

BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus

D. A. Isenberg¹, A. Rahman¹, E. Allen¹, V. Farewell², M. Akil³,
I. N. Bruce⁴, D. D'Cruz⁵, B. Griffiths⁶, M. Khamashta⁵, P. Maddison⁷,
N. McHugh⁸, M. Snaith⁹, L. S. Teh¹⁰, C. S. Yee¹¹, A. Zoma¹²
and C. Gordon¹¹

Objective. To devise a more discriminating version of the British Isles Lupus Assessment Group (BILAG) disease activity index and to show that it is reliable.

Methods. A nominal consensus approach was undertaken by members of BILAG to update and improve the BILAG lupus disease activity index. The index has been revised following intense consultations over a 1-yr period. It has been assessed in two real-patient exercises. These involved patients with diverse clinical features of SLE, including gastrointestinal, hepatic and ophthalmic problems, which the earlier versions of the index did not fully take into account. Reliability in terms of the ability to differentiate patients was assessed by calculating intraclass correlation coefficients. The level of agreement between physicians was determined by calculating the ratio of estimates of the standard error (SE) attributable to the physicians to the SE attributable to the patients.

Results. Good reliability and high levels of physician agreement were observed in one or both exercises in the constitutional, mucocutaneous, neurological, cardiorespiratory, renal, ophthalmic and haematological systems. In contrast, the musculoskeletal system did not score as well, although providing more clear-cut glossary definitions should greatly improve the situation.

Conclusions. Some significant changes in the BILAG disease activity index to assess patients with SLE are proposed. The process of demonstrating validity and reliability has started with these two exercises assessing real patients. Further validation studies are under way. BILAG 2004 is likely to be valuable in clinical trials assessing new therapies for the treatment of SLE, as it provides a more comprehensive system-based disease activity measure than has been available previously.

KEY WORDS: Systemic lupus erythematosus, Disease activity, Assessment, Index, BILAG.

The British Isles Lupus Assessment Group (BILAG) has been meeting regularly since 1984. The group devised a disease activity index for patients with systemic lupus erythematosus (SLE) that was based upon the principle of the physicians' intention to treat [1]. Thus it was developed and subsequently validated by making agreed assumptions about the likely treatment that will be given to patients with particular groups of clinical features in eight organs or systems. The advantages of this approach are that it provides a testable hypothesis, offered a more discerning view of disease activity in patients with SLE compared with the usual global activity score and might be particularly useful in therapeutic trials. The initial report, published 16 yr ago, described the development of the original index. Subsequently [2], some minor modifications were tested and its reliability and validity as an instrument for the accurate measurement of clinical disease activity were

demonstrated in each of the organs or systems, with the possible exception of the central nervous system.

In the ensuing years members of BILAG and some others have used the index in various ways, including attempts to correlate disease activity with serological abnormalities [3, 4], determining the occurrence and rate of flare in patients with SLE [5, 6] and in clinical trial settings [7].

Further modifications to the index have been made especially with regard to the renal system. The group commissioned a state-of-the-art piece of computer software that incorporates a large amount of demographic information, the clinical information to determine a BILAG activity index and two global [Systemic Lupus Activity Measure (SLAM) and Safety of Estrogens in Systemic National Assessment-Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI)] disease activity

¹University College London, London, ²University of Cambridge, Cambridge, ³Sheffield Centre for Rheumatic Disease, Sheffield, ⁴University of Manchester, Manchester, ⁵St Thomas' Hospital, London, ⁶The Freeman Hospital, Newcastle, ⁷University of Wales, Bangor, ⁸Royal National Hospital for Rheumatic Diseases, Bath, ⁹Derbyshire Royal Infirmary, Derbyshire, ¹⁰Blackburn Royal Infirmary, Blackburn, ¹¹University of Birmingham, Birmingham, ¹²Hairmyres Hospital, East Kilbride, Scotland, UK.

Submitted 8 September 2004; revised version accepted 25 February 2005.

Correspondence to: D. Isenberg, Centre for Rheumatology, University College London, 4th Floor, Arthur Stanley House, 40–50 Tottenham Street, London W1T 4NJ, UK. E-mail: d.isenberg@ucl.ac.uk

indices; in addition the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index and the patient self-assessment SF-36 index can be recorded. In addition, the software allows the recording of patient medication, DNA binding, C3 and other serological tests, which do not form part of the BILAG index, and a very useful graphing capacity. The system, known as BLIPS (British Lupus Integrated Prospective System), and the more recent minor modifications of the BILAG index are described elsewhere [8].

During the past 5 yr, however, the group has become increasingly concerned that some aspects of the division of organs and systems for the purposes of activity assessment are unsatisfactory. In particular, some clinical features relating to the abdomen tend to be rather scattered in the present BILAG index and few ophthalmological problems are taken into account. The terms and definitions used for neuropsychiatric manifestations have become out of date in the light of the American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes [9]. Some patients with active lupus are treated with high-dose anticoagulation, as thrombosis as well as inflammation may underlie certain clinical features [6].

In addition, one of the unique features of the BILAG index is that it is a transitional index, as items that are improving are scored less severely than those that are new, worse or the same. Similarly, changes in renal manifestations, such as proteinuria and creatinine, determine the renal score. It should be stressed that features should only be recorded if the physician attributes the feature to active lupus and not to some other process. In the new index, features that contribute to an A score when recorded as being the same, worse or new will contribute to a B score when improving, as these features are still significant. At present, all items that are improving can only contribute to a C score, which does not reflect the appropriate level of disease activity for the more severe manifestations.

Finally, some of the items in the present BILAG system, such as avascular necrosis, are, in reality, damage items and therefore should not be in an activity index.

Thus it is now timely to optimize the BILAG index, particularly as an era of new therapies for patients with lupus is dawning. As clinical trials of B-cell depletion, anti-Bly and CTLA4-Ig (amongst others) are being planned, we believe that an index that offers an immediate across-the-board view of activity in individual systems in patients with lupus has much more to offer than validated global score indices.

For these reasons, members of BILAG have engaged in intense discussions over the past year to refine the BILAG index and have undertaken two real-patient exercises. Each of these exercises has involved eight patients with SLE and eight physicians, in an attempt to provide initial validation and reliability testing for the BILAG 2004 index of disease activity in patients with lupus.

Methods

The revised BILAG index (BILAG 2004) has been developed from the original index, based, as described above, on the principle of a physician's intention to treat using a nominal consensus approach. In this revised index, the original section on vasculitis has been removed and the nine systems (not organs) considered are: constitutional, mucocutaneous, central nervous system, musculoskeletal, cardiovascular/respiratory, abdominal, renal, ophthalmic and haematological.

Each of the items included in the index has been carefully considered by members of the group. The most active score in each organ or system, grade A, is defined as the individual clinical features, or combinations of features, which the group believes would lead to the prescriptions of medium/large doses of corticosteroids (>20 mg prednisolone or equivalent) and/or starting or increasing immunosuppressive drugs or high-dose

anticoagulation [International Normalized Ratios (INR) >3]. Grade B is given to those patients with known disease activity requiring somewhat lower doses of immunosuppressives (e.g. <20 mg prednisolone) and/or specific drugs, such as antimalarial, anti-epileptic, antidepressant and non-steroidal anti-inflammatory drugs (NSAIDs) or topical steroids. The C grade in each system defines patients with mild persistent activity only requiring symptomatic therapy (e.g. analgesics or NSAIDs). D grade implies the organ or system was once active but is no longer so and grade E indicates that the organ or system has never been active. As with the original BILAG index, BILAG 2004 provides a testable hypothesis. Studies have been established to collect data to determine whether patients meeting the clinical criteria for grades A, B and C really do receive the treatment envisaged and to compare the index to other measures of disease activity (construct validity) and to demonstrate that the index is sensitive to change. The purpose of the two real-patient exercises described in this report was to determine the level of agreement between the eight physicians using BILAG 2004 in the assessment of eight SLE patients and to demonstrate the ability of the new index to capture the level of activity in the various organs or systems in a clinically meaningful way.

The patients involved in the assessments gave written, informed consent. The study was approved by the University College London ethics committee.

Real-patient exercises

In both real-patient exercises the order of assessment was randomized according to an 8 × 8 Latin square design. For this design the appropriate statistical model is additive and any interaction between order, physicians or patients is identified with error. During each of the patient exercises the assessors were asked to complete the BILAG 2004 form. They were provided with a one-page synopsis of the patient's history, current haematological results, serum creatinine and, where relevant, urine protein-creatinine ratios. The patients were chosen to reflect a mixture of clinical features due to a range of activity and damage items. Seven out of eight assessors were the same in both exercises. The assessors were all members of BILAG.

The first exercise took place in May 2003, when eight adult patients with SLE (seven female, one male) were assessed by eight rheumatologists. Each consultation took up to 50 min. Following a detailed assessment of the results from this exercise, a number of minor changes in the revised index and the glossary were agreed. During the second real-patient exercise in March 2004, eight adult patients (seven female, one male) were assessed by eight rheumatologists. Two of the patients participated in both exercises. The consultations again lasted up to 50 min.

Statistical considerations

The BILAG index can be converted into a numerical score (A grade = 9 points, B = 3, C = 1, D = 0, E = 0) [5] and is treated as continuous for the purpose of this analysis. In line with the approach taken in a similar exercise for the assessment of myositis outcomes [10], we have chosen to use two summary measures of agreement: an intraclass correlation coefficient (ICC) and the ratio of the estimates of the standard deviation attributable to the physicians to the standard deviation (SD) attributable to the patients themselves ($\sigma_{\text{phys}}/\sigma_{\text{pat}}$). The numerical values must be interpreted with some caution but should provide qualitative guidance for the comparison of the behaviour of the different tools.

A three-way model appropriate for the Latin square design was used, and following the approach of Shrout and Fleiss [11], an appropriate ICC with 95% confidence interval, was defined based on physician, patient and error variation. The ICC

is equivalent to ICC [2, 1] as given by Shrout and Fleiss [11]. Although order was adjusted for in the analysis, it can be considered to be an artefact of the design and has not been incorporated into the ICC. A 95% confidence interval was defined for $\sigma_{\text{phys}}/\sigma_{\text{pat}}$.

Analysis of variance was used to estimate the variance components under the assumption that patients and physicians were randomly chosen from larger populations. This assumption allows the results to be generalized beyond the physicians who took part in the real-patient exercises.

For each system, both summary measures are presented together, as it was felt that to assess the performance of each system in the index it is necessary to consider both its reliability and the level of physician agreement. Reliability, as measured here by the ICC, refers to the ability of the index to differentiate between patients. In contrast, the level of physician agreement, assessed here by the examination of the standard deviation (SD) of measurement, refers to the level of agreement between the physicians. It is possible for a system to have a high ICC, indicating that it differentiates well between patients, together with a high SD of measurement attributable to the physicians, indicating poor agreement. It is also possible, given a homogeneous population, for a system to have a low ICC, indicating poor ability in differentiating between patients, but a low standard deviation attributable to the physicians, indicating good agreement.

Both measures have been used to classify the results from both real-patient exercises into three categories. For the purpose of this classification we have considered an ICC > 0.60 (not 0.65) as high, indicating that the index differentiates well between the patients. We have considered agreement among physicians to be high if $\sigma_{\text{phys}}/\sigma_{\text{pat}} < 0.40$. (These boundaries are somewhat arbitrary but facilitate classification of the results.)

The first category consists of those systems where both the ICC is high and $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ is low, indicating that the index is differentiating well between patients with a high level of physician agreement. These results have been categorized as Good.

The second category consists of those systems that demonstrate a good performance in only one of the two measures. These results have been categorized as Good*, as in a previous reliability study in myositis [10]. Among these, when $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ is low, indicating a high level of agreement among the physicians, it appears that the low ICC is generally due to little or no variation among the patients, and these systems can be considered to be performing reasonably well. However, in this category, when the ICC is moderate or high, indicating an ability to differentiate between the patients, the high value of $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ indicates that there was some variability among the physicians.

The third group consists of those systems in which both the ICC is low and $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ is high, indicating that the index is not differentiating well between patients and there is a poor level of physician agreement. These systems have been classified as Poor.

Results

Results of the two exercises are shown in Table 1. The analysis of the results from both exercises suggests that the BILAG 2004 performs well.

In the first exercise the assessment of disease activity in the mucocutaneous, nervous system, renal, ophthalmic and haematological organs and systems exhibited good reliability (ICC > 0.60) and a high level of physician agreement ($\sigma_{\text{phys}}/\sigma_{\text{pat}} < 0.40$). With the renal and haematological systems it was important to show that the physicians interpreted the laboratory data appropriately in terms of attribution to lupus disease. In addition, the BILAG 2004 led to a high level of agreement between the physicians for the constitutional and cardiovascular/respiratory and abdominal

systems. However, for these systems the ability to distinguish patients was reduced (ICC 0.40–0.60).

In the second exercise, the assessment of disease activity in the constitutional, renal, cardiorespiratory and haematological organs/systems exhibited good reliability and a high level of agreement. However, despite a high level of physician agreement, reliability in terms of distinguishing patients could be improved in the mucocutaneous and nervous system (ICC < 0.60).

In the first exercise only one patient had abdominal disease and in the second exercise only one patient had abdominal disease and only one patient had ophthalmic disease. Consequently (despite the classifications in Table 1), it is difficult to interpret the performance of BILAG 2004 in these organs/systems.

In both exercises only the assessment of the musculoskeletal system exhibited both poor reliability and a low level of physician agreement.

In both exercises the total BILAG score was calculated. In the first exercise the total BILAG performed poorly, demonstrating both poor reliability and a low level of physician agreement. In the second exercise the total BILAG demonstrated better reliability in discriminating patients and a reasonably high level of agreement among the physicians.

But it should be noted that the index was not designed to be used as a global score and the values used for the numerical scoring have not been validated yet against the gold standard of treatment prescribed by the physician.

Discussion

The two real-patient exercises described in this study are the first steps along the road that, we believe, will lead to a more discriminating version of the BILAG disease activity index than the current version, which is based on the one that was originally devised over 15 yr ago [1]. Over time it had become clear that several items, whilst contributing little to the score (e.g. the presence of avascular necrosis and tendon contracture) were in reality damage items and should not have been included. There was, however, an increasing concern that, although rare, it should be possible to record disease activity affecting the eyes using the BILAG index. Doubts had also been expressed about the tendency of the original index to scatter items relating to the abdomen to different parts of the index, whereas vasculitis, which had originally been considered an individual system, should more properly be distributed to the variety of systems that it can truly affect. Over the years it also became apparent that there is a need for a detailed glossary and greater use of imaging and other investigations to support clinical impressions, particularly for the use of the BILAG index in multicentre clinical trials.

What is not changed in the BILAG 2004 index is the essential principle on which the index was established, namely the physician's intention to treat. This principle provides the establishment of a testable hypothesis and was used to establish the index initially [1, 2]. We envisage that the same rigorous proof of principle will be undertaken with the major revision of the index that we are now describing. Studies confirming this and other measures of construct validity are already under way. The new concept included in this index is that an A score can reflect the need to use high-dose anticoagulation, not just potent immunosuppression, for complex severe manifestations of lupus, in which the predominant pathological mechanisms are often uncertain.

Equally important, we believe, is the provision of an index that is not, primarily, intended to provide a global score. Whilst global scores have their place in a disease as complex and subtle as lupus, it is of paramount importance to establish an index that offers an at-a-glance review of the disease activity across the whole spectrum of systems that can be affected. It is also necessary to use a transitional index that is sensitive to change. This notion seems particularly apposite, given that several new therapies

TABLE 1. Reliability in terms of the ability to differentiate patients, assessed by calculating the ICC

| Organs/systems | Good | | Good* | | Poor | |
|------------------------------|----------------------|--|----------------------|--|----------------------|--|
| | ICC | $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ | ICC | $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ | ICC | $\sigma_{\text{phys}}/\sigma_{\text{pat}}$ |
| First real-patient exercise | | | | | | |
| Constitutional | | | 0.567 (0.307, 0.85) | 0.340 (0, 0.901) | | |
| Mucocutaneous | 0.705 (0.462, 0.914) | 0.238 (0, 0.642) | | | | |
| Central nervous system | 0.778 (0.564, 0.939) | 0.187 (0, 0.513) | | | | |
| Musculoskeletal | | | | | 0.173 (0.013, 0.558) | 0.697 (0, 2.051) |
| Cardiovascular/respiratory | | | 0.401 (0.160, 0.765) | 0.284 (0, 0.969) | | |
| Abdominal | | | 0.430 (0.179, 0.784) | 0 (0, 0.740) | | |
| Renal | 0.984 (0.960, 0.996) | 0 (0, 0.082) | | | | |
| Ophthalmic | 0.794 (0.589, 0.944) | 0.054 (0.343) | | | | |
| Haematological | 1.00 | 0 | | | | |
| Total BILAG | | | | | 0.485 (0.230, 0.813) | 0.568 (0.102, 1.375) |
| Second real-patient exercise | | | | | | |
| Constitutional | 0.833 (0.653, 0.956) | 0.117 (0, 0.371) | | | | |
| Mucocutaneous | | | 0.364 (0.131, 0.740) | 0.258 (0, 0.99) | | |
| Central nervous system | | | 0.439 (0.191, 0.788) | 0.339 (0, 1.003) | | |
| Musculoskeletal | | | | | 0.112 (0, 0.473) | 1.009 (0, 2.924) |
| Cardiovascular/respiratory | 0.621 (0.363, 0.881) | 0.233 (0, 0.689) | | | | |
| Abdominal | 0.613 (0.353, 0.878) | 0.154 (0, 0.594) | | | | |
| Renal | 1.00 | 0 | | | | |
| Ophthalmic | | | 0.285 (0.073, 0.679) | 0 (0, 1.109) | | |
| Haematological | 1.00 | 0 | | | | |
| Total BILAG | | | 0.509 (0.252, 0.827) | 0.351 (0, 0.961) | | |

The level of agreement between physicians was determined by calculating the ratio of estimates of the standard deviation (SD) attributable to the physicians to the SD attributable to the patients: $\sigma_{\text{phys}}/\sigma_{\text{pat}}$, with 95% confidence intervals from both real-patient exercises.

are coming to clinical trial, and it is entirely feasible that individual drugs will help improve disease activity in some, but not all, systems. A global score may not capture such partial improvement so easily, whereas the BILAG 2004 index will enable the detection of improvement (or deterioration) very accurately in individual systems.

The principle purpose of running these two real-patient assessments was to start the process of confirming the reliability and validity of what we are now proposing. As the members of BILAG have changed, especially over the last 5 yr, we thought that these exercises would provide an important forum for both discussing activity in patients with lupus in general and for harmonizing definitions of clinical items in particular.

We were unable in these relatively limited exercises to explore the full range of activity in the ophthalmic and abdominal systems but it has afforded us an opportunity to make a practical start for both these and the other, more traditional organ/system assessments. Training sessions emphasizing the terms to be used and the glossary definitions preceded both real-patient exercises. We feel strongly that the use of the BILAG 2004 index (or indeed the original BILAG index) in a multicentre clinical trial or longitudinal outcome study really does require the use of such sessions. This ensures that the participating physicians are fully cognisant with the glossary and have some understanding of the way that the index is constructed.

As indicated previously, numerous changes were made before the first patient exercise to the items in each system on the BILAG form to the scoring and the glossary. Relatively minor changes in the wording of the glossary were made between the first and second patient exercises. After the second patient exercise, a detailed discussion of the results resulted in a number of further adjustments to the glossary. The final version of what we are currently proposing is shown in Appendix 1.

The use of the Good/Good*/Poor distinction was developed in a similar exercise, described in the development of international consensus measures for patients with 'inflammatory myositis' [10]. The boundaries used to ease classification are a convenience designation. Readers should look at the confidence intervals

around the summary measures to make their own judgement on the performance of the tools.

In the present two exercises, assessment of disease activity in the constitutional, mucocutaneous, nervous system, cardiorespiratory, renal, ophthalmic and haematological organs/systems were deemed to be Good/Good* in one or both assessments. In contrast, we had concerns about the musculoskeletal system assessments.

The discrepancies between physicians in the musculoskeletal system related to interpretation of the glossary for the degree of arthritis present. Physicians varied in their scoring depending on the duration of symptoms and whether or not synovitis had been observed, rather than the extent of the arthritis in terms of the number of joints affected. In lupus, transient inflammatory arthritis lasting a day or two is not unusual and can be quite severe, even disabling. To resolve this issue we have now defined 'severe polyarthritis' as 'observed active synovitis in at least 2 joints with significant impairment of activities of daily living and which has been present on several days (cumulatively) over the last 4 weeks'. In contrast, 'arthritis' or 'tendonitis' is defined as 'active synovitis in 1 or more joints' or tendons with some impairment of function, which has been present on at least several days over the last 4 weeks. This does not have to be observed at the assessment. All other forms of inflammatory joint pain are considered to be 'arthralgia', which is defined as 'inflammatory joint pain that does not fulfil the above criteria for arthritis'.

We are well aware that undertaking these two real-patient exercises has allowed only a relatively small number of patients to be studied using the new index. However, we consider that these exercises, together with the considerable amount of discussion that preceded and has followed them, do now provide the basis of a very useful tool to assess patients with active lupus. A research fellow supported by the Arthritis Research Campaign has now been appointed to undertake and coordinate further validity and reliability assessments using larger numbers of patients and physicians, to compare BILAG 2004 with other measures of disease activity, and to demonstrate that the index is sensitive to change.

In conclusion, we propose that BILAG 2004 provides a timely update of this unique and comprehensive disease activity index to assess patients with lupus. It is likely to be of considerable value in the assessment of disease activity in patients with lupus participating in trials of new therapies. Although the scoring can be done manually, a computer program is being designed to facilitate the scoring process.

| | |
|---------------------|--|
| <i>Rheumatology</i> | Key messages |
| | <ul style="list-style-type: none"> • The BILAG 2004 index provides a comprehensive activity index for the assessment of lupus patients. • This initial validation study shows that it is reliable when used by different physicians and can discriminate between patients. |

Supplementary data



Supplementary data are available at *Rheumatology* Online.

D. D'C. has received honoraria and is a consultant to Aspera Pharmaceuticals. The other authors have declared no conflicts of interest.

References

1. Symmons DPM, Coppock JS, Bacon PA *et al.* Development of a computerised index of clinical disease activity in systemic lupus erythematosus. *Q J Med* 1988;69:927–32.
2. Hay EM, Bacon PA, Gordon C *et al.* The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. *Q J Med* 1993;86:447–58.
3. Isenberg DA, Garton M, Reichlin MW, Reichlin M. Long term follow up of autoantibody profiles in black female lupus patients and clinical comparison with Caucasian and Asian patients. *Br J Rheumatol* 1997;36:229–33.
4. Ravirajan CT, Rowse L, MacGowan JR, Isenberg DA. An analysis of clinical disease activity and nephritis-associated serum antibody profiles in patients with systemic lupus erythematosus: a cross sectional study. *Rheumatology* 2001;40:1405–12.
5. Ehrenstein MR, Conroy SE, Heath J, Latchman DS, Isenberg DA. The occurrence, nature and distribution of flares in a cohort of patients with systemic lupus erythematosus: a rheumatological view. *Br J Rheumatol* 1995;34:257–60.
6. Gordon C, Sutcliffe N, Skan J, Stoll T, Isenberg DA. Definition and treatment of lupus flares measured by the BILAG index. *Rheumatology* 2003;42:1372–9.
7. Griffiths B, Emery P, Isenberg DA *et al.* A multicentre randomised controlled trial of cyclosporin A (CYA) versus azathioprine (AZA) in patients with severe SLE: an interim analysis. *Rheumatology* 2004;43(Sii):106.
8. Isenberg DA, Gordon C. From BILAG to BLIPS – disease activity assessment in lupus, past, present and future. *Lupus* 2000; 9:651–4.
9. Liang MH, Corzillius M, Bae SC *et al.* The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. *Arthritis Rheum* 1999;42:599–608.
10. Isenberg DA, Allen E, Farewell V *et al.* International consensus outcome measures for patients with idiopathic inflammatory myopathies. Development and initial validation of myositis activity and damage indices in patients with adult onset disease. *Rheumatology* 2004;43:49–54.
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–48.