Randomized Clinical Trials and Longitudinal Observational Studies in Systemic Lupus Erythematosus: Consensus on a Preliminary Core Set of Outcome Domains

JOSEF S. SMOLEN, VIBEKE STRAND, MARIO CARDIEL, STEVEN EDWORTHY, DANIEL FURST, DAFNA GLADMAN, CAROLINE GORDON, DAVID A. ISENBERG, JOHN H. KLIPPEL, MICHELLE PETRI, LEE SIMON, PETER TUGWELL, and FREDERICK WOLFE

ABSTRACT. The OMERACT module on systemic lupus erythematosus (SLE) dealt with the definition of preliminary core sets of outcome domains for randomized clinical trials (RCT) and longitudinal observational studies (LOS). After lectures introducing the problems and addressing the key issues, 6 discussion groups, 3 each for LOS and RCT, discussed and weighted more than a dozen possible items for use as outcome domains. The means of the respective 3 groups were calculated. For both RCT and LOS the same outcome domains received more than 10 of a total maximum of 100 points: disease activity, health related quality of life (HRQOL), damage, and toxicity/adverse events. However, the weights for HRQOL and damage were different for LOS and RCT. A final vote led to the acceptance of these 4 variables as a preliminary core set for outcome in SLE by more than 80% of the participants. This core set will allow for improved design, performance, and evaluation of future studies in SLE. However, a number of domains not included in the core set were regarded as important for further analysis and research. (J Rheumatol 1999;26:504–7)

Key Indexing Terms:
SYSTEMIC LUPUS ERYTHEMATOSUS CLINICAL TRIALS OUTCOME

The OMERACT Meeting of the lupus (SLE) investigators was held to define preliminary core sets of outcome domains to be recommended for randomized controlled trials (RCT) and longitudinal observational studies (LOS).

Participants were first provided with background information on outcome domains that have been previously developed. These presentations related to current knowledge on measures of disease activity, organ damage, quality of life (QOL), treatment approaches, and other items potentially related to outcome in SLE[1–4]. The participants were then assigned to one of 6 discussion groups, 3 each attempting to define outcome core variables for RCT and for LOS. The 6 groups were asked to independently (a) judge the items listed for their principal value as outcome domains; (b) add domains deemed important in the context of the respective task, RCT, or LOS; (c) define those items that should be included in a core set for outcome measures in RCT and LOS, respectively; and, finally, (d) determine the outcomes that needed further research. Table 1 lists all items deemed possibly important as outcome measures and includes those recommended by the OMERACT Steering Committee, those previously identified by the Steering Committee as having potential value for inclusion after further discussion, those raised during the discussions after plenary presentations and particularly in the small discussion groups.

MATERIALS AND METHODS

Within each discussion group, a nominal group process allowing free discussion by participants on each item was encouraged. Each group first supplemented the original list with items to be considered in the discussion. Then the individual items were evaluated and particular emphasis was put on the precise definition of the items, evidence for their utility as outcome measures for the individual segments, and the availability of previously validated measures. For example, "damage" was regarded as important and

From the University of Vienna, Austria; Stanford University, San Francisco, USA; Istituto Nacionale de Nutricion, Mexico City, Mexico; University of Calgary, Calgary, Canada; Virginia Mason Research Center, Seattle, USA; University of Toronto, Toronto, Canada; University of Birmingham, Birmingham; University College, London, UK; National Institutes of Health, Bethesda; John Hopkins University School of Medicine, Baltimore; Harvard School of Medicine, Boston, USA; University of Ottawa, Ottawa, Canada; Arthritis Research and Clinical Centers, Wichita, USA.

J.S. Smolen, MD, Professor of Internal Medicine, University of Vienna; V. Strand, MD, Associate Professor of Medicine, Stanford University; M. Cardiel, MD, Istituto Nacional de Nutricion; S.M. Edworthy, MD, Associate Professor of Medicine and Community Health Sciences, University of Calgary; D. Furst, MD, Director Arthritis Clinical Research, Virginia Mason Research Center; D. Gladman, MD, Deputy Director, Center for Prognosis Studies in Rheumatic Diseases, University of Toronto; C. Gordon, MD, Medical School, University of Birmingham; A. Isenberg, MD, ARC Professor of Rheumatology, University College London; J.H. Klippel, MD, Clinical Director, NIAMS, National Institutes of Health; M. Petri, MD, Professor of Medicine, Johns Hopkins University School of Medicine; L.S. Simon, MD, Director, Postgraduate Medical Education, Harvard School of Medicine; P. Tugwell, MD, Professor of Medicine, University of Ottawa; F. Wolfe, MD.

Address reprint requests to Dr. J.S. Smolen, Division of Rheumatology, Department of Internal Medicine III, University of Vienna, Waehringer Guertel 18-20, A-1090 Vienna, Austria.

Table 1. Items suggested for inclusion into a core set and items recommended for discussion and possible inclusion by the steering committee.

Domains Recommended for Inclusion	Domains Recommended for Discussion
Disease activity	Disease severity
Damage	Fatigue
Health related quality of life	Fibromyalgia
Toxicity/adverse events	Patient global activity
Economic costs	Patient global severity
	Physician global activity
	Physician global severity
	Hypertension
	Disability
	Psychosocial factors
	Serology
	Work status
	Comorbidity
	Use of drugs/steroid sparing
	Osteoporosis
	Pregnancy

has been sufficiently defined and validated by the Systemic Lupus International Collaborating Clinic/American College of Rheumatology (SLICC/ACR) Damage Index[3]. The toxicity and adverse event domain included death, since reporting of mortality is required in all trials. In contrast, assessment of flares, although regarded as possibly important, is hampered by lack of agreement on the definition and validation of a "flare."

After ample discussion time, each of the 6 groups prepared a final list of domains recommended for inclusion in the preliminary core set by assigning relative weights to each item. A total of 100 points could be used by every participant, and everyone was free to assign as many points to an item as they wished: it was possible to assign all 100 points to a single item, to distribute the points evenly among all items (e.g., giving 4 points each to 25 different items), or to distribute the points in a weighted fashion among just a few, selected items, never exceeding a total of 100 points per individual participant. The scores of all group members were then summed and an average calculated. The averages for the 3 groups discussing RCT and for the 3 discussing LOS were combined and a final average score determined. This allowed a final ranking of the domains list for RCT and LOS, respectively.

The results were briefly presented by the discussion group coordinators at the plenary session and then subjected to a final electronic voting process by all participants. A cutoff value of 10 points was determined as being the minimal level to enter the final voting process.

RESULTS

Core set of outcome domains for RCT and LOS. Table 2A lists the results obtained by the groups discussing RCT, while Table 2B displays the core domains derived for LOS. Interestingly, for both types of studies disease activity, health related quality of life (HRQOL), damage, and toxicity/adverse events reached > 10 average points as the most important measures for outcome.

The sum of points for these 4 domains was 74 for RCT and 75 for LOS. Thus, from a total of 100 points that could be allocated to individual variables, the discussion groups allocated only 26 or 25 points, respectively, to more than a dozen other possible variables.

The domains received slightly different weights, comparing RCT to LOS (Table 2): for RCT, disease activity was given the greatest weight, twice that of the QOL domain (which was deemed the second most important variable for RCT), and almost 3 times the weight of damage and toxicity. In contrast, for LOS, damage was believed to be the second most important domain and had an average weight very close to that of disease activity. Indeed, disease activity for LOS had obtained only two-thirds of the weight it had been given for RCT,while damage obtained a 50% higher weight for LOS than RCT; HRQOL was considered equally important for both types of studies.

Other domains. Many other items were regarded as potentially important (Table 3). Thus, economic factors were deemed very important, but it was felt that a refinement of instruments was still needed and that their assessment required too much effort to be recommended for all trials. Work status was regarded as imprecisely defined, since the occupational status of women historically has not been acknowledged as employment in the same sense as employment outside the home. Fibromyalgia may be differently perceived in different regions, and therefore a cross cultural use of such an item may not be helpful until further research into its definition and geographical, cultural, and racial variations has been performed.

Some domains were believed to be included already in

Table 2. Results of weighting domains by the discussion groups. Indicated are the means of the individual results from 3 discussion groups for each RCT and LOS. In every discussion group, the means were calculated from the individual assignments by up to 15 participants.

A. Randomized Clinical Trials (Items)	Points	B. Longitudinal Observational Studies (Items)	Points
Disease activity	34	Disease activity	22
Health related quality of life	15	Damage	18
Damage	13	Health related quality of life	14
Toxicity/adverse events*	12	Toxicity/adverse events*	11
TOTAL (of 100 points)	74	TOTAL (of 100 points)	75
Final vote: "Can we agree that all 4 domains be assessed in all RCT in SLE?" 85% Yes, 13% No, 2% Don't know		Final vote: "Can we agree that all 4 domains be assessed in all LOS in SLE?" 83% Yes, 15% No, 2% Don't know	

*Includes death

Table 3. Additional items, with respective weight points, believed to be important for further discussion and/or research.

A. Domains for Further Discussion (RCT)		B. Domains for Further Discussion (LOS)	
Physician global activity	8	Patient global activity	6
Economics/utility	5	Comorbidities	4
Patient global activity	5	All others (see A)	< 3
Fatigue	4		
Patient global severity	4		
Psychosocial	< 3		
Flare	< 3		
Steroid sparing	< 3		

the selected domains: variables associated with serology are included in some disease activity domains; nevertheless, questions about the validity and reliability of individual serological variables remain and further definition may be needed. The physician's global activity assessment was felt to be important, and is already included as part of overall disease activity in some, but not all, instruments. Patient global activity was similarly reflected in the (patient-generated) QOL instruments. The measurement of disability in SLE requires further research and this element could be included in the QOL instruments.

Furthermore, the concept of toxicity/adverse effects recorded separately was not completely clear, because some of the items might already be included in the damage index and there should be no duplication of items. Nonetheless, it was felt that demographics, death, hospitalization, infection, hypertension, and side effects of specific drugs should be considered.

The concept of disease severity was deemed important, but its definition was unclear and there were not well validated instruments available to report it as yet. In particular, the distinction between the extent of disease activity and disease severity when occurring in critical anatomical locations, such as the major organs, requires further elucidation. The definition of flare is also still unclear. Flare is already semantically partly included in disease activity, but beyond that, use of such a domain requires further definition and validation.

Ranks of 0 were given to some items by each of the groups; these items were eliminated from the original domain list or were scored as "0" when included in other domains.

Final voting. The results of the group discussions on both RCT and LOS were then presented to the plenary. As noted above, not only was the relative ranking of the top 4 items among all 3 groups dealing with RCT or with LOS consistent (data not shown), but the RCT and the LOS groups recommended the same items as the core set for studies (Table 2).

More than 60 delegates then cast electronic ballots indicating they were in favor, against, or did not know whether the domains recommended through the nominal group process should be accepted. Of the delegates 85% voted in favor of adopting the 4 items disease activity, damage, QOL, and toxicity/adverse events as core domains to be included in studies of outcome; the vote was similar for both RCT and LOS. It was further recommended that research priorities would include the consideration of economic costs, the role of serology, fatigue, fibromyalgia, disease severity, steroid sparing effects, physical disability, and psychosocial components.

DISCUSSION

SLE is a heterogeneous disease, and outcome is affected by numerous factors. The results obtained in the course of the OMERACT SLE process constitute a consensus on the recommendation of a core set of outcome measures for use in studies of SLE. Interestingly, the same core set was determined for randomized clinical trials and for longitudinal observational studies. It consists of the variables disease activity, damage, health related QOL, and adverse effects/toxicity (including death). In spite of the identity of core domains, the weights given to the individual variables by the discussion groups dealing with RCT or LOS were substantially different. In particular, damage was regarded as much more important for LOS than it was for RCT, and this decision clearly has face validity, given the usually much shorter duration of RCT. Nevertheless, recording of disease activity was deemed most important for both RCT and LOS, and HRQOL was perceived as equivalently important in both RCT and LOS.

These core set domains obtained a surprisingly high number of total weight points (about 75); thus only about 25 points remained for assignment to more than a dozen other variables for both LOS and RCT. This weighting of the result is also reflected by its acceptance by more than 80% of the participants supporting the perceived importance and relative unanimity ascribed to the SLE core set. Perhaps not surprisingly, the domains corresponded almost perfectly with domains recommended in the general LOS module (Health Status, Disease Process, Damage, Mortality, and Toxicity/Adverse Reactions), with the difference that there Death was considered a separate domain[5].

The 4 domains are recommended for inclusion in all future RCT and LOS, either as part of the selected outcome variables or in addition to the ones chosen for evaluation driven by the major hypothesis. This will not only ensure that such studies allow the evaluation of the 4 domains deemed most important for overall outcome in SLE, but also will ensure some degree of comparability between studies with respect to such major outcomes.

It should be borne in mind that the 4 domains recommended are not more than a minimal core set. Other items may be important and it was also suggested that research

efforts concentrate on some of the other items deemed to be of importance for outcomes in SLE. In addition, items not dealt with at all within the OMERACT context may be of significant value when particular hypotheses are tested.

The major focus of RCT is typically related to treatments directed toward specific elements of disease such as lupus nephritis or central nervous system disease. However, other trials may be aimed at maintenance of remission or prevention of deterioration of certain systems. Therefore, each RCT will have a unique set of specific outcome variables selected on the basis of the primary hypothesis being evaluated. The same is true for LOS, which may study longterm outcomes of organ involvement, but also safety factors such as risk of or means of protection against certain aspects of the disease or comorbidity (e.g., osteoporosis or side effects of glucocorticoids). The specific outcome measures will not only be different among such studies, but they usually will also differ in some respects from the 4 general domains selected through the OMERACT SLE process.

A number of outcome domains were identified by the participants to be of particular interest for future research. To this end, data need to be collected and specific hypotheses validated. It was suggested that such potentially useful outcome variables should be included in the investigative efforts of clinical researchers and that they should be discussed and evaluated in future OMERACT conferences. Among such items were economic outcomes and instruments for their assessment, fatigue, and use of steroids, as well as serology, physical disability, and psychosocial components. Finally, an issue to be considered in future workshops is the question of measuring overall "response" to a certain therapeutic procedure by generating a "responder index" based upon outcome domains.

Some factors may be important for stratifying the trial population prior to randomization. For example, if a new drug were tested for its ability to reduce fatigue, it might be relevant to stratify according to the presence or absence of fibromyalgia. Other trials may need stratification according to the organ damaged or the degree of damage that has occurred.

None of the instruments in the domains selected for RCT have been tested extensively in clinical trials. Indeed, the available instruments are currently only valid in LOS and, with the possible exception of disease activity, have not been sufficiently field tested in clinical trials. This is even more true for the combination of the 4 domains making up the core set. On the other hand, particularly for LOS, several instruments are available for 3 of the domains. [Validated instruments to assess disease activity: the British Isles Lupus Assessment Group (BILAG), European Consensus Lupus Activity Measure (ECLAM), Lupus Activity Index (LAI), NIH SLE Index Score (SIS), SLE Activity Measure (SLAM), and SLE Disease Activity Index[2,6]; all correlate well with each other and can be used interchangeably.]

However, whereas SLEDAI, SLAM, SIS, ECLAM, and LAI are essentially global scores, most of which have been compared on "paper" and real patients and have been shown to be comparable, BILAG provides activity scores in 8 organs/systems but can be converted into a global score that also correlates well with other scores. There is one validated instrument for the assessment of damage (SLICC/ACR Damage Index). The Medical Outcome Survey Short Form (SF-36) has been commonly used for the assessment of health status/QOL[1–4]. Obviously, for some of the instruments reliability may be dependent on the experience of the observer. In any case, the responsiveness of most instruments needs further study.

To date there have been relatively few longitudinal observational studies and even fewer randomized controlled clinical trials in SLE. A great impediment has been the inability to measure the success or failure of interventions as well as subtle or even major differences in the evolution of the disease in a reliable manner. The OMERACT SLE core set of outcome domains will allow improved assessment of efficacy of new therapeutic agents or regimens with greater validity and comprehensiveness. Regulatory agencies such as the US Food and Drug Administration and the European Medicines Evaluation Agency, EAMA, will then have greater confidence in the proposals for new agents, and both physicians and patients will obtain new evidence on the efficacy of certain remedies in the short as well as in the long term. This will significantly improve our potential to manage this multisystem and potentially life threatening disease.

ACKNOWLEDGMENT

We thank Dr. Maarten Boers, Amsterdam, for discussion and his very helpful suggestions.

REFERENCES

1. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Outcome measures to be used in clinical trials in systemic lupus erythematosus. J Rheumatol 1998;26:490–7.
2. Bencivelli W, Vitali C, Isenberg DA, Smolen JS, Snaith M, Bombardieri W. Disease activity in systemic lupus erythematosus. Report of the Consensus Study Group of the European Workshop for Rheumatology Research. III. Development of a computerised clinical chart and its application to the comparison of different indices of disease activity. Clin Exp Rheumatol 1992;10:549–54.
3. Gladman DD, Ginzler E, Goldsmith C, et al. The development and initial validation of the SLICC/ACR Damage Index for SLE. Arthritis Rheum 1996;39:363–9.
4. Stoll T, Gordon C, Seifert B, et al. Consistency and validity of patient administered assessment of quality of life by MOS SF-36: first association with disease activity and damage in patients with systemic lupus erythematosus. J Rheumatol 1997;24:1608–14.
5. Wolfe F, Lassere M, van der Heijde D, et al. Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. J Rheumatol 1999;26:484–9.
6. Vitali C, Bencivelli W, Mosca M, Carrai P, Sereni M, Bombardieri S. Development of a clinical chart to compute different disease activity indices for systemic lupus erythematosus. J Rheumatol 1998;26:498-501.