eight tablespoons) and urine. A self assessment questionnaire will also be completed.

3) INSTRUCTIONS ON HOW TO TAKE THE STUDY DRUG:

I will consistently take the prescribed number of study drug (methotrexate or placebo) pills once a week and will take 2.5 mg (½ tablet) of folic acid daily for the rest of the week (except on the day that I take the MTX). I will consistently take all of my other medications as instructed by my physician. I understand that I am asked to use a reliable form of birth control.

4) RISKS AND BENEFITS:

Side-effects of MTX include a decrease in my blood cell production, irritation of my liver, stomach, lungs, mucus membranes in my mouth or nose, and rashes, nausea, vomiting or diarrhea. Recently, the use of folic acid after the intake of MTX has been shown to decrease its potential side-effects. Elderly patients are not at higher risk of developing toxicity unless they have reduced liver or kidney function. A very rare potential effect of prolonged MTX therapy is that it may increase slightly the likelihood of developing cancer (blood, liver, breast, prostate). MTX has been used to treat arthritic and skin conditions in many thousands of patients since the 1950s. In all of that time, only a handful of patients have developed cancers for which MTX was thought to have possibly played a role.

While there is no assurance that I will receive any benefit from participating in this study, I may have a better control of my lupus activity and may be able to decrease my dose of prednisone. It is possible that this regimen might not benefit me at all or that the side effects produced by taking MTX may outweigh the potential benefits. There may be risks to me or to an embryo that are currently unforeseeable. Future study may be developed to improve treatment of lupus with the results of the present study.

5) STANDARD THERAPY AND NEW RESULTS FROM THIS STUDY:

<u>STANDARD THERAPY:</u>

Due to the design of the present study, standard therapy with antimalarial medications and steroids will be continued. The use of nonsteroidal anti-inflammatory medications will be avoided whenever possible. However, it will be allowed in severe arthritis or serositis cases. Finally, should my lupus become so active as to require immunosuppressive drugs such as azathioprine or cyclophosphamide, the study would be stopped to allow me to receive these drugs.

<u>NEW RESULTS:</u>

NEW RESULTS:

An interim analysis will be done halfway through the study to analyze whether MTX has such a significant effect that everyone should be receiving it immediately rather than at the end of the planned study period. These results would be communicated to me immediately along with any new treatment that would be discovered or developed during the study period.

6) CONFIDENTIALITY:

My research record will remain confidential unless under circumstances required by law. In order to verify study data, monitors from the Health Protection Branch (HPB-CANADA) may need some specific records such as medical charts (including but not limited to those maintained during the course of the study), lab tests and other records related to my health. Research records at the Montreal General Hospital and Montreal General Hospital Research Institute can be reviewed by the Research Ethics and Clinical Trials Committees to monitor compliance with institutional regulations regarding research involving human subjects. No individual identities will be used on any scientific reports or publications resulting from this study.

7) COSTS OF TREATMENT AND COMPENSATION:

There will be no charge for the study drug or for office visits, physicians fees and laboratory fees directly associated with the study activities. I will not be reimbursed for the cost of any other medications (beside the study drug) that I am currently taking and will continue to take during the study. There will be no payment to patients for participating in this study.

8) CONTACTS FOR QUESTIONS:

This study has been explained to me by Dr Paul R. Fortin and my questions were answered. If I have any additional questions about the study, I may call Dr Fortin at (514) 937-6011 ext 2437. If I want to know more about my rights as a participant in a study from the Montreal General Hospital, I can contact the office of Dr. N. Blair Whittemore at (514) 937-6011, ext 3013.

9) HOSPITAL REGULATION:

Any participant into drug trials at the Montreal General Hospital is asked to obtain a Montreal General Hospital medical record unit number (if not already available) and that for my protection, a copy of the study summary with appropriate emergency contact phone numbers along with the consent form, will be forwarded to my Montreal General Hospital chart.

When starting my participation in this study, I will receive an information card providing the name of the study and of the drug with emergency contact phone numbers.

10) STATEMENT OF CONSENT AND VOLUNTARY PARTICIPATION:

I have been given copies of this consent form to keep. By signing this consent form I have indicated that I wish to participate in the study and that I agree to the terms and conditions herein. I also acknowledge that I have reviewed this document and that before giving my consent, I have read and understood the document. I acknowledge that I have carefully weighed and considered the consequences of such participation, and after having done so, I wish to participate in that study as described above.

Furthermore, I understand that my participation in this study may be terminated without my consent should my physician judge it would be more appropriate to do so.

11) VOLUNTARY PARTICIPATION:

PARTICIPATION IN RESEARCH IS VOLUNTARY. I have the right to decline to participate or to withdraw at any point from this study without jeopardy to my medical care.

12) CONSENT:

I have read the explanations of this study and have been given the opportunity to discuss it and ask questions. I am entirely free to participate or not to this study and this will in no way affect my standard medical care. I hereby consent to participate in this multicenter study of MTX utility in the treatment of systemic lupus erythematosus that will be conducted on 100 patients throughout Canada.

_____	_____
Patient Signature	(Please print name)
_____	_____
Investigator Signature	(Please print name)
_____	_____
Witness Signature	(Please print name)
_____	
Date	

APPENDIX: CONSENT TO HAVE BLOOD KEPT FROZEN

A MULTICENTER, RANDOMIZED, DOUBLE-BLIND PLACEBO-CONTROLLED STUDY OF METHOTREXATE AND FOLIC ACID IN SYSTEMIC LUPUS ERYTHEMATOSUS A PHASE III TRIAL

Sponsor: Canadian Network for Improved Outcome In Systemic Lupus Erythematosus (CANIOS). This study is done with no pharmaceutical support. Faulding (CANADA) Inc. will be providing the methotrexate and placebo at no extra cost and Novopharm Quebec will provide us with the folic acid at no cost.

Principal Investigator: Paul R Fortin, MD, MPH, FRCP(C)

Institution: The Montreal General Hospital
1650 Cedar
Montreal, Qc, H3G 1A4
Tel.: (514)-937-6011 (2437)
FAX : (514)-934-8293

---

13) BANKING OF BLOOD SAMPLES:

SLE is a relatively uncommon disease and several blood test abnormalities may be associated with it. In the event that some of your blood tests would show such abnormalities, we may want to store your blood for further studies in the future. In that eventuality, the banked sample will be kept anonymously confidential. By giving your consent to the present study, you understand that some of your blood samples may be kept frozen for further study related to lupus research. I can withdraw any blood samples that would have been stored on me at any time and upon my immediate request without any other justification for such an action.

I agree to have some of my blood frozen for possible future research under the conditions described above.

_____	_____
Patient Signature	Investigator signature
_____	_____
(Please print name)	(Please print name)

_____
Witness signature

_____
(Please print name)

_____
Date

HÔPITAL GÉNÉRAL DE MONTRÉAL

FORMULE DE CONSENTEMENT

ÉTUDE MULTICENTRIQUE RANDOMISÉE A DOUBLE-INSU DU METHOTREXATE AVEC ACIDE FOLIQUE DANS LE LUPUS ÉRYTHÉMATEUX DISSÉMINE ESSAI - PHASE III

Présenté par: le réseau canadien pour l'amélioration des résultats dans le lupus érythémateux disséminé (ou CANIOS: "Canadian Network for Improved Outcome in Systemic Lupus Erythematosus"). Cette étude n'est pas supportée par une compagnie pharmaceutique. Faulding (CANADA) Inc. fournira le méthotrexate et son placebo sans frais supplémentaires. Novopharm Québec fournira les comprimés d'acide folique sans frais supplémentaires.

Chercheur principal: Paul R Fortin, MD, MPH, FRCP(C)

Institution: L'Hôpital Général de Montréal
1650 Cedar
Montréal, Qc, H3G 1A4
Tél.: (514) 937-6011 (2437)
FAX: (514) 934-8293

Nous vous invitons à participer à un projet de recherche à l'Hôpital Général de Montréal, l'un des centres à travers le Canada qui cherche à identifier un traitement plus efficace dans votre maladie, le lupus érythémateux disséminé (LED). Votre participation à cette étude est purement volontaire. Il se peut que vous n'ayez aucun bénéfice personnel supplémentaire en dehors de ceux dû au traitement traditionnel en participant à cette étude, mais des informations importantes peuvent être obtenues qui seront utiles à d'autres. Vous êtes libre de vous retirer de cette étude en tout temps sans que cela n'entraîne de pénalité ou n'affecte vos soins futurs. La nature de cette étude, les risques, les inconvénients, les inconforts et les bienfaits potentiels sont discutés ci-dessous. Nous vous demandons de discuter avec l'un d'entre nous toutes les questions que ce formulaire peut soulever.

1) BUT DE L'ÉTUDE, PROCÉDURE ET DURÉE:

BUT:

Ce projet de recherche a été développé pour évaluer la réponse au méthotrexate (MTX) avec l'acide folique des personnes atteintes d'un lupus érythémateux disséminé (LED) actif. Le MTX ressemble à une vitamine nommée folate et il inhibe une enzyme nommée la "dihydrofolate réductase". On croit que cette enzyme est importante pour la production de certaines de nos cellules inflammatoires, celles qui contrôlent certains phénomènes inflammatoires et immunitaires. Comme le LED est une maladie auto-immune et inflammatoire chronique, le MTX pourrait entraîner et maintenir une rémission chez les personnes atteintes d'un lupus actif. De plus, il pourrait avoir pour effet de diminuer les doses requises de stéroïdes. Le MTX se prend par la bouche une fois par semaine. Son utilisation est déjà approuvée dans une autre maladie auto-immune/inflammatoire: l'arthrite rhumatoïde. Des essais contrôlés randomisés ont démontré que le MTX contrôlait significativement mieux la maladie rhumatoïde qu'un placebo.

PROCÉDURE:

Nous vous demandons de participer à cette étude parce que votre médecin a jugé votre lupus plus actif et/ou parce que vous avez à prendre de hautes doses de stéroïdes (prednisone) pour le contrôler. Parce qu'un usage prolongé de hautes doses de stéroïdes a été associé à des effets secondaires sérieux (ostéoporose qui augmente les risques de fractures; intolérance au glucose semblable au diabète sucré; changements de votre aspect physique; un risque plus élevé d'élévation de votre pression artérielle ou de développer une maladie athérosclérotique dans les vaisseaux de votre coeur ou de votre cerveau), nous pensons que le MTX pourrait bien être une meilleure alternative que le traitement standard.

Nous voulons déterminer si le traitement avec MTX et acide folique conduira à un meilleur contrôle de votre maladie et s'il permettra à vos médecins de diminuer et/ou cesser l'utilisation des stéroïdes plus rapidement qu'avec le traitement standard. Même si nous ne savons pas à l'heure actuelle si cela sera le cas, nous savons cependant que le MTX a été utile dans d'autres maladies inflammatoires afin de contrôler la maladie et d'utiliser moins de stéroïdes.

Pour établir si le MTX sera utile dans le LED, nous devrons comparer des patients traités de façon conventionnelle [anti-inflammatoires non-stéradians (AINS), stéroïdes et/ou antimalariques comme le Plaquenil™ (hydroxychloroquine)] avec d'autres patients traités avec MTX et acide folique en plus du traitement standard.

Si vous acceptez de participer, l'on déterminera dans quel groupe vous serez par une méthode basée purement sur le hasard (comme si l'on jouait à pile ou face); les deux groupes sont soit le traitement standard avec placebo, soit le traitement standard avec MTX. Le placebo est une pilule identique au MTX mais dans laquelle il n'y a aucun médicament (c'est comme une pilule de sucre). De cette façon, vous ne serez pas biaisé dans votre interprétation des résultats puisque vous ne saurez pas si vous recevez le MTX ou un placebo.

Vous verrez deux médecins à chaque visite: le premier évaluera vos progrès cliniques par rapport au traitement que vous recevez alors que le second supervisera tous les aspects de votre état de santé général. Ni l'un ni l'autre ne sera au courant du traitement que vous recevrez mais le deuxième médecin saura quels sont les autres médicaments que vous prenez de même que les résultats de vos tests de laboratoire. Ce médecin ajustera vos doses de médicament selon des normes pré-établies. Une troisième personne, notre coordonnateur de recherche, vous donnera le MTX ou le placebo et aura accès à la clef pour décoder le médicament que vous recevez si cela devenait nécessaire avant la fin du projet.

Au départ, tous les patients recevront le traitement standard et vous recevrez en plus une dose de MTX ou de placebo une fois par semaine, et d'acide folique (2.5 mg/jour) pour les six autres jours. Une fois que vous montrez des signes de réponse au traitement, les stéroïdes seront diminués selon une stratégie que l'on vous donnera par écrit. Pour un certain nombre de participants, la dose de MTX (ou de placebo) devra être modifiée selon votre tolérance à ce médicament, vos symptômes et les tests de laboratoires. Nous déterminerons l'utilité du MTX en comparant:

1) la diminution de l'activité moyenne de la maladie (telle que mesurée sur un formulaire que votre médecin remplira à chaque visite) dans le groupe "MTX" en

comparaison avec le groupe "traitement standard seulement",
2) Le temps avant d'induire une rémission entre les deux groupes,
3) la diminution des doses de stéroïdes entre les deux groupes,
4) la fréquence des rechutes du lupus,
5) la fréquence des effets secondaires dus au MTX.

Si un patient compte deux rechutes ou plus lors de son traitement (avec ou sans MTX), il sera retiré de l'étude et d'autres traitements pourront alors être proposés selon l'expertise du médecin traitant.

Vous trouverez ci-après quelques situations qui augmentent les risques de développer des effets secondaires au MTX. Si l'une de ces situations se retrouve chez vous, vous ne serez alors pas éligible pour cette étude. Ces situations incluent:

1) une insuffisance rénale avec une créatinine sérique de $\geq$ 175 µmol/l (SI unités) or 2.0 mg/dl.
2) des globules blanc ou des plaquettes bas.
3) la présence d'anomalies hépatiques connues ou détectées par une prise de sang.
4) une ingestion d'alcool de plus de 3 col sec (1.5 onces) de spiritueux, 3 bouteilles de bière ou l'équivalent par semaine. Il est impossible de réglementer la prise d'alcool mais nous décourageons fortement tout usage d'alcool tout au long de cette étude parce que l'alcool augmente les risques de toxicité hépatique (du foie).
5) une combinaison d'obésité importante (plus d'un 1/3 au-dessus du poids idéal) avec la présence d'un diabète requérant de l'insuline.
6) L'utilisation de médicaments contenant du sulfa. L'acétaminophene (Tylenol™) sera préféré aux anti-inflammatoires non-stéradians lorsque cela est possible.
7) Une histoire antérieure d'hypersensibilité ou d'intolérance au MTX ou à l'acide folique.
8) Une fibrose interstitielle pulmonaire définie par une radiographie pulmonaire anormale ou une diffusion diminuée (DLCO < 70% de la valeur prédite) sans évidence d'hypertension pulmonaire.

DURÉE:

Tous les patients recevront le médicament expérimental (placebo ou MTX) pour au moins douze mois. Si un patient est sous stéroïdes, tous les efforts seront faits pour diminuer les stéroïdes selon le plan pré-déterminé. A la fin de cette étude à double-aveugle, il y aura une étude ouverte pendant laquelle les patients qui ont répondus au MTX et ceux qui recevaient un placebo se verront offrir de continuer/commencer le MTX pour une période additionnelle de six mois.

2) MESURE DE LA RÉPONSE ET DES PROGRÈS:

Je comprends que je devrai venir à l'Hôpital Général de Montréal pour une évaluation préliminaire de triage afin de déterminer si je suis éligible à cette étude. Je discuterai avec mon médecin de mon histoire médicale et des médicaments que j'ai pris dans le passé. J'aurai un examen physique et des tests de laboratoires, soit une prise de sang (pas plus de huit cuillerées-à-soupe) et une analyse des urines. Une radiographie (rayons-X) de mes poumons sera faite si je n'en ai pas eu dans les trois mois précédant.

Parce que le MTX peut entraîner des malformations sévères à l'embryon/foetus pendant une grossesse, les femmes devront s'assurer de ne pas être enceintes pendant toute la durée de cette étude et elles devront prendre des mesures de contraception reconnues efficaces. En effet, les hormones que l'on retrouvent dans les anovulants oraux pourraient être associées à un risque plus élevé d'exacerbation de votre lupus. Cette forme de contraception parfois déconseillée chez les femmes atteintes du lupus, sera acceptée dans cette étude. Le choix de la méthode de contraception sera laissé à votre discrétion après consultation avec votre médecin. Un test de grossesse sera fait chez toute les femmes pré-ménopausée pour s'assurer qu"elle ne sont pas enceinte au moment de commencer l'étude.

Je comprends que l'on me demandera de revenir à l'Hôpital Général de Montréal moins de dix jours après la Visite d'Évaluation pour une Visite de Qualification au cours de laquelle je terminerai l'évaluation de mon cas. Mon médecin remplira alors plusieurs formulaires sur mon état de santé, et l'on me demandera de remplir un questionnaire d'auto-évaluation sur ma santé générale et sur l'importance que j'attribue à mon état de santé présent. Si je suis éligible à participer à cette étude et que je le désire, l'on me désignera dans lequel des deux groupes thérapeutiques je serai d'après un tirage au sort (ou en d'autres mots le médicament que je recevrai me sera décerné purement par chance). C'est à ce moment que l'on me donnera mon premier paquet du médicament expérimental (des pilules dans un emballage libellé) de même que des instructions sur comment prendre ces pilules.

La première visite contrôle sera faite deux semaines plus tard, puis une fois par mois pour un total de quartorze visites. J'aurai des tests sanguins (pas plus de quatre cuillerées-à-soupe) et urinaire (vous n'avez pas à être à jeun cette fois-ci) à chaque visite de contrôle. Au moment de la visite de contrôle, j'aurai un examen physique, on me questionnera sur ma santé et j'aurai à compléter un questionnaire d'auto-évaluation. Les médicaments de l'étude seront distribués à intervalles de quatre semaines, et je dois rapporter aux visites suivantes tout médicament qui n'a pas été utilisé de même que toutes les bouteilles et emballages.

A chacune des visites de contrôle mensuelles, mon état de santé sera réévalué afin de déterminer s'il y a eu une amélioration ou une détérioration de mon lupus. On pourra me demander de diminuer les doses de quelques-uns de mes médicaments. Si je développe des effets indésirables sévères, je pourrai cesser le traitement et je reviendrai à l'hôpital pour une Visite Finale. On me demandera aussi de cesser cette étude si je suis incapable de me soumettre aux exigences de celle-ci.

Si les doses de mon médicament expérimental (méthotrexate ou placebo) doivent être ajustées, on me demandera de revenir pour une Visite Intérimaire deux semaines après la Visite Contrôle pour une autre prise de sang et pour vérifier la dose de médicament que je devrai prendre.

A la fin de cette étude ou dans le cas où je devrais me retirer de cette étude plus tôt, je reviendrai pour une Visite Finale. J'aurai un examen physique, on me questionnera sur mon état de santé et l'on procédera à des tests sanguins (pas plus de huit cuillerées-à-soupe) et urinaires. Je remplirai le questionnaire d'auto-évaluation et je rapporterai la dernière bouteille de pilules que j'ai utilisée. A la fin de la période de traitement initiale, le suivi dépendra du fait que j'ai complété l'étude ou non et de ce que je sois intéressé à recevoir du méthotrexate pour une période additionnelle de six mois. De cette façon, ceux qui reçoivent déjà du méthotrexate avec une amélioration de leur symptômes et qui le désirent pourront continuer à en prendre. De même, ceux qui recevaient le placebo dans

la première partie de l'étude, auront la possibilité d'essayer le MTX sous supervision médicale. Si je n'ai pas complété cette étude, on me demandera de revenir à l'hôpital pour une visite de contrôle 12 mois après la Visite de Qualification pour un examen physique, une histoire médicale, et des tests de sang (pas plus de huit cuillerées-à-soupe) et d'urine. On vous demandera également de remplir un formulaire d'auto-évaluation.

3) MODE D'UTILISATION DU MÉDICAMENT EXPÉRIMENTAL:

Je prendrai le nombre exact de pilules du médicament expérimental (méthotrexate ou placebo) tel que prescrit une fois par semaine et je prendrai un comprimé d'acide folique à chaque jour pour le reste de la semaine (sauf le jour où je prends le MTX). Je continuerai tous mes autres médicaments tels qu'ils ont été prescrits par mon médecin traitant. Je comprends que l'on me demande d'utiliser une forme fiable de contraception si je suis une femme en âge d'avoir des enfants.

4) RISQUES ET INCONFORTS:

Les effets secondaires du MTX peuvent être: une diminution de la production des cellules de votre sang, une irritation du foie, de l'estomac, des poumons, des muqueuses de votre bouche ou de votre nez, un rash, des nausées, des vomissements ou de la diarrhée. Récemment, il a été démontré que l'utilisation d'acide folique après la prise du MTX peut diminuer ces effets secondaires. Un effet possible mais très rare relié à une utilisation prolongée du MTX est qu'il pourrait augmenter légèrement des risques de développer un cancer (sang, foie, sein, prostate) dans l'avenir. Le MTX a été utilisé comme médicament anti-arthritique et pour les maladies de la peau sur des milliers de patients depuis les années 1950. Pendant toute cette période de temps, quelques patients seulement ont développé des cancers pour lesquels le MTX aurait possiblement joué un rôle.

Cependant il n'y a aucune garantie que je n'ai quelque bénéfice que ce soit en participant à cette étude, il se peut que l'activité de la maladie lupique soit mieux contrôlée et il se peut que l'on puisse diminuer la dose de prednisone. Il est également possible que ce traitement n'apporte rien de plus que le traitement standard ou que les effets secondaires dus au MTX soient plus importants que les bénéfices possibles. Il pourrait aussi y avoir des risques pour moi et pour un embryon qui ne sont pas reconnus aujourd'hui. Des études futures seront probablement développées à partir des résultats de l'étude présente.

5) TRAITEMENT STANDARD ET RÉSULTATS DE LA PRÉSENTE ÉTUDE:

TRAITEMENT STANDARD:

En raison du "design" de cette étude, le traitement standard avec les antimalariques et les stéroïdes sera continué. L'usage d'anti-inflammatoires non stéradians sera évité si possible. Cependant, leur usage sera autorisé dans les arthrites sévères et les cas de pleuro-péricardites. Finalement, si mon lupus devait empirer au point de nécessiter des médicaments immunosuppresseurs encore plus puissants comme l'azathioprine ou la cyclophosphamide, l'étude sera terminée pour que je puisse recevoir ces médicaments.

RÉSULTATS NOUVEAUX:

Une analyse intérimaire sera faite à mi-chemin au cours de cette étude pour vérifier si le MTX a un

effet significatif si important qu''il faudrait le donner immédiatement à tous les participants. Ces résultats vous seront transmis immédiatement si c'est le cas. Toute autre découverte qui surviendrait au cours de cette étude vous seraient également transmise immédiatement.

6) CONFIDENTIALITÉ:

Mon dossier de recherche demeurera confidentiel sauf si la loi en décide autrement. Pour vérifier les données de cette étude, des évaluateurs du Bureau de Protection de la Santé (CANADA-BPS) pourraient avoir à réviser certains dossiers médicaux (incluant mais non pas limité aux données recueillies pendant cette étude), des résultats de tests de laboratoire et d'autres dossiers en rapport avec ma santé. Les dossiers de recherche à l'Hôpital Général de Montréal et à l'Institut de Recherche de l'Hôpital Général de Montréal pourront être révisés par les membres des Comités d'Éthique en Recherche et des Essais Cliniques afin de vérifier et de contrôler que tous les règlements qui s'appliquent à la recherche chez l'humain soient respectés. En aucun cas, l'identité des participants ne sera dévoilée dans les communications et publications scientifiques découlant de cette étude.

7) COUT DU TRAITEMENT ET COMPENSATIONS:

Il n'y aura pas de coûts associés au médicament expérimental ni aux visites à la clinique, aux services des médecins ou aux tests de laboratoire demandés pour cette étude. Vous ne serez pas remboursé pour le coût de tout autre médicament (en dehors du médicament expérimental) que vous prenez actuellement et continuerez à prendre pendant l'étude. Vous ne recevrez pas de compensations monétaires pour votre participation à cette étude.

Si je devais avoir un accident ou devenir malade comme conséquence directe de ma participation à cette étude, des traitements seront disponibles à l'Hôpital Général de Montréal.

8) PERSONNE A CONTACTER SI VOUS AVEZ DES QUESTIONS:

Cette étude m'a été expliquée par le docteur Paul R. Fortin et l'on a répondu à mes questions. Si je devais avoir d'autres questions au sujet de l'étude, je peux téléphoner au Dr Paul R. Fortin au (514) 937-6011, ext. 2437. Si je veux en savoir plus quant à mes droits comme participant à une étude à l'Hôpital Général de Montréal, je peux téléphoner au bureau du Dr. Blair N. Whittemore au (514) 937-6011, ext. 3013.

9) REGLEMENT HOSPITALIER:

Tout participant à un essai médicamenteux à l'Hôpital Général de Montréal doit obtenir un numéro de dossier médical à l'Hôpital Général de Montréal (si cela n'a pas déjà été fait). Pour votre protection, une copie de ce résumé de l'étude avec les numéros de téléphone d'urgence appropriés, de même qu'une copie de ce formulaire de consentement sera inclue dans votre dossier médical hospitalier.

Au début de votre participation à cette étude, vous recevrez une carte d'information avec le nom de cette étude, de même que les noms des médicaments que vous recevrez peut-être et des numéros de téléphones d'urgence.

10) DECLARATION DE CONSENTEMENT:

J'ai reçu une copie de ce formulaire de consentement que je pourrai conserver. En signant ce formulaire de consentement, je confirme que j'accepte de participer à cette étude et que j'accepte les termes et conditions ci-dessus. Je confirme aussi que j'ai revisé ce document et que je l'ai lu et compris avant de le signer. Je reconnais que j'ai attentivement pese le pour et le contre de ma participation, et qu'après l'avoir fait, je désire participer à l'étude telle que décrite ci-dessus.

De plus, je comprends que ma participation à cette étude pourrait être terminée prématurément sans mon consentement si mes médecins devaient juger qu'il est plus adéquat d'agir ainsi.

11) PARTICIPATION VOLONTAIRE:

MA PARTICIPATION A CE PROJET DE RECHERCHE EST PUREMENT VOLONTAIRE. J'ai le droit de refuser d'y participer ou de me retirer n'importe quand au cours de l'étude sans que cela n'affecte en rien mes traitements habituels.

12) CONSENTEMENT:

J'ai lu les explications de ce formulaire de consentement, on m'a donné la chance d'en discuter et l'on a répondu à mes questions. Je suis entièrement libre de participer ou non à cette étude et cela n'affectera en rien mes soins médicaux habituels. Je consens donc à participer à cette recherche sur l'utilité du méthotrexate dans le traitement du lupus érythémateux disséminé qui sera menée sur 100 sujets à travers le Canada.

_____

Signature du patient

_____

(lettres moulées)

_____

Signature du chercheur

_____

(lettres moulées)

_____

Signature du témoin

_____

(lettres moulées)

_____

Date

APPENDICE: CONSENTEMENT POUR GARDER DU SANG CONGELÉ

ETUDE MULTICENTRIQUE RANDOMISEE A DOUBLE-INSUE DU METHOTREXATE AVEC ACIDE FOLIQUE DANS LE LUPUS ERYTHEMATEUX DISSEMINE
ESSAI - PHASE III

Présenté par: le réseau canadien pour l'amélioration des résultats dans le lupus érythémateux disséminé (ou CANIOS: "Canadian Network for Improved Outcome in Systemic Lupus Erythematosus"). Cette étude n'est pas supportée par une compagnie pharmaceutique. Faulding (CANADA) Inc. fournira le méthotrexate et son placébo sans frais supplémentaires.

Chercheur principal: Paul R Fortin, MD, MPH, FRCP(C)

Institution: L'Hôpital Général de Montréal
1650 Cedar
Montréal, Qc, H3G 1A4
Tél.: (514) 937-6011 (2437)
FAX: (514) 934-8293

---

13) BANQUE DE PRELEVEMENT SANGUINS:

Le LED est une maladie relativement rare et plusieurs anomalies peuvent étre décelées dans les tests de sang. Dans l'éventualité que vos tests sanguins devait démontrer de telles anomalies, nous aimerions pouvoir conserver votre prise de sang avec l'intention de l'étudier de façon plus approfondie dans le futur. En donnant mon consentement à cette étude, je comprends que mon échantillon sanguin pourra être conservé pour des études reliées au lupus. Je peux demander que l'on détruise les prises de sang que j'aurai donné et ce en tout temps et sans que je n'ai à fournir aucune explications pour une telle demande.

J'accepte à ce que mon sang soit gardé congelé pour des recherches dans le future selon les termes décrits ci-dessus.

_____	_____
Signature du patient	Signature du chercheur
_____	_____
(lettres moulées)	(lettres moulées)
_____	
Signature du témoin	
_____	
(lettres moulées)	
_____	
Date	

Appendix D

Inclusion and exclusion criteria for SMILE, taken from the SMILE protocol. This section contains 2 pages.

SMILE Inclusion Criteria

1. Men and women with a diagnosis of systemic lupus erythematosus according to the American College of Rheumatology Criteria (ACR, formally ARA or American Rheumatology Association) (Tan et al., 1982).
2. Age of at least 18 years.
3. Female subjects of child-bearing potential must have a negative pregnancy test and must practice an effective means of contraception during and for one month after the end of the study. They should accept that they should not attempt to get pregnant during the period of the study.
4. Subjects with active disease as defined by a total score of the Systemic Lupus Activity Measure (SLAM) (Liang et al) of at least 8. This SLAM score justifies further treatment.
5. Subjects with damage as defined by a total score of the Systemic Lupus Collaborating Clinics/American College of Rheumatology (SLICC/ACR) (Gladman et al) of less or equal to 15. This amount of damage suggest that no therapeutic intervention is likely to improve the subject's overall condition or function.
6. Subjects on nonsteroidal anti-inflammatory drugs (NSAIDs), prednisone, or antimalarial drugs (chloroquine sulfate or hydroxychloroquine) must be on stable doses for at least four weeks preceding the study.
7. Subjects admitted into the study can have other medical conditions as long as these conditions or their treatment will not interfere with the experimental medications and assessments.
8. Subjects who can understand either French or English and who give a written informed consent according to local regulation or ethic committee recommendations and in accordance with the provisions of the Declaration of Helsinki, Hong Kong, 1989.

SMILE Exclusion Criteria

1. A previous history of hypersensitivity or intolerance to either methotrexate (MTX) or folic acid (FA).
2. Subjects with either a total SLAM score of less than 8 or a total SLICC/ACR score of more than 15.
3. Inability to comply with instructions or a history of medical non-compliance. Inability to comply will be defined by the patient's inability to correctly answer written questions based on

information provided in the written consent document. Determinations of prior non-compliance will be left to the discretion of the treating physician.

4. Subjects who have received intra-articular or intramuscular corticosteroids in the four weeks prior to study entry.

5. Clinically significant acute or chronic liver disease with the exception of autoimmune liver disease. Significant liver disease is defined by reproducible abnormal liver function tests (LFTs) (twice above the upper limit of normal) immediately prior to entry. Autoimmune liver disease is an exclusion criteria once ethanol-induced, viral or other metabolic causes of liver disease are excluded.

6. Alcohol use in excess of 2 ounces of 100 proof liquor or its equivalent per week. (All alcohol use is to be discouraged during this study.)

7. Insulin requiring diabetes mellitus with morbid obesity (>33% ideal body weight).

8. Renal impairment such that the serum creatinine is $\geq$ 175 μmol/l (SI units) or 2.0 mg/dl.

9. Interstitial lung disease as defined by an abnormal chest x-ray or decrease diffusion capacity (DLCO < 70% of predicted) without evidence of pulmonary hypertension.

10. White blood cell (WBC) count < 3,000/mm^3 and/or platelet count < 80,000/mm^3.

11. Prior use of methotrexate to treat systemic lupus erythematosus.

12. Use of sulfa drugs that may potentiate the folate antagonistic effects of MTX.

13. Non steroidal anti-inflammatory drugs will be allowed throughout the trial unless there is evidence of renal failure or other contra-indications to these drugs. Their concomitant use with methotrexate is routine in patients with rheumatoid arthritis.

14. Use of another cytotoxic or immunosuppressive drug such as cyclophosphamide, azathioprine, chlorambucil , cyclosporin or Trimetoprime currently or in the preceding six months.

15. Current participation in any other drug trial or participation in such a trial in the previous one month.

16. Serologic evidence of infection with human immunodeficiency virus (HIV).

17. Biologic potential for pregnancy and not utilizing effective means of contraception.

18. Recently ($\leq$ 6 months) diagnosed malignancy.

19. Vitamin B12 deficiency.

Appendix E

The investigators and co-investigators of SMILE and their affiliations. This section is 1 page long.

Investigators and co-investigators in the Study of Methotrexate in Lupus Erythematosus (SMILE)

Principal Investigator: Paul R. Fortin, MD, MPH; *University Health Network, Toronto, Ontario (formerly at McGill University Health Centre, Montreal, Quebec)*

Co-Investigators (in order of number of patients entered): Michal Abrahamowicz, PhD; Diane Ferland, RN, BScN; Ann E. Clarke, MD, MSc; John Penrod, PhD; *McGill University Health Centre, Montreal, Quebec.* Diane Lacaille, MD, MHSc; John M. Esdaile, MD, MPH; Howard B. Stein, MD; *Arthritis Research Centre of Canada, Vancouver, BC.* Douglas Smith, MD; *Ottawa Hospital, General Campus, Ottawa, Ontario.* Michel Zummer, MD; Jean-Pierre Mathieu, MD; Line Duchesne, MD; Suzanne Mercille, MD; Pierre Dagenais, MD, PhD; *Hôpital Maisonneuve-Rosemont, Montreal, Quebec.* Janet E. Pope, MD, MPH; *St. Joseph's Hospital, London, Ontario (formerly at Victoria Hospital, London, Ontario).* Steven Edworthy, MD; Susan Barr, MD; Garry Morris, MD; *Calgary Health Sciences Centre, Calgary, Alberta.* Micheal Starr, MD; *St. Mary's Hospital, Montreal, Quebec.* Vivian Bykerk, MD; *Mount Sinai Hospital, Toronto, Ontario (formerly at Credit Valley Hospital, Mississauga, Ontario).* S. Bernatsky, MD; Janice Canvin, MD; Hani S. El-Gabalawy, MD; Christine Peschken, MD; *Winnipeg Health Sciences Centre, Winnipeg, Manitoba.* Alfred Cividino, MD; *MacMaster-Cherooke Hospital, Hamilton, Ontario.* Jean-Luc Senécal, MD; Jean-Pierre Raynauld, MD; Eric Rich, MD; *Hôpital Notre-Dame, Montreal, Quebec.* C. Kirk Osterland, MD; Carol A. Yeadon, MD; *Royal Victoria Hospital, Montreal, Quebec.* André Beaulieu, MD; Simon Carette, MD; *Centre Hospitalier de l'Université Laval, Quebec City, Quebec.* Gilles Boire, MD, MSc; *Centre Hospitalier de l'Université de Sherbrooke, Sherbrooke, Quebec.*

Research Assistants: Rhondda Morrison, RN, MSN; Elaine S. Clark, Karen Rangno, RN; *Arthritis Research Centre of Canada, Vancouver, BC.* Paula Dale; *Ottawa General Hospital, Ottawa, Ontario.* Deborah L. Fenlon, BScN; Wendy Curran, Lynda Bere, RN; *Victoria Hospital, London, Ontario.* Rosario Talavera, Beverly Green, BN, BScN, Elisia Teixeira, RN, BN; *Calgary Health Sciences Centre, Calgary, Alberta.* Jackie O'Farrell, RN; *Credit Valley Hospital, Mississauga, Ontario.* Sharyn Wood, RN, Ann Huggard, RN; *Winnipeg Health Sciences Centre, Winnipeg, Manitoba.* Joy Zahavich, RN; *MacMaster-Cherooke Hospital, Hamilton, Ontario.* Mariette Prave, RN; *Centre Hospitalier de l'Université Laval, Quebec City, Quebec.*