Easy-BILAG
Carter, Lucy M; Gordon, Caroline; Yee, Chee-Seng; Bruce, Ian N; Isenberg, David; Skeoch, Sarah; Vital, Edward M

DOI: 10.1093/rheumatology/keab883
License: Creative Commons: Attribution-NonCommercial (CC BY-NC)

Citation for published version (Harvard):

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
Easy-BILAG: a new tool for simplified recording of SLE disease activity using BILAG-2004 index

Lucy M. Carter, Caroline Gordon, Chee-Seng Yee, Ian Bruce, David Isenberg, Sarah Skeoch and Edward M. Vital

Abstract

Objective. BILAG-2004 index is a comprehensive disease activity instrument for SLE but administrative burden and potential frequency of errors limits its use in routine practice. We aimed to develop a tool for more accurate, time-efficient scoring of BILAG-2004 index with full fidelity to the existing instrument.

Methods. Frequency of BILAG-2004 items was collated from a BILAG-biologics registry (BILAG-BR) dataset. Easy-BILAG prototypes were developed to address known issues affecting speed and accuracy. After expert verification, accuracy and usability of the finalized Easy-BILAG was validated against standard format BILAG-2004 in a workbook exercise of 10 case vignettes. Thirty-three professionals ranging in expertise from 14 UK centres completed the validation exercise.

Results. Easy-BILAG incorporates all items present in ≥5% BILAG-BR records, plus full constitutional and renal domains into a rapid single page assessment. An embedded glossary and colour-coding assists domain scoring. A second page captures rarer manifestations when needed. In the validation exercise, Easy-BILAG yielded higher median scoring accuracy (96.7%) than standard BILAG-2004 documentation (87.8%, *P* = 0.001), with better inter-rater agreement. Easy-BILAG was completed faster (59.5 min) than the standard format (80.0 min, *P* = 0.04) for 10 cases. An advantage in accuracy was observed with Easy-BILAG use among general hospital rheumatologists (91.3 vs 75.0, *P* = 0.02), leading to equivalent accuracy as tertiary centre rheumatologists. Clinicians rated Easy-BILAG as intuitive, convenient, and well adapted for routine practice.

Conclusion. Easy-BILAG facilitates more rapid and accurate scoring of BILAG-2004 across all clinical settings, which could improve patient care and biologics prescribing. Easy-BILAG should be adopted wherever BILAG-2004 assessment is required.

Key words: systemic lupus erythematosus, BILAG, disease activity, assessment, index

Rheumatology key messages

- Easy-BILAG is a high-accuracy, time-efficient tool for recording BILAG-2004 disease activity in SLE.
- It is the new recommended format for scoring BILAG-2004 index in clinical practice.
- Easy-BILAG and its training material is available free of charge for use in routine care at https://licensing.leeds.ac.uk/products/healthcare-questionnaires.

Introduction

Disease activity measurements in SLE are necessary for optimal patient care. They are central to clinical guidelines [1, 2] and treat-to-target approaches, which have been shown to improve outcomes in SLE [3, 4], and rely on specifically defining and measuring low disease activity and remission [5]. Furthermore, national commissioning policies for biologic agents also increasingly stipulate measured baseline and response disease activity criteria [2, 6].

National Hospital for Rheumatic Diseases, Royal United Hospitals Bath NHS Trust, Bath, UK
Submitted 30 July 2021; accepted 16 November 2021
Correspondence to: Edward Vital, Leeds Institute for Rheumatic and Musculoskeletal Medicine, Chapel Allerton Hospital, Leeds LS7 4SA, UK. E-mail: e.m.vital@leeds.ac.uk

© The Author(s) 2022. Published by Oxford University Press on behalf of the British Society for Rheumatology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Recording complex multisystem manifestations longitudinally is a significant challenge in SLE. As a result, formalized disease measures have been less readily embedded in routine care [7] compared with other rheumatic diseases such as rheumatoid arthritis. Composite disease activity instruments, including the SLEDAI, ECLAM and the British Isles Lupus Assessment Group (BILAG)-2004 index, all have proven validity and reliability [2, 8, 9]. The BILAG-2004 index is the most comprehensive available instrument. It replaced the original ‘classic’ BILAG index [10] and the current version includes numerical scoring and updates to haematology items [11, 12]. BILAG-2004 index comprises 97 discrete clinical manifestations of SLE across nine organ domains. The activity in each domain is graded separately from A, highly active and likely to necessitate escalation in therapy, to E, no current or previous disease activity [13]. BILAG-2004 captures several important disease features such as haemolysis, lymphadenopathy, polyneuropathy, interstitial pneumonitis, gastrointestinal and significant although rare ophthalmological items that do not feature in SLEDAI. Additionally, all organ domains carry potentially equal weighting. Unlike SLEDAI, BILAG-2004 index differentiates between disease features that have partially, but not completely improved, those that have not changed, and those that are worse. It is thus more sensitive than SLEDAI to changes in disease activity over time and better detects partial response to therapy and the exacerbation of already active disease features [14, 15]. These differences may be particularly important in research studies.

The current BILAG-2004 documentation relies on an index case report form, a detailed glossary of clinical items and a separate scoring algorithm for each of the nine organ domains such that the overall A–E domain scores, notably renal and haematological, are frequently not available at the point of completing the case report form. The requirement to rate items as improving, worse, or the same requires assessment of the status 1 month earlier. Formal training is also recommended. Therefore, despite its advantages, BILAG-2004 index may be difficult or time-consuming to complete during routine clinic visits, particularly for those not familiar with the glossary, the layout of the case report form and scoring document, even though most patients have relatively few abnormal items present at any single visit.

The Easy-BILAG project aimed to develop and validate a simplified tool to record and score the current published version of the BILAG-2004 index [11] more rapidly and accurately for use in routine clinical care.

Methods

Development process

Easy-BILAG was registered as a multicentre quality improvement initiative with the Quality Assurance and Governance department of Leeds Teaching Hospitals NHS Trust and at the relevant governance departments at individual participating trusts. No real patient data was used in Easy-BILAG validation material and specific research ethics approval was therefore not obtained. BILAG-BR is an ongoing prospective study with research ethics approval from NRES Committee North West–Greater Manchester West (REC: 09/H1014/64) and Health Research Authority approved on 9 November 2009 (IRAS ref. 24407).

Reasons underlying inaccuracy or difficulty completing BILAG-2004 assessment were discussed in meetings of the BILAG group of expert clinicians, experienced in delivering BILAG-2004 training and adjudicating clinical trials. The frequency of BILAG-2004 clinical items was evaluated in an active SLE cohort. Based on these insights, a series of Easy-BILAG prototypes was developed to address the key problems identified. The finalized Easy-BILAG was validated in a workbook exercise. Minor changes to the wording in the constitutional and renal domains were made after validation as directed by feedback from participants. A separate self-adjudication checklist was also added, to assist with use in a clinical trial setting, after the validation.

Determining the frequency of BILAG-2004 SLE manifestations

Pseudonymized BILAG-2004 disease activity scores, from individual SLE patients enrolled in the UK BILAG-Biologics Registry (BILAG-BR) between March 2010 and November 2019, were available for evaluation. The majority of enrolled patients had moderate to severe SLE disease activity and were commencing biologic therapy [16]. The frequency with which each of the 97 BILAG-2004 clinical items was recorded as ‘new’, ‘improving’, ‘same’, or ‘worse’ was quantified. Quantitative items (i.e. full blood count, creatinine, eGFR, proteinuria and blood pressure) were recorded as raw numerical values in this dataset, independent of attribution to SLE disease activity and these were therefore excluded from this analysis.

Validation exercise

Consultant rheumatologists including members of BILAG as well as those without sub-speciality interests in SLE, rheumatology specialty trainees and experienced lupus specialist nurses from 14 UK centres were invited by BILAG members to complete a timed validation workbook of 10 short case vignettes. Workbooks were distributed to participating centres by mail. Each participant was provided with a workbook that randomly assigned them to Easy-BILAG or standard format. Each clinician thereby scored BILAG-2004 disease activity for the same 10 cases using either Easy-BILAG or standard format BILAG-2004 index form, glossary and scoring algorithm [11]. No prior BILAG-2004 training or experience was prerequisite. The validation workbook was designed to test clinicians in scoring both frequent and uncommon SLE manifestations, and test longitudinal scoring of items in flare and remission. All workbooks contained an introductory overview and detailed instructions on how...
to use the BILAG-2004 index. Individuals self-reported
the time taken to score the 10 cases. They were also
asked to report their level of prior BILAG-2004 experi-
ence, job role and whether they worked within a general
or tertiary hospital setting. Overall perception and usabil-
ity of the scoring format was assessed by four Likert-
item survey questions. Clinicians returned their
workbooks in hard copy to the central study team and
could opt to submit anonymously. Item and domain level
accuracy was calculated against a model answer scheme
verified in advance by original BILAG-2004 authors.

Statistical analysis

After demonstrating non-normality, variables were com-
pared between two groups in SPSS using Wilcoxon
rank-sum test and between multiple groups by Kruskall–
Wallis followed by relevant pairwise comparisons. Inter-
rater agreement for all domain scores was evaluated
according to BILAG-2004 format by Fleiss’ kappa coeffi-
cient using R Studio irr package [17] and the result catego-
rized as previously described [18]. Statistical signifi-
cance was considered \( P < 0.05 \) on one-tailed
testing.

Results

Key problems identified in BILAG-2004 scoring

Consultation among BILAG-2004 lead authors and ex-
pert clinicians identified features of the current published
recording format that pose barriers to its accurate and/
or time-efficient use in routine practice, many of which
are also relevant to clinical trials. Strategies to resolve
priority issues were developed. These are collectively
summarized in Supplementary Table S1, available at
Rheumatology online.

Item level frequency of BILAG-2004 SLE manifesta-
tions in a UK biologics registry

Item level data from 2395 BILAG-BR disease activity
records revealed that the six most frequent items were
each present in >20% of records, namely: mild arthritis
(72%), mild skin eruption (47%), moderate arthritis
(38%), mild mucosal ulceration (34%), mild alopecia
(34%) and pleurisy/pericarditis (22%). Twenty-two dis-
crete items were present in 5% or more of cases and
no Ophthalmic or Gastrointestinal domain items were
required grade A to C scores was examined separately.
Rating of active domains retained high levels of accur-
cy with use of Easy-BILAG \([94.6 (89.1, 97.2); \text{Fig. 3B}]\) as compared with the
standard BILAG-2004 format \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (Q_1, Q_3)]\) was signifi-
cantly shorter among clinicians using Easy-BILAG
\([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-
mature disease is more prone to error, rating of domains
requiring grade A to C scores was examined separately.
Rater of active domains retained high levels of accuracy
with use of Easy-BILAG \([89.1, 90.2); \text{Fig. 3B}]\), which was significantly above that observed using
standard format BILAG-2004 \([80.4 (65.2, 90.2), \text{U} = 48.0, \text{P} = 0.001; \text{Fig. 3B}]\). Self-reported completion time for
the 10-case workbook \([\text{min; median} (59.5, 69.5); \text{Fig. 3C}]\) than standard for-

Annotated Easy-BILAG page 1 in full demonstrates the structure and scoring mechanism for the most frequently encountered SLE features in addition to the constitutional and renal domains in full. This single page is sufficient to assess the majority of patients. Items are organized within tables by organ domain. Each is anchored to its relevant colour-coded key to the left. Standard BILAG-2004 principles apply throughout. The highest tariff item triggers the final domain score. Scoring requires clinicians to circle each listed item as ‘not present’ (0), ‘improving’ (1), ‘same’ (2), ‘worse’ (3), or ‘new’ (4), and colour-coding translates to overall organ domain score from A (blue, highly active), B (green, moderate active), C (yellow, low active), D (pink, inactive), E (white, normal).

For the following systems, circle all that apply and total the points to determine the domain score.

Update the final domain scores with any items recorded on page two.

No other items? BILAG COMPLETE

(continued)
Easy-BILAG matches general hospital and tertiary centre clinicians for accuracy and speed

The performance of Easy-BILAG outside centralized, subspecialty and research-orientated tertiary centres was evaluated. Accuracy in grading active domains [% accuracy; median (Q1, Q3)] was high among tertiary centre clinicians using standard format BILAG-2004 [n = 9; 86.9 (70.7, 95.7) Fig. 3D] but still trended to higher accuracy among those using Easy-BILAG [n = 11; 95.7 (90.2, 96.7), P = 0.06 Fig. 3D]. In contrast, general hospital clinicians using standard format BILAG-2004 returned significantly lower accuracy [n = 8; 75.0 (64.1, 83.7)] than those testing Easy-BILAG [n = 5; 91.3 (90.2, 96.7), P = 0.02; Fig. 3D] while the latter achieved accuracy comparable to tertiary centre colleagues (P = 0.70). Among general hospital clinicians, workbook completion was significantly faster [time in min; median (Q1, Q3)] with Easy-BILAG [n = 5; 59.0 (51.0, 63.0), Fig. 3E] than with standard format [n = 8; 97.0 (76.0, 113.0), P = 0.01; Fig. 3E]. Easy-BILAG achieved similarly high and matched levels of accuracy between clinicians with and without prior BILAG-2004 training, and between clinicians who regularly or infrequently use BILAG-2004 in their existing practice (Supplementary Fig. S3, available at Rheumatology online).

Easy-BILAG improved inter-rater agreement on disease activity

Overall inter-rater agreement in workbook disease activity grading using standard BILAG-2004 format was classified as good, though levels of agreement in assigning active grade B disease were lower (Table 1). Clinicians testing Easy-BILAG demonstrated higher levels of inter-rater agreement overall and across each level of active disease graded A–C where agreement was classified as very good (Table 1).

Fig. 1 Continued
(pink, moderately active), C (yellow, mild stable disease) to D (white, no disease activity, but domain previously active) or E (white, no current or prior disease activity). Free text space is included to the right of each domain for the clinician to assign the final domain score from A–E. An abbreviated glossary definition is included immediately adjacent to each item. Example shows ‘Skin eruption—severe’ recorded as ‘improving’ (pink) triggering overall mucocutaneous domain grade B; ‘Arthritis (mild)/arthralgia’ recorded as ‘same’ (yellow) triggering overall musculoskeletal domain grade C. Cardiorespiratory domain example is currently in remission prompting grade D when history indicates the domain has previously been active. Grade E is assigned in the neuropsychiatric domain where there is no current or prior disease activity. The haematological domain requires clinicians to circle where each full blood count item is located and indicate whether ‘Other features’ are present. Haemoglobin is scored separately according to whether haemolysis is present or absent. The severity of anaemia/cytopenia translates to overall domain score based on the same colour-coded key (left) used throughout. Example shows overall grade C triggered by ‘haemoglobin 8–10.9 g/dl without haemolysis’ and ‘lymphocyte count <1.0 x 10⁹/L’ (both yellow) with no ‘Other features’ present. Gastrointestinal and Ophthalmic domains prompt clinicians to check page 2 of Easy-BILAG for relevant features or assign grade D or E (see Fig. 2). Constitutional and Renal domains derive domains scores using a tally of arbitrary points assigned to each item. A separate key anchored to both domains directs overall scoring according to points tally. Example shows constitutional domain grade C triggered by ‘Anorexia’ and ‘Weight loss’ rated as ‘improving’ (points tally 20, yellow) and renal domain grade A triggered by ‘Biopsy nephritis past 3 months’ and ‘urine protein : creatinine ratio >100 and unimproved’ (points tally 2000, blue). If no items signpost to page 2, as in this example, the assessment is complete.

Easy-BILAG has high usability for routine practice

Clinician feedback was surveyed by Likert scale responses on completion of the validation exercise. All clinicians testing Easy-BILAG agreed that the format was intuitive and sufficiently detailed, while 12/16 (75.0%) agreed Easy-BILAG was a convenient format for disease activity recording (Fig. 3F). By visual analogue scale [VAS 0—highly unlikely; 10—highly likely; median, (Q1, Q3)], clinicians from both tertiary and general hospitals testing Easy-BILAG reported significantly higher likelihood of adopting the tool in regular practice [n = 16; VAS: 8.5 (7.5, 10.0); Fig. 3G] than those using standard format BILAG-2004 [n = 16; VAS: 5.0 (3.0, 8.5); U = 68.0, P = 0.01].

Discussion

We report a novel tool for recording the BILAG-2004 disease activity index that combines full fidelity to the current instrument [11] with a condensed, data driven, colour-coded design adapted to assist clinicians to assess SLE in routine practice. In this validation exercise it yielded superior accuracy and time efficiency over current standard format BILAG-2004 index using paper forms, with a particularly marked advantage among clinicians based outside tertiary centres. It showed improved inter-rater reliability and perceived usability for routine practice among clinicians of varying levels of prior experience across different hospital settings. Thus Easy-BILAG offers a tool with which clinicians can confidently, accurately and time-efficiently integrate BILAG-2004 index into routine practice.

Treat-to-target principles have demonstrated wide-ranging benefits on patient outcomes across rheumatic diseases and it is clear that this approach reduces both flare rates and damage accrual in SLE [4]. To target
Fig. 2 Easy-BILAG page 2 captures infrequent and rare BILAG-2004 items

An annotated example Easy-BILAG page 1 Neuropsychiatric domain (A) shows ‘Mononeuropathy’, ‘Polyneuropathy’ and ‘Seizure disorder’ rated as ‘not present’ but evidence of ‘Other features’ prompts review of the full Neuropsychiatric domain on Easy-BILAG page 2 (B). Easy-BILAG page 2 in full (B) shows the scoring template for uncommon and rare SLE features not captured by page 1. Items are organized within tables by organ domain. Gastrointestinal and Ophthalmic domains are represented in full, followed by less frequently scored spillover items from the Mucocutaneous (Mucocutan), Musculoskeletal (MSK), Cardiorespiratory, Neuropsychiatric and Haematological (Haem) domains, signposted on page 1. Colour-coding translates to overall organ domain scores from A (blue), B (pink), C (yellow) to D or E (white) as directed by the key on Easy-BILAG page 1. Example shows ‘Psychosis’ rated as ‘new’ (grey arrowheads) thereby triggering overall grade A in the Neuropsychiatric domain free text space on the main page (A). The rater need only complete specific sections of page 2 if directed to by prompts on page 1, remaining sections can be left unmarked.
remission and low disease activity states, clinicians must be equipped with tools that allow them to measure disease activity accurately and consistently. BILAG-2004 index is a highly comprehensive disease activity instrument which assures clinicians that they have performed a thorough SLE assessment. It can therefore support and prompt clinical decision making, but its administrative burden is high and inadequately adapted to routine practice. The other major disease activity instrument, SLEDAI, is often felt to be the quicker and easier instrument [9] as it offers a limited item, fixed scoring format. It does, however, have different limitations, for example, some features such as autoimmune haemolysis are not scored at all, and due to the fixed points weighting, features such as lupus rash or thrombocytopenia, however severe, can never in isolation translate to the high disease activity score. Further, arthritis always scores twice as highly as rash even if a mild arthritis and severe rash are present. Being a binary system it does not distinguish items that are partially improved from those that are unchanged or worse. Therefore SLEDAI does not always align well with the patient experience or physician’s intention to treat [15]. By re-ordering items by their frequency and screening questions for rarer manifestations Easy-BILAG offers a balance between the simplicity and speed of the SLEDAI for common features while retaining the sensitivity to change and scope of BILAG-2004.

User feedback indicates Easy-BILAG had substantially greater appeal and usability for routine practice than standard format BILAG-2004 index. Thus Easy-BILAG permitted significant time saving for clinicians. Using Easy-BILAG, professionals with a range of prior BILAG-2004 experience completed 10 validation case assessments in an average time of just below 6min per case. This time accounts for reading and evaluating the case material, and the cases were designed to cover both rare and common manifestations. Although real-world time required to document and derive a score for a patient has not been evaluated here, the present data suggest a potential time saving in clinical settings. Completing BILAG-2004 index accurately in standard published format requires consulting a clinical glossary and a separate scoring algorithm for each organ domain. Easy-BILAG makes use of colour-coding and abridged glossary descriptions to reduce reliance on separate reference documents. Using data-driven design, the same BILAG-2004 assessment would require only a single page assessment for the majority of UK registry patients. Since its design was informed by item frequency in a biologics registry it is likely that far more routine outpatient visits would be less complex and quicker.

Easy-BILAG facilitated excellent scoring accuracy when tested against model case vignettes. Crucially, a similarly high level of accuracy was achieved by clinicians based in general hospital practice compared with those in tertiary centres where more subspecialist and research-focused activities using BILAG-2004 are typically concentrated. We also found that similar accuracy was maintained irrespective of prior training or current use of BILAG-2004 index. It is important to note that our validation exercise included a detailed overview and scoring instructions for BILAG-2004 index. Therefore Easy-BILAG is not a replacement for appropriate training, but our findings do suggest it is an accessible format for clinicians across all major areas of practice and can be more readily applied by those with less prior experience. Training material used in the current validation exercise is also made available alongside Easy-BILAG through the University of Leeds licensing platform.

Easy BILAG not only showed higher scoring accuracy, but also substantially less variability. We found that in this group of real-world clinicians with a range of prior BILAG-2004 experience, Easy-BILAG achieved better inter-rater agreement across all disease activity grades than standard format scoring. Importantly variation was not simply a function of clinical role, practice setting or prior experience as Easy-BILAG appeared to show an advantage among all professional categories. In practice this has implications for continuity of care where a team of various clinicians may review or manage an individual during follow-up and make treatment decisions such as whether to continue a biologic therapy. In the validation exercises, scoring with standard format appeared particularly vulnerable to inconsistencies assigning scores to grade B disease, which has also been observed in defining moderate flares [19]. Easy-BILAG was particularly beneficial in these cases and among these mixed-experience clinicians. Since grade B disease in two or more organ domains defines moderate disease activity in current EULAR, BSR and NHS England guidelines [1, 2, 6], this is a particularly decisive aspect of scoring where inconsistencies could conceivably introduce unwarranted variation in access to therapies. Improved recording accuracy might improve the identification of flares in clinical practice and the appropriate referral for new therapies.

Although our main goal in development of the Easy-BILAG was for use in routine clinical practice, our findings may have implications for clinical studies. Robust disease activity grading is central to trial outcome measures and the success or failure of new therapies in SLE. While detailed training in BILAG-2004 is essential for clinical investigators, we found that even among experienced lupus clinicians BILAG-2004 accuracy can be enhanced and variability reduced by formatting the assessment in a novel way. Investigators should be mindful of how data collection tools are presented to clinicians and the example of Easy-BILAG may be relevant to disease areas beyond SLE. Specific validation of Easy-BILAG in clinical trials has yet to be undertaken and is beyond the scope of this study but we would suggest that its potential advantage in scoring accuracy could reduce the burden of disease activity adjudication required in clinical trials. We would also anticipate that integrating BILAG-2004 into more
Box plots (A–C) show performance of Easy-BILAG against standard format BILAG-2004 in self-timed validation workbook exercise for all professional categories (overlay jitter points). Scoring accuracy (%) against model workbook answers for all organ domains (A) and active organ domains (B) was significantly higher with Easy-BILAG. Time taken to complete workbook exercise (C) was significantly shorter with Easy-BILAG. Box plots (D–E) show Easy-BILAG (blue) versus standard format BILAG-2004 (white) across general and tertiary practice. Scoring accuracy (D) was significantly higher and time taken (E) reduced among general hospital rheumatologists using Easy-BILAG and equivalent to tertiary rheumatologists. Stacked horizontal bar chart (F) shows Easy-BILAG rated favourably for intuitiveness, detail and convenience on 5-point Likert-scale. Box plot (G) shows likelihood to adopt instrument in routine practice was significantly higher by 10-point visual analogue scale (10—highly likely, 0—highly unlikely) for Easy-BILAG among clinicians across general and tertiary practice (overlay jitter). *** $P \leq 0.001$; ** $P \leq 0.01$; $P \leq 0.05$; ns: non-significant, $P > 0.05$

Routine practice through use of Easy-BILAG could facilitate more robust data collection in clinical practice and better enable clinicians to take part in clinical trials and identify eligible patients.

The current work has some limitations. First, due to constraints on space not all glossary items could be incorporated for rare items scored on Easy-BILAG page 2 and clinicians would still need to refer to core BILAG-
2004 official glossary if assessing these features. Second, as validation work was conducted on expert standardized training material, Easy-BILAG is yet to be evaluated in real-world practice with real patients. Its ability to maintain accuracy in capturing disease flares and response to therapy requires further validation. Similarly the time advantage in completing training material may not reflect use in the clinical or research setting and this would require confirmation in a further study. Additionally, applying criteria for improvement and worsening criteria as well as the ‘trickle-down’ rule all still require appropriate understanding and training on BILAG-2004 index. An optional self-adjudication checklist at the end of Easy-BILAG has been devised as a prompt for these areas. Computer-assisted formats with inbuilt quality assurance on all aspects of BILAG-2004 can offer additional error checking, which is particularly valuable in clinical trials [10]. As with the current published BILAG-2004 format, Easy-BILAG has been written in English, which could pose a barrier to use among clinicians and investigators for whom English is not first language. The validation work undertaken did not specifically characterize Easy-BILAG usability in non-first language English speakers.

In conclusion, Easy-BILAG should help to support the use of BILAG-2004 index in routine practice, assisting comprehensive and consistent assessment of patients and detection of disease activity, which in turn could facilitate compliance with biologic and other treatment guidelines. It is our recommended format for recording BILAG-2004 disease activity assessments in future clinical practice.

Acknowledgements

This research was supported by the United Kingdom National Institute for Health Research Leeds Biomedical Research Centre based at the Leeds Teaching Hospitals NHS Trust. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

We thank rheumatology clinicians from the following centres for their participation in the Easy-BILAG validation exercise: Leeds Teaching Hospitals NHS Trust, Sandwell and West Birmingham NHS Foundation Trust, Doncaster & Bassetlaw Teaching Hospitals NHS Foundation Trust, Royal National Hospital for Rheumatic Diseases, Oxford University Hospitals NHS Foundation Trust, Royal Berkshire NHS Foundation Trust, Frimley Health NHS Foundation Trust—Wexham Park, University College London Hospitals NHS Foundation Trust, University Hospitals Birmingham NHS Foundation Trust, Manchester University NHS Foundation Trust, East Lancashire Hospitals NHS Trust—Royal Blackburn Teaching Hospital, Belfast Health and Social Care Trust—Musgrave Park, Cambridge University Hospitals NHS Foundation Trust and Sheffield Teaching Hospitals NHS Foundation Trust.

Easy-BILAG is copyright of Edward Vital, University of Leeds. It is available for free use for routine clinical practice and academic studies, and for commercial purposes under license at https://licensing.leeds.ac.uk/products/healthcare-questionnaires.

Funding: This work was not supported by any dedicated grant funding.

Disclosure statement: Caroline Gordon reports personal fees for honoraria from consultancy work from the Center for Disease Control and Prevention, AbbVie, Amgen, AstraZeneca, MGP, Sanofi and UCB, and personal fees for Speakers Bureau from UCB.

David Isenberg has consulted for AstraZeneca, Merck Serono, Eli Lilly, Servier and Genentech. The honoraria offered are passed onto a local arthritis charity.

Sarah Skeoch reports speaker fees from Pfizer and grant income from Innovative Medicines Initiatives 2 Joint Undertaking under grant agreement number

<table>
<thead>
<tr>
<th>Easy-BILAG</th>
<th>Standard BILAG-2004 Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 15</td>
<td>n = 17</td>
</tr>
<tr>
<td>Fleiss’ kappa coefficient, $\kappa$</td>
<td>Fleiss’ kappa coefficient, $\kappa$</td>
</tr>
<tr>
<td>Overall</td>
<td>0.892 Very good</td>
</tr>
<tr>
<td>BILAG grade</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.877 Very good</td>
</tr>
<tr>
<td>B</td>
<td>0.833 Very good</td>
</tr>
<tr>
<td>C</td>
<td>0.882 Very good</td>
</tr>
<tr>
<td>D</td>
<td>0.751 Good</td>
</tr>
<tr>
<td>E</td>
<td>0.970 Very good</td>
</tr>
</tbody>
</table>

TABLE 1 Inter-rater agreement in BILAG-2004 domain scoring by validation workbook format
116106. This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and EFPIA.

Edward Vital has received research grants paid to his employer from AstraZeneca and Sandoz, and personal fees from Genentech, Aurinia, Lilly, AstraZeneca, ILTOO and Modus Therapeutics.

Lucy Carter and Chee-Seng Yee declare no conflicts of interest.

**Data availability statement**

The data underlying this article will be shared on reasonable request to the corresponding author.

**Supplementary data**

Supplementary data are available at Rheumatology online.

**References**


17 Garner M, Lemon J, Fellows I, Singh P. irr: Various Coefficients of Interrater Reliability and Agreement. 2019 [R package version 0.84.1]; https://CRAN.R-project.org/package=irr (19 September 2021, date last accessed).