Assessing patients with lupus: towards a drug responder index

D. Isenberg and R. Ramsey-Goldman

Centre for Rheumatology/Bloomsbury Rheumatology Unit, Middlesex Hospital, London and 1 Division of Arthritis and Connective Tissue Diseases, Northwestern University, Chicago, IL 60611-3008, USA

The remarkable diversity of clinical features that are evident in patients with systemic lupus erythematosus (SLE) demands particular vigilance on the part of the physician attempting to assess and treat this complex, but fascinating, disease. Although virtually any symptom one cares to think of may be due to active lupus (particularly in the central nervous system), the physician must also guard against the fallacy that every clinical problem experienced by a patient with lupus is due to the disease. In this review, we will put forward the view that in order to understand the totality of the effect of a disease like lupus upon a patient, measures are needed that distinguish disease activity (implying that the problems can be corrected), damage (meaning permanent change) and the patient’s own perception of their health status.

Practical assessment of patients with lupus in the clinic or hospital ward

A comprehensive approach in a chronic disease characterized by exacerbations and remissions is needed in all settings in which clinicians assess lupus patients. Therefore, patients seen in the out-patient clinic or hospital, and those participating in a clinical drug trial, need a careful evaluation that assesses the current status of the patient, but also looks to the future in order to prevent or minimize irreversible damage from the disease or its treatment.

What should be assessed?

The protean manifestations of lupus require a thorough search for signs and symptoms of disease. The clinical review includes probing for constitutional symptoms (e.g., fatigue, fever, weight loss of \( >5\% \) of body mass) and screening by history and examination for mucocutaneous, musculoskeletal, cardiopulmonary, gastrointestinal, lymphoreticular, neuropsychiatric and ocular manifestations.

Laboratory monitoring includes a full blood count and differential white cell count (lymphopenia is a common feature in lupus patients). Since renal disease may be silent, a urinalysis looking for red cells, white cells, protein and cellular casts should be performed on a regular basis. If abnormalities are detected, a 24 h urine collection should be obtained to assess the total...
daily protein loss. The glomerular filtration rate is most often assessed by measuring the creatine clearance. This is not entirely accurate as creatinine is secreted actively by the renal tubules. An alternative is $^{51}$Cr-labelled EDTA, the clearance of which can be assessed from the plasma radioactivity at one time point following an i.v. dose (rather than collecting timed urine samples). Autoantibody determinations including antinuclear antibody, anti-Ro/SSA, anti-La/SSB, anti-RNP, anti-Sm and antiphospholipid antibodies facilitate the diagnosis of lupus or its subsets. Only anti-double-stranded DNA antibodies are thought to be helpful by some, but not all, investigators [1]. Similarly, changes in complement components [notably C3, C4 or, if available, their breakdown products (C3d or C4d)] may assist with prognosis and monitoring of therapy in some patients [1].

Organ damage can be assessed on a yearly basis with the Systemic Lupus International Collaborative Clinics/American College of Rheumatology (SLICC/ACR) damage index, which is discussed in detail in a subsequent section. A strategy for monitoring drug toxicity in lupus patients was suggested in a recent review [1]. Briefly, the authors suggested using the practice guidelines for monitoring drug toxicity in rheumatoid arthritis developed by the American College of Rheumatology [3] as a guide for monitoring a patient with SLE. Clearly, one must adapt the approach for the individual patient.

Assessment of quality of life or the patient’s perception of disease is also important. In the last few years, various indices have been utilized to assess this important perspective. From the patient’s standpoint, these instruments are easy to use and require little time to complete. Valuable information on the functional aspect of a patient’s current condition is available to the clinician. The scales and domains measured in these instruments are described in detail in a subsequent section.

Formal assessment of disease activity in SLE

As Liang et al. [4] have pointed out, between the mid-1950s and mid-1980s some 60 attempts were made to develop a disease activity index for patients with SLE. None of these attempts were adequately validated or even shown to be reliable. The situation has changed radically in the last 15 yr, however, with the description of genuinely validated clinical activity measures. Ideally, such measures should include the following components:

- **Individual variables, ascertained in a generally acceptable way (case validity).**
- **An adequate number of variables chosen in an appropriate manner (content validity).**
- **Agreement with an external criterion considered to be a superior measure of disease activity standard (criterion validity).**
- **Positive correlation with other clinical scales or with laboratory markers of disease activity and the capacity to differentiate between patient groups who would be expected to have differing levels of disease activity (construct validity).**

Two main types of activity measures have been evolved. The global score systems are aimed at providing a simple overall measure of activity. In contrast, individual organ/system assessment scales seek to emphasize the diverse nature of lupus and thus to avoid a single score on the grounds that they are too reductionist.

The best known global disease activity measures are the SLAM (Systemic Lupus Activity Measure) [4], SLEDAI (Systemic Lupus Erythematosus Disease Activity Index) [5] and ECLAM (European Community Lupus Activity Measure) [6]. Derived from a very different philosophical standpoint, the BILAG (British Isles Lupus Assessment Group) index was developed initially in 1984 in an attempt to move away from the global score approach. It is based upon the ‘intention-to-treat’ premise [7]. Each index has been shown to have validity using the criteria set out in the previous paragraph.

The original SLEDAI index was developed from a meeting in Toronto in 1985, through a nominal group process, whereby a list of 24 variables was produced, thought to be important by the participants in describing disease activity in SLE. These variables were then used to evaluate disease activity in a cohort of patient profiles selected from the database of the Toronto lupus clinic. The variables (or descriptors of disease activity) were grouped into nine organ systems. Multiple regression models were then used to derive weights for the selected variables in predicting disease activity. The final scale adopted included 24 items using weights ranging from 1 to 8, with a total maximum possible score of 105. This index measures the disease activity in the 10 day period prior to the assessment, within which the manifestation must be recorded. It is a descriptive index and has been found to be reliable in naive observers [8], between countries [9] and in routine use [10]. It is sensitive to change over time [11] and has been used in prognosis studies [12]. The SLEDAI has been modified slightly for use in particular situations, e.g. the Mexican version of SLEDAI—known as the MEXSLEDAI [13]. In an ongoing study in the USA, known as the SELENA (Safety of Estrogens in Lupus Erythematosus, National Assessment) trial, another modification of SLEDAI is being used in which several of the descriptors have been changed. For example, vertigo has been added in ‘cranial nerve disorder’ and for both pleurisy and pericarditis classic and severe pain is sufficient to be recorded in the absence of objective data such as a pleural effusion (J. Buyon, personal communication).

The SLAM index, in contrast, is designed to allow some assessment of disease severity [4]. It was developed in Boston with input from members of the Lupus Council of the American College of Rheumatology. It includes 32 items divided into 11 organ systems and includes a scoring for severity as the variables are not only scored as present, but graded on a scale of 1–3 based on severity, giving a possible total score of 86. Although this index includes the concept of activity and
severity, some of the measures included, such as arthralgia and fatigue, may not represent true disease activity, but rather the patient’s perception of disease activity. This index has been found to be reliable [13] and has been used as an outcome measure in therapeutic trials. A modified version of SLAM, SLAM-R, omits scoring for pneumonia and truncates several scales.

The ECLAM index has been described following a study in which 29 centres, from 14 European countries, evaluated and chose 15 items from an analysis of 704 patients [6]. These items were selected at a consensus meeting of representatives from several of the participating centres. It differs from the other activity indices as it was directly derived from a study of a large number of real patients and the analysis of a large amount of data was collected in a standardized manner during the multicentre effort.

In essence, each of these global indices scores a varying number of points for involvement in particular organs or systems. By simply adding up the points scored, a global total is arrived at. These indices have shown sensitivity to change [10] and in a number of studies (reviewed elsewhere [2]) have been shown to compare well with each other.

The BILAG index was derived following detailed discussions among a group of rheumatologists who achieved a consensus about when to treat lupus patients with disease-modifying therapy such as high doses of corticosteroids or immunosuppressives [7]. The index underwent relatively minor changes, but in its current format has shown a high degree of between-rater variability and validity [14].

The BILAG index now includes a total of 86 items in eight organs or systems, each item is scored as present or absent within the previous month, with many of the items being identified as new, improved, the same, or worse. For an item to be recorded in one of these categories, the assumption is made that the problem is due to lupus. Thus, a lupus patient with coincident asthma would not have shortness of breath recorded if the clinician felt that it was due to the coincident disease. As with the global score indices, laboratory tests make up relatively little of the final score, apart from the haematological system, and to a lesser extent the renal system, for obvious reasons.

Each organ or system is given a score of A–E, where A denotes disease thought to be sufficiently active to require disease-modifying treatment, prednisolone $< 20$ mg/day or immunosuppressants, B refers to problems requiring only symptomatic therapy such as antimalarials or non-steroidal anti-inflammatory drugs or prednisolone $< 20$ mg/day, C indicates stable, mild disease, D indicates a previous affected but currently inactive system and E indicates that the system/organ has never been involved previously. However, it must be emphasized that the individual scores are determined on the basis of the presence/absence of the items referred to above.

BILAG has been shown to function well for each of these organs or systems, with the possible exception of the central nervous system. There has been particular difficulty amongst physicians in deciding whether neurological features were due to lupus or simply coincidental findings, and also neurological disease may be treated with specific therapy such as anticonvulsants in those with epileptic attacks or anticoagulants following strokes, rather than by corticosteroids and immunosuppressants.

Although it was not designed for this purpose, the BILAG index can be converted into a global score system, thus $A = 9$, $B = 3$, $C = 1$, and $D$ and $E = 0$, resulting in a potential range of $0–72$. In a comparative study of 75 patients with SLE, a global score derived as above from BILAG was shown to correlate very highly with the ‘genuine’ global score indices, SLAM, SLEDAI and ECLAM [15].

Assessment of damage in SLE

Fifty years ago, when lupus had a 5 yr survival estimated at $< 50\%$ [16], mortality was an important, and largely sufficient, measure of outcome in patients with lupus. With the substantially improved survival figures for patients with lupus, it has become necessary to develop a more subtle method of assessing the overall cumulative effect of the disease. The SLICC/ACR index was published in 1996 after 5 yr of preparatory work [17]. The index does not seek to attribute the cause of damage to the disease or its treatment. For example, it merely records the number of items of permanent change that have affected an individual patient since the onset of the disease. The damage score can thus only remain the same or increase with time. In order for a feature to be regarded as due to damage, the change must have persisted for at least 6 months. It is ascertained by clinical assessment or simple investigations such as urinalysis and plain radiographs, which are widely available. Damage is distinguished in 12 organs or systems, including ocular, neuropsychiatric, renal, pulmonary, cardiovascular, peripheral vascular, gastrointestinal, musculoskeletal, skin, gonadal and endocrine damage, and the occurrence of malignancies. Weighting was not shown to improve the ability of the index to record damage, but the index has been shown to discriminate between changes in disease damage in patients with both active and inactive disease, and to have good interobserver reliability.

In a 10 yr retrospective study of 80 patients with SLE using the damage index, it was shown that renal damage at 1 yr was predictive of end-stage renal failure [18]. Pulmonary damage at 1 yr was predictive of death by 10 yr of follow-up. The study also demonstrated that Afro-Caribbean and Asian patients had significantly higher mean renal and total damage scores at 10 yr than the Caucasian population. A high neuropsychiatric damage score was also evident in the Asian group.

Other studies [19, 20] have confirmed the usefulness and validity of this index. For example, Nossent [20] undertook a case note review of 90 Afro-Caribbean patients with lupus who had been followed for a mean period of 6 yr with ‘periodic assessment’ of the SLICC/ACR damage index. The mean damage score 6 months after diagnosis was 0.6 and at last assessment 2.4. The damage score correlated with a weighted aver-
age of SLEDAI disease activity scores and the number of disease exacerbations, but not with age or steroid dose.

In contrast, Alarcon et al. [21], in a cross-sectional study of Hispanic, Afro-American and Caucasian patients (70–80 patients in each group) whose disease duration did not exceed 5 yr, did not find a statistically significant difference in the damage score between these three groups.

The SLICC/ACR damage index has won widespread acceptance in the lupus research ‘community’. Thus, Bootsma et al.’s [22] comment that the index appeared useful in SLE ‘because it has the capacity to measure change over time; it offers the opportunity to compare treatment arms in clinical trials and reflects cumulative damage’. It should be emphasized that long-term studies are needed. In short-term studies (up to 3 yr), it is likely that only 50% of patients with lupus will register any damage.

Assessing health status in SLE

An important principle to remember is that the patients’ perceptions of their lupus, in particular its most troublesome features, are frequently different from the clinician’s.

Although no lupus-specific measure has been designed to date, two measures have been used in a wide variety of studies. These measures were developed from the questionnaires devised initially by Ware and his colleagues at the Rand Corporation [23]. Over 200 items were initially selected for study, before the Medical Outcome Survey Short Form-20 [24] (literally 20 questions) and subsequently the Short Form-36 were developed [25, 26]. The Short Form 20, however, lacks a particular question about fatigue, which is most important in many patients with lupus. Thus, a number of studies have been reported using the so-called SF-20+ (i.e. the SF-20 questionnaire plus an additional question concerning fatigue) [27, 28].

Although the SF-20 appears to be a valid measure of quality of life, the instrument used most often and internationally to assess health status is the SF-36. This index may be more appropriate as it records the continuum from healthy to severely ill very well [29, 30]. The SF-36 is almost as quick and easy to complete as the SF-20. It has 36 questions concerning eight static domains predominantly occurring over the previous month, namely, physical function, role limitations: physical problems, role limitations: emotional problems, social function, mental health, general health perception, vitality and pain. The SF-36 is broader than the SF-20, asking questions about vitality and general health (over the previous year) not found in the SF-20. Other domains have a broader and deeper perspective, as, for example, physical function includes specific questions about lifting groceries, climbing one or several flights of stairs and walking different longer distances. It is scored in a similar way to the SF-20, with a score of 0–10–0; higher values indicating better health.

Both measures have been compared in a study of 150 lupus patients [31], with significant associations between the corresponding domains in each measure and with global disease activity measured by BILAG, although this association is not strong, suggesting that quality of life and disease activity measured by BILAG are distinct. In the SF-36 index, different disease activity levels were significantly associated with different quality of life scores and showed an excellent ability to record the continuum from good health to serious illness. Disease activity had a greater effect on quality of life than age, cumulative damage or disease duration. A previous study had shown that in the SF-20+, limitations in physical functioning were associated with damage to the musculoskeletal system, measured by the SLICC/ACR damage index, and renal disease was inversely associated with fatigue, but there were no other significant associations between this health questionnaire and the damage index. In view of the comprehensiveness of the SF-36, its widespread use and international validation for a wide range of medical conditions, this is currently the optimal choice as the patient life-impact assessment, although the degree of change of an SF-36 score that is clinically important is not yet known.

Future developments

We stand on the brink of a revolution in the way in which we treat patients with SLE. At the time of writing this article, as well as the recent introduction of myco-phenolic acid to the immunosuppressive armamentarium, we are aware of a major clinical study in the USA assessing the effects of a B-cell collagen LJP394, a safety study of oestrogens given to post-menopausal women, and two trials of monoclonal anti-CD40 ligand in the treatment of SLE. In addition, a double-blind placebo-controlled trial of the adrenal steroid hormone dehydroepiandrosterone (DHEA) has just been published [32]. It appears to have a protective effect with respect to steroid-induced osteoporosis, but was of little benefit in patients with severe disease. There is thus an urgent need to agree a format for assessing the response of a patient with lupus to these exciting new forms of therapy. Intense discussions have been going on during the past 2 yr among members of the SLICC and OMERACT groups, and more recently the Food and Drug Administration in the USA. We are hopeful that consensus will be reached and consider it likely that agreement requires recognition of the importance of recording a disease activity measure, a damage measure, a patient health perception index, and methods of recording toxicity and economic costs. Only by assessing each of these factors will we be able to determine whether new drugs provide overall benefit for a patient. It seems logical to us that clinical researchers around the world, who are interested in lupus, should compare the outcome of their studies in the same way. The optimal way to do this is to ensure that everyone uses the same drug responder index.
Acknowledgements

DAI and RG-G are grateful for the support of their colleagues in the SLICC group. RG-G is supported by grants from the National Institute of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases (AR-30692), Lupus Foundation of America Illinois Chapter, the Arthritis Foundation Illinois Chapter.

References