



Clinical science

Predicting EQ-5D full health state in systemic lupus erythematosus using machine learning algorithms

João Botto^{1,2}, Nursen Cetrez^{1,2}, Dionysis Nikolopoulos^{1,2}, Malin Regardt^{3,4}, Emelie Heintz⁵, Julius Lindblom ^{1,2}, Ioannis Parodis ^{1,2,6,*}

¹Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

²Department of Gastroenterology, Dermatology and Rheumatology, Karolinska University Hospital, Stockholm, Sweden

³Department of Neurobiology, Care Sciences and Society, Division of Occupational Therapy, Karolinska Institutet, Stockholm, Sweden

⁴Women's Health and Allied Health Professionals Theme, Medical Unit Occupational Therapy and Physiotherapy, Karolinska University Hospital, Stockholm, Sweden

⁵Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden

⁶Department of Rheumatology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

*Correspondence to: Ioannis Parodis, Rheumatology, Karolinska University Hospital Solna, SE-171 76 Stockholm, Sweden. E-mail: ioannis.parodis@ki.se

Abstract

Objectives: To determine factors associated with reports of EuroQol 5-Dimensions (EQ-5D) full health state (FHS) before and after a trial intervention in patients with SLE, resorting to machine learning algorithms.

Methods: We conducted a post hoc analysis of two phase 3 clinical trials of belimumab (BLISS-52, BLISS-76). Demographic, laboratory and clinical features were retrieved and the Monte Carlo Feature Selection algorithm was employed, then further refined upon consideration of collinearity and clinical relevance. We used support vector machine with radial basis function kernel (SVMRadial), least absolute shrinkage and selection operator (LASSO), neural network (NNet) and logistic regression (LR) to capture both linear and non-linear relationships while ensuring interpretability and robustness.

Results: Among 1642 SLE patients, 12.9% reported FHS at baseline and 23.1% at week 52. Selected features were age, sex, Asian ancestry, baseline cSLEDAI-2K, SELENA-SLEDAI PGA, and urine protein:creatinine ratio (UPCR) and baseline EQ-5D 3-Levels (EQ-5D-3L) index score (week 52 models only). The models predicting FHS demonstrated comparable performance at baseline and week 52. A maximum area under the curve of 0.73 was seen for the baseline LASSO and LR models and a maximum of 0.77 for the week 52 LASSO and NNet models. Negative predictive values were high for all models (0.88–0.94). Calibration showed marginal improvement in week 52 models.

Conclusion: Machine learning identified older age, female sex, non-Asian ancestry, high disease activity and low UPCR to be associated with a lack of FHS experience in SLE patients at baseline and week 52. High baseline EQ-5D-3L index scores constituted the strongest predictor of FHS at week 52.

Trial registration: The BLISS-52 and BLISS-76 trials are registered at www.clinicaltrials.gov (NCT00424476 and NCT00410384, respectively).

Lay Summary

What does this mean for patients?

This research provides valuable insights into the factors that influence how patients with systemic lupus erythematosus (SLE) perceive their health over time. Using a type of artificial intelligence called machine learning, researchers analysed clinical trial data to determine which factors affect the likelihood of reporting a 'full health state' (FHS), i.e. an idealized state of complete health, based on the EQ-5D health questionnaire, both at the start of treatment and after 52 weeks. The study found that older age, female sex, non-Asian ancestry, higher disease activity and lower protein levels in the urine were linked to a lower likelihood of reporting an FHS. In contrast, patients who reported better perception of health status at the start of the trial were more likely to maintain this perception 1 year later. For patients, this means that individual characteristics and baseline health status can influence treatment outcomes and perceived quality of life. These findings support a more person-centred approach to care, helping clinicians personalize treatment strategies and integrate patient-reported experiences into disease evaluation and decision-making. By recognizing the factors that impact health perception, healthcare providers can better address patients' needs and expectations, ultimately improving their overall well-being.

Keywords: systemic lupus erythematosus, patient-reported outcomes, quality of life, EQ-5D, machine learning.

Key messages

- We used machine learning to determine factors associated with the EQ-5D full health state in SLE.
- Older age, female sex, non-Asian ancestry, high disease activity and low UPCR predicted non-FHS.
- High baseline EQ-5D-3L index scores constituted the main predictor of FHS experience at week 52.

Introduction

SLE is a complex autoimmune disease that can cause inflammation and damage to multiple organ systems. It predominantly affects women of childbearing age and poses a significant burden on patients' lives [1, 2]. In the assessment of this burden, patients' health-related quality of life (HRQoL) constitutes an important parameter. Patients with SLE often experience significantly lower HRQoL compared with healthy individuals [3] and those with other chronic diseases [4]. Researchers have strived to understand the impact of SLE on HRQoL and which factors are associated with its variations [3, 4]. It is important to know that a good clinical response does not always correspond to a favourable HRQoL outcome [5]. Therefore, evaluating HRQoL has become increasingly important in SLE research and clinical practice [6, 7] and was proposed as one of the principal outcome domains proclaimed during the Outcome Measures in Rheumatology (OMERACT) IV consensus conference [8].

The EuroQol 5-Dimensions (EQ-5D) health questionnaire is widely used for measuring HRQoL [9] and its psychometric properties have been validated in SLE [10]. The EQ-5D consists of a visual analogue scale (VAS) for self-assessing health state on a scale ranging from 0 (worst possible health state) to 100 (best possible health state) and a descriptive system comprising five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension is represented by one question, with three levels of severity in the EQ-5D-3L version: no problems (level 1), some problems (level 2) and extreme problems (level 3). The responses are put together to form an individual health profile (e.g. 21322) and can be converted into a single index score using preference-based population-specific scoring algorithms. This overall score can range from <0 (where 0 represents experience equal to death) to 1. A score of 'no problems' in all five dimensions yields an index score of 1, corresponding to an idealized state of complete health, which is termed the full health state (FHS) [11].

While therapeutic advancements have been made in the management of SLE [12], little is known about the factors that influence a patient's ability to achieve and maintain an EQ-5D FHS. Identification of variables with the ability to predict attainment of an EQ-5D FHS in SLE could provide clinicians with more realistic expectations and with valuable insights for individualized patient care and optimization of the allocation of healthcare resources.

In a recent study, we reported an EQ-5D FHS prevalence of 23% among SLE patients from a clinical trial setting at week 52 from baseline and disclosed factors such as female sex, increasing BMI and non-Asian ethnicity to be negatively associated with an EQ-5D FHS at week 52 [13]. The present study aimed to determine factors that are associated with an EQ-5D FHS experience or its absence in patients with SLE from the same clinical trial setting, both at baseline and after 52 weeks from commencement of the trial intervention. This article pioneers the application of machine learning algorithms for addressing this question, which constitutes a novel approach in this context.

Methods

Study population

Data analysis was conducted using information from two phase 3 clinical trials assessing the efficacy of belimumab, a

monoclonal antibody that hampers B cell survival and proliferation by binding to soluble B cell activating factor (BAFF), in patients with active SLE [BLISS-52 (NCT00424476; $n = 865$) [14] and BLISS-76 (NCT00410384; $n = 819$) [15]. All individuals enrolled in these trials satisfied the revised classification criteria for SLE established by the ACR [16], were adults and had an ANA titre $\geq 1:80$ and/or a serum anti-dsDNA antibody level ≥ 30 IU/ml at screening. Additionally, they exhibited a Safety of Estrogens in Lupus National Assessment–SLE Disease Activity Index (SELENA-SLEDAI) [17] score ≥ 6 at screening. All participants had been on a stable non-biologic standard therapy for at least 30 days prior to the trial baseline. Patients were not eligible for participation if they had active severe LN or CNS involvement.

The study participants were randomly assigned to one of three arms: belimumab 1 mg/kg, belimumab 10 mg/kg or placebo, administered intravenously, on top of standard therapy. The initial doses were given at weeks 0, 2 and 4 and were followed by subsequent doses every fourth week [14, 15].

Data from the BLISS trials were provided by GSK (Uxbridge, UK) through the Clinical Study Data Request consortium. Forty-two patients were excluded due to missing primary outcome data (EQ-5D FHS at baseline or week 52), resulting in a study population of 1642 SLE patients, to ensure robust outcome assessments.

Ethics

All study participants provided written informed consent. The trial protocols received approval from regional ethics review boards across all participating centres. The protocol of the present analysis received approval from the Swedish Ethical Review Authority (2019-05498). The BLISS study protocols and the protocol of the present investigation adhered to the ethical guidelines outlined in the Declaration of Helsinki.

Measures of disease activity and damage

The SLEDAI evaluates overall SLE activity using 24 descriptors, each assigned a specific weight based on its relative significance [18]. The SLEDAI-2000 (SLEDAI-2K) modification allows ongoing disease activity in the descriptors alopecia, mucosal ulcers, proteinuria and rash to be scored [19], unlike the original SLEDAI where only new occurrences are scored [18]. The clinical version of SLEDAI-2K (cSLEDAI-2K), which omits the serological descriptors (anti-dsDNA and complement) [20], was the one employed for assessing clinical disease activity in the present investigation.

The SELENA-SLEDAI physician global assessment (PGA) captures overall SLE activity on a scale ranging from 0 (none) to 3 (severe) [21].

Lastly, the Systemic Lupus International Collaborating Clinics (SLICC)/ACR damage index (SDI) measures irreversible organ damage that occurs from the time of SLE diagnosis onwards, irrespective of whether the damage is directly linked to SLE. The SDI comprises 39 items, grouped into 12 organ systems [22].

Study outcome

The outcome in this study was a patient-reported FHS (*vs* non-FHS) assessed using the EQ-5D-3L instrument at baseline and at week 52. A FHS was defined as a report of no

problems in all dimensions of the descriptive system of the EQ-5D-3L (health profile 11111).

Data collection

Demographic parameters, baseline clinical and laboratory features and concomitant medications were retrieved. EQ-5D data were retrieved from the baseline and week 52 visits. EQ-5D-3L index scores were calculated based on the US population scoring algorithm [23].

Feature selection and dataset split

Machine learning models were developed with the goal of predicting a patient-reported EQ-5D FHS at baseline and week 52. Feature selection was conducted using the Monte Carlo Feature Selection (MCFS) algorithm, applied through the *rmcfs* R package version 1.3.5 (R Foundation for Statistical Computing, Vienna, Austria). The MCFS was chosen due to its robustness and capacity to handle feature selection across multiple permutations. This approach was favoured over alternatives due to its ability to generate a prioritized list of relevant features and manage complex interdependencies. The only difference between the pools of variables input into the algorithm for the baseline and week 52 models was the inclusion of the baseline EQ-5D index scores in the algorithm for the latter models.

With the MCFS, a prioritized list of relevant features was generated and served as the determinants of the outcome. Selection of the optimal number of features for the models was conducted by assessing their relative importance from a ranked list, using the permutation method [24]. Next, further refinement of the feature selection was made based on collinearity and clinical relevance.

Patients with incomplete data on the selected features and outcome variables were removed from the datasets during subsequent analysis. Preprocessing steps included data normalization, conducted internally by the *caret* R package (see below) during training, ensuring features were scaled appropriately for algorithms like SVM and neural networks. Upon finalization of the datasets for the baseline and week 52 analysis, each dataset was randomly split into a training and a test set at a ratio of 4:1, respectively, using the *caret* R package version 6.0-94 [25]. Since there was a class skew in the outcome variable, oversampling was applied to the training set to address this issue, using the R package ROSE version 0.0-4 [26].

Training and testing of the machine learning models

The following algorithms were employed: support vector machine with radial basis function kernel (SVMRadial), least absolute shrinkage and selection operator (LASSO), neural network (NNet), and logistic regression (LR). SVMRadial and NNet were selected for their capacity to model non-linear relationships, while LASSO and LR offered complementary insights into linear relationships with improved interpretability. Tree-based methods such as random forests or gradient boosting machines, although suitable for capturing feature interactions, were not employed, prioritizing model simplicity and interpretability given the dataset size and scope. Predictions were classified as positive (i.e. EQ-5D FHS experience) if their respective probabilities were ≥ 0.5 . To prevent overfitting, hyperparameter tuning was automatically done through 10 times 10-fold cross-validation (CV).

Both training and CV were conducted using the *caret* R package [25]. The optimal hyperparameters for the models were as follows: SVMRadial: $\sigma = 0.2498819$, $C = 1$ for baseline and $\sigma = 0.1994953$, $C = 0.25$ for week 52; LASSO: $\alpha = 1$, $\lambda = 0.002$ for baseline and $\alpha = 1$, $\lambda = 0.04$ for week 52; NNet: size = 5, decay = 0.1 for baseline and size = 1, decay = 0.1 for week 52.

After having obtained and trained the final models, these were evaluated using the test sets. The performance metrics encompassed accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), receiver operating characteristics (ROC) curves for illustrative purposes and area under the curve (AUC). The two latter performance measures were calculated using the *pROC* R package version 1.18.0 [27], while the others were assessed through the *caret* R package [25]. Finally, calibration curves were plotted and smoothed (resorting to the B-spline function) to determine the degree of alignment between predicted and observed outcome probabilities using the following Python libraries: *pandas* (version 1.5.3) [28], *scikit-learn* (version 1.2.1) calibration module [29], *NumPy* (version 1.23.5) [30], *SciPy* (version 1.10.0) interpolate module [31] and *Matplotlib* (version 3.7.0) *pyplot* module [32].

Statistical analysis

For normal distributions, descriptive statistics are indicated as number (percentage) or mean (s.d.), while for non-normal distributions, data are reported as median [interquartile range (IQR)]. Descriptive statistics and comparisons between patients who experienced an EQ-5D FHS and patients who did not were conducted using the *tableone* R package version 0.13.2 [33]. Differences yielding P -values < 0.05 were considered statistically significant.

All analyses were performed in RStudio version 2023.03.0 + 386 (R Foundation for Statistical Computing), with the exception of calibration curves, which were plotted and smoothed using Python version 3.10.9 (Python Software Foundation, Wilmington, DE, USA).

Results

Table 1 details demographic and baseline clinical characteristics for the overall study population as well as comparisons between patients who reported an EQ-5D FHS and patients who did not at baseline and at week 52. An EQ-5D FHS was reported by 12.9% of the patients at baseline and by 23.1% at week 52. The FHS subgroup (both at baseline and week 52) had more patients of Asian ancestry, while the non-FHS subgroup had more women and patients of older age or higher baseline BMI, SELENA-SLEDAI PGA, SDI or cSLEDAI-2K scores and lower baseline urine protein:creatinine ratio (UPCR). Patients in the FHS subgroup at week 52 reported greater EQ-5D index scores at baseline.

Feature selection

Results from the MCFS algorithm are presented in Supplementary Fig. S1, available at *Rheumatology Advances in Practice* online. The algorithm deemed 17 and 41 features informative for the baseline and week 52 models, respectively. Most of them were related to disease activity and organ damage. Other informative features included age, sex, Asian ancestry and baseline BMI, as well as baseline EQ-5D data for the week 52 models, which also constituted the most informative group of features for these models.

Table 1. Patients' demographic and baseline clinical features, comparing the EQ-5D FHS and non-FHS subgroups at baseline and at week 52.

Characteristics	Overall (N = 1642)	Baseline			Week 52		
		FHS (n = 212)	Non-FHS (n = 1430)	P-value	FHS (n = 380)	Non-FHS (n = 1262)	P-value
Female sex, n (%)	1544 (94.0)	192 (90.6)	1352 (94.5)	0.033	347 (91.3)	1197 (94.8)	0.015
Ancestry, n (%)							
Asian	332 (20.2)	76 (35.8)	256 (17.9)	<0.001	114 (30.0)	218 (17.3)	<0.001
Black/African American	142 (8.6)	8 (3.8)	134 (9.4)	0.010	22 (5.8)	120 (9.5)	0.031
Indigenous American	383 (23.3)	59 (27.8)	324 (22.7)	0.115	101 (26.6)	282 (22.3)	0.101
White/Caucasian	785 (47.8)	69 (32.5)	716 (50.1)	<0.001	143 (37.6)	642 (50.9)	<0.001
SDI score ≥ 1 , n (%)	694 (42.3)	70 (33.0)	624 (43.7)	0.004	111 (29.2)	583 (46.2)	<0.001
Age, years, mean (s.d.)	37.78 (11.52)	34.51 (11.36)	38.26 (11.47)	<0.001	34.19 (10.78)	38.86 (11.53)	<0.001
BMI, mean (s.d.)	25.55 (5.91)	23.97 (4.69)	25.78 (6.04)	<0.001	24.08 (4.85)	25.99 (6.13)	<0.001
EQ-5D index score, mean (s.d.)	0.74 (0.19)	1.00 (0.00)	0.70 (0.17)	NA	0.86 (0.15)	0.70 (0.18)	<0.001
SELENA-SLEDAI PGA, mean (s.d.)	1.42 (0.48)	1.26 (0.47)	1.45 (0.48)	<0.001	1.34 (0.49)	1.45 (0.48)	<0.001
SDI score ^a , mean (s.d.)	0.79 (1.24)	0.53 (1.05)	0.82 (1.26)	0.001	0.47 (0.95)	0.88 (1.30)	<0.001
SLEDAI-2K score, mean (s.d.)	9.94 (3.82)	9.46 (3.76)	10.01 (3.82)	0.051	9.67 (3.99)	10.02 (3.76)	0.127
cSLEDAI-2K score, mean (s.d.)	7.31 (3.63)	6.21 (3.65)	7.47 (3.60)	<0.001	6.66 (3.84)	7.51 (3.54)	<0.001
SLE disease duration, years, median (IQR)	4.43 (1.48–9.32)	5.05 (1.56–9.46)	4.38 (1.47–9.28)	0.892	5.84 (5.64)	6.56 (6.54)	0.054
UPCR (mg/mg), median (IQR)	0.15 (0.09–0.36)	0.17 (0.10–1.10)	0.14 (0.09–0.33)	<0.001	0.56 (0.93)	0.45 (0.91)	0.046

NA: not applicable.

P-values in bold are statistically significant ($P < 0.05$).

^a This variable is not normally distributed; however, values are presented as means since medians would not be informative about differences between subgroups.

With the posterior refinement of feature selection, the following features were made final for both models: age, sex, Asian ancestry, baseline cSLEDAI-2K score, baseline SELENA-SLEDAI PGA and baseline UPCR. For the week 52 models, apart from these features, EQ-5D index scores were also included.

Training and test sets

Each training set contained 80% ($n = 1314$) of the total dataset and each test set contained the remaining 20% ($n = 328$). No statistically significant differences were found for any of the final features between the training and test sets, at either time point (Supplementary Table S1, available at *Rheumatology Advances in Practice* online). This suggests a similar feature distribution and indicates that the models were trained on datasets that are representative of the data encountered during testing.

Performance of the models

Regarding the training sets, the AUCs for the baseline SVMRadial, LASSO, NNet and LR models were 0.83, 0.73, 0.78 and 0.73, respectively. The corresponding values for the week 52 models were 0.83 for the SVMRadial model and 0.80 for the other three.

Regarding the test sets, the performance metrics for the baseline and week 52 models are presented in Table 2. Overall, the performance of the models was similar in the baseline and week 52 analyses. However, comparing the baseline models with the week 52 models, a slightly better performance was documented based on AUC metrics for the week 52 models, i.e. a maximum of 0.73 was seen for the baseline LASSO and LR models *vs* a maximum of 0.77 for the week 52 LASSO and NNet models. Additionally, the sensitivity and PPV were overall better in the week 52 models (sensitivity: 0.52–0.71 for baseline and 0.67–0.72 for week 52; PPV: 0.21–0.25 for baseline and 0.39–0.44 for week 52). Regarding NPVs, these were particularly high for all models (0.88–0.94). The ROC curves from the models at both time

points are shown in Fig. 1, and the confusion matrices are presented in Supplementary Tables S2–S9, available at *Rheumatology Advances in Practice* online. Calibration curves, displayed in Fig. 1, show that all models exhibited poor calibration results at both time points, overestimating their predictions. Nevertheless, the week 52 models were slightly better calibrated, with greater predicted probabilities.

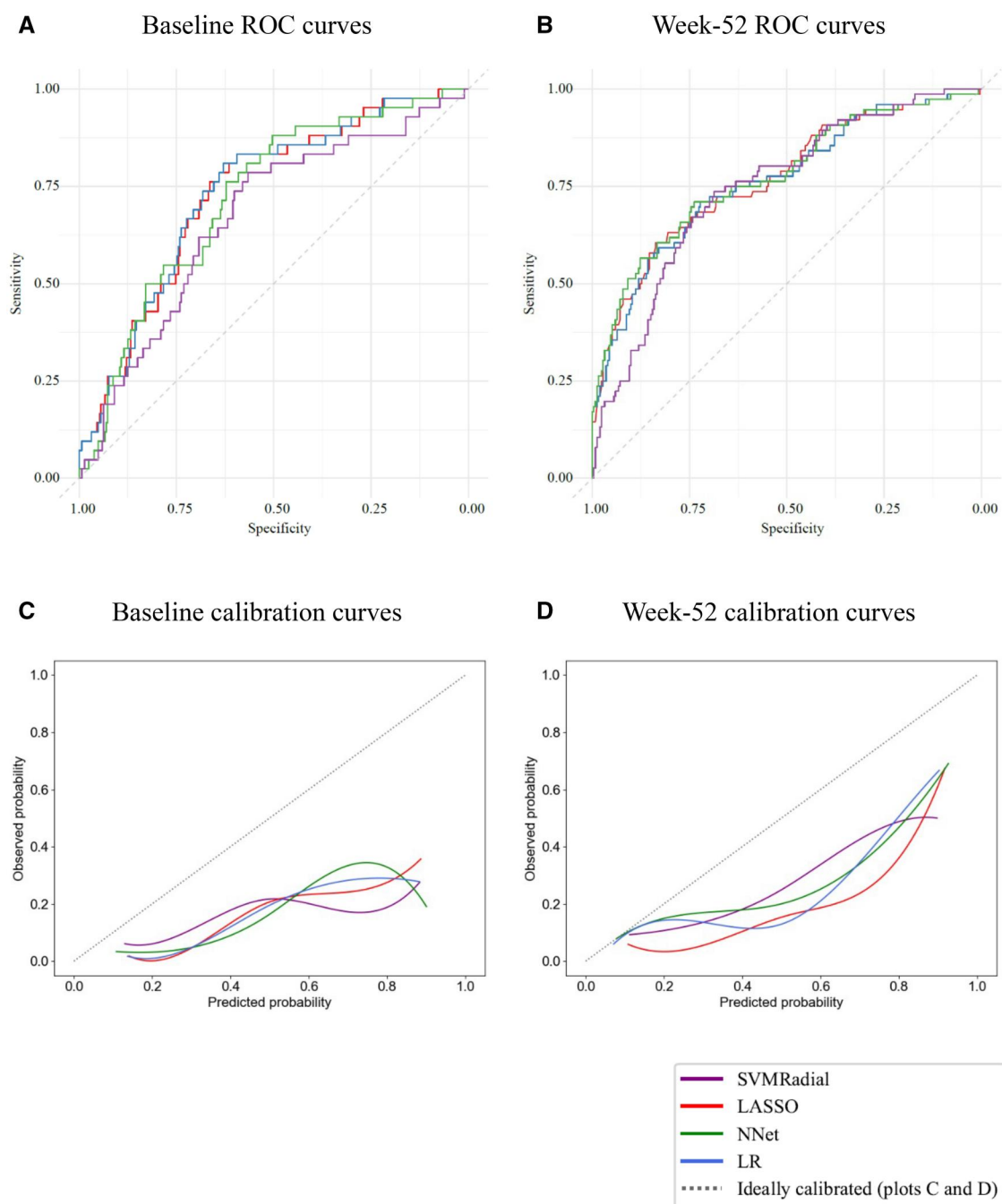
Discussion

The main purpose of this post hoc analysis of the BLISS-52 and BLISS-76 trials was to identify factors that are associated with EQ-5D FHS reports at baseline and 52 weeks after commencement of a therapeutic intervention with belimumab or placebo on top of standard therapy, using machine learning algorithms. Following feature selection by means of the MCFS algorithm, mitigation was attempted based on collinearity and clinical relevance. The final models demonstrated that older age, female sex, non-Asian ancestry, high SELENA-SLEDAI PGA, high cSLEDAI-2K scores and low UPCR were negatively associated with an EQ-5D FHS experience both at baseline and after the trial intervention, while high baseline EQ-5D index scores constituted the most important determinant of an EQ-5D FHS experience at week 52.

Relating to sex, our findings are in line with previous studies demonstrating women with SLE report worse HRQoL compared with men [34–36]. The relationships between age or disease activity and HRQoL have not been consistent in previous SLE literature [34–39]; we herein found that advancing age and greater clinical disease activity were negatively associated with an EQ-5D FHS experience. Discrepancies in the literature may be multifactorial, from differences across populations to the large variation in instruments used to measure disease activity or estimate HRQoL. The most common instrument for estimating HRQoL has probably been the Medical Outcomes Study 36-item Short Form [39], with only a few studies hitherto having used EQ-5D in SLE populations [13, 38]. Despite the large interrater variability for PGA scoring, higher PGA scores have

Table 2. Performance of baseline and week 52 models on the test sets.

Test	Baseline				Week 52			
	SVMRadial	LASSO	NNet	LR	SVMRadial	LASSO	NNet	LR
Accuracy (95% CI)	0.69 (0.64, 0.74)	0.69 (0.64, 0.74)	0.69 (0.63, 0.74)	0.69 (0.64, 0.74)	0.72 (0.67, 0.77)	0.67 (0.62, 0.72)	0.72 (0.67, 0.77)	0.69 (0.63, 0.74)
AUC	0.67	0.73	0.72	0.73	0.75	0.77	0.77	0.76
Sensitivity	0.52	0.71	0.55	0.69	0.67	0.72	0.71	0.72
Specificity	0.72	0.69	0.71	0.69	0.73	0.65	0.73	0.67
PPV	0.21	0.25	0.21	0.25	0.43	0.39	0.44	0.40
NPV	0.91	0.94	0.91	0.94	0.88	0.89	0.89	0.89

**Figure 1.** ROC curves on the test datasets from the (A) baseline and (B) week 52 models and smoothed calibration curves on the same datasets from the (C) baseline and (D) week 52 models

consistently been negatively associated with HRQoL in patients with SLE [40–42], which highlights that PGA captures features of disease activity beyond the span captured by the SLEDAI.

Although patients of Asian origin with SLE are known to have a more severe disease burden compared with White/Caucasian patients [43], we herein demonstrated that Asian patients more frequently reported an EQ-5D FHS compared with the non-Asian population. In fact, this is in line with previous literature that documented a tendency of Asian people with SLE to report better HRQoL [44], with social and cultural determinants influencing reporting trends likely underlying this observation.

There is limited information in previous studies regarding the relationship between renal involvement and SLE patients' HRQoL. Daleboudt *et al.* [45] reported a moderate inverse correlation between proteinuria and HRQoL in a study population of proliferative LN, where substantial reductions in proteinuria were seen after successful treatment. Higher UPCr levels were herein associated with an EQ-5D FHS experience, both at baseline and week 52, which contrasts with the study by Daleboudt *et al.* [45] and is seemingly counterintuitive. Importantly, the BLISS trials excluded patients with active severe LN during the screening phase, resulting in relatively low proteinuria levels in the overall study population, thus limiting the generalizability of this finding to SLE populations with more active renal disease. Nonetheless, our observation contrasts with what one would consider an expected finding and is therefore of some interest, as it underscores that subclinical proteinuria is not sensed by SLE patients in a way that affects patient-reported HRQoL. This calls for attentive surveillance of renal manifestations with frequent urinalysis and creatinine levels in people with SLE.

Finally, baseline EQ-5D-3L index scores constituted the most important feature in the week 52 models, with greater scores predicting an EQ-5D FHS experience at week 52. This corroborates previous findings by means of logistic regression in a study by Lindblom *et al.* [13].

The selected machine learning algorithms were chosen to balance model performance with interpretability. While our study did not include tree-based methods such as random forests or gradient boosting machines, the use of neural networks and support vector machines with radial basis function kernels allowed for the capture of non-linear relationships in the data. This approach was deemed appropriate for the study objectives while avoiding the complexity and potential overfitting associated with tree-based methods, given the dataset size. Well-performing and easy-to-apply models for predicting an EQ-5D FHS experience in SLE would be useful, particularly since EQ-5D assessment has shown the ability to distinguish therapy responders from non-responders [13, 46] and FHS experience has been linked to a lower hazard to accrue organ damage [47]. Furthermore, the EQ-5D has demonstrated favourable psychometric properties in terms of validity and reliability [10], it is easy to fill (yielding a high rate of acceptability and completion) and it is simple to implement in clinical practice [48]. For all these reasons, the EQ-5D may be considered a useful tool for assessing HRQoL among currently used patient-reported outcome measures in SLE [46, 47], acknowledging that it provides important complementary information to current definitions of remission and low disease activity that solely rely on clinical and laboratory elements [49].

Regarding the performance of the machine learning models, no substantial differences were seen between AUCs deriving from the training and test sets, indicating an absence of overfitting issues. Moreover, AUCs from all models in the training sets were comprised of reasonable values, indicating the absence of underfitting as well. It is important to underscore that our models yielded very high NPVs, $\approx 90\%$, ensuring robust predictions of which patients are not likely to experience an EQ-5D FHS. In contrast, the downside of the models lies in the relatively low PPVs and the calibration curves, where a likely overestimation of predictions is noticed, with the week 52 models performing slightly better. This overestimation could stem from the class imbalance inherent in the dataset. Although oversampling using the ROSE package improved model training, it may have introduced synthetic data artifacts, limiting calibration accuracy. When executing oversampling, the algorithm randomly chooses examples from the minority class and replicates them, creating a balanced training set. This means that the positive examples (in our case the minority class) will yield a collection of repeated examples that does not provide sufficient information about the different possible distributions of features in positive examples. Moreover, if the oversampling technique does not adequately represent the variety of these examples in the original dataset (due to the randomness of example selection for replication), the models may not learn the nuanced patterns that are associated with the positive class.

The study has some limitations that should be acknowledged, including the post hoc nature of the analysis and the lack of data on comorbidities, socio-economic status and daily physical activity, features that are likely to contribute to SLE patients' HRQoL experience. Limitations also included the lack of explicit statistical comparisons between model performances. Our focus was on identifying associated features rather than optimizing model performance. While the absence of an external validation dataset limits generalizability, our internal validation strategy ensured that the training and test sets were representative of the overall dataset. Additionally, while tree-based methods such as random forests or gradient boosting machines could provide further insights into feature interactions and complex relationships, our selected models offered a balanced approach between interpretability and performance. This approach aligns with our aim to focus on clinically interpretable outputs and avoid overfitting given the dataset size. We did not handle outliers—our analysis focused on utilizing the data as fully as possible without extensive cleaning. To mitigate the potential impact of outliers, secondary binary variables were included for some continuous features, which demonstrated consistency with the trends observed in the primary data. The low number of positive training examples (i.e. EQ-5D FHS reports) and the oversampling technique used to address this issue may have negatively affected the performance of our models, especially regarding the prediction of positive instances as explained above. Moreover, the generalizability of our findings is hindered by the exclusion of patients with active severe LN or neuropsychiatric lupus. Nonetheless, the study also has some important strengths, including the large and diverse SLE population and the extensive longitudinal data collection with a low degree of missing data, contributing to the robustness of findings. Another strength is the application of machine learning algorithms, which, to our knowledge, are

used for the first time for the prediction of patient-reported outcomes, particularly the EQ-5D FHS experience.

Machado Escobar *et al.* [50] suggested that an EQ-5D index score ≥ 0.739 is indicative of good HRQoL among SLE patients, and a recent study by Hua *et al.* [46] advocated that an EQ-5D index score ≥ 0.800 distinguishes responders to treatment from non-responders. Thus an analysis focusing on predictive factors for achieving EQ-5D health states below FHS could offer valuable insights, as would study of determinants to differential responses across the different EQ-5D dimensions.

Conclusions

Through a machine learning approach, we determined factors associated with an EQ-5D FHS experience at baseline and 52 weeks after the commencement of therapeutic intervention in a clinical trial setting. We found older age, female sex, non-Asian ancestry, high disease activity and low UPCR levels to be negatively associated with EQ-5D FHS reports. High baseline EQ-5D index scores constituted the most important predictor of an EQ-5D FHS at week 52. Our study contributes insights into determinants of patient-reported outcomes, a current major need towards person-centred care and integration of patient-reported health experiences in disease evaluation tools and treatment decisions.

Supplementary material

Supplementary material is available at *Rheumatology Advances in Practice* online.

Data availability

The datasets used and analysed in the present study can be made available through the Clinical Study Data Request consortium.

Authors' contributions

J.B., N.C. and I.P. were responsible for the conception and design of the work. N.C., J.L. and I.P. were responsible for the acquisition of data. J.B., N.C., D.N., M.R., E.H., J.L. and I.P. were responsible for statistical analysis and interpretation of data. J.B. and I.P. wrote the original draft. All authors were responsible for critical revision of the manuscript for important intellectual content, reviewed and approved the final version of the manuscript prior to submission and agree to be accountable for all aspects of the work.

Funding

This work was supported by the GSK Investigator-Sponsored Studies (ISS) program and grants from the Swedish Rheumatism Association (R-969696), King Gustaf V's 80-year Foundation (FAI-2020-0741), Swedish Society of Medicine (SLS-974449), Nyckelfonden (OLL-974804), Professor Nanna Svartz Foundation (2021-00436), Ulla and Roland Gustafsson Foundation (2021-26), Region Stockholm (FoUI-955483) and Karolinska Institutet. The funders were not involved in the design of the study; the collection, analysis or interpretation of data; writing of the manuscript or the decision to submit it for publication.

Disclosure statement: I.P. has received research funding and/or honoraria from the EuroQol Research Foundation, Amgen, AstraZeneca, Aurinia, BMS, Lilly, Gilead, GSK, Janssen, Novartis, Otsuka and Roche. The other authors declare no conflicts of interest.

Acknowledgements

The authors wish to express their gratitude to GSK (Uxbridge, UK) for sharing the data from the BLISS-52 (NCT00424476) and BLISS-76 (NCT00410384) trials with the Clinical Study Data Request Consortium, as well as to all participating patients.

References

1. Barber MRW, Drenkard C, Falasinnu T *et al.* Global epidemiology of systemic lupus erythematosus. *Nat Rev Rheumatol* 2021;17:515–32.
2. Shannon G, Eileen F, Robert O, Michael P, Sandra Sze-jung W. Real-world burden of systemic lupus erythematosus in the USA: a comparative cohort study from the Medical Expenditure Panel Survey (MEPS) 2016–2018. *Lupus Sci Med* 2022;9:e000640.
3. McElhone K, Abbott J, Teh LS. A review of health related quality of life in systemic lupus erythematosus. *Lupus* 2006;15:633–43.
4. Jolly M. How does quality of life of patients with systemic lupus erythematosus compare with that of other common chronic illnesses? *J Rheumatol* 2005;32:1706–8.
5. Gomez A, Qiu V, Cederlund A *et al.* Adverse health-related quality of life outcome despite adequate clinical response to treatment in systemic lupus erythematosus. *Front Med* 2021;8:651249.
6. Annapureddy N, Devilliers H, Jolly M. Patient-reported outcomes in lupus clinical trials with biologics. *Lupus* 2016;25:1111–21.
7. Yen JC, Neville C, Fortin PR. Discordance between patients and their physicians in the assessment of lupus disease activity: relevance for clinical trials. *Lupus* 1999;8:660–70.
8. Strand V, Gladman D, Isenberg D *et al.* Endpoints: consensus recommendations from OMERACT IV. *Lupus* 2000;9:322–7.
9. Gusi N, Olivares PR, Rajendram R. The EQ-5D health-related quality of life questionnaire. In: VR Preedy, RR Watson, eds. *Handbook of disease burdens and quality of life measures*. New York: Springer, 2010:87–99.
10. Aggarwal R, Wilke CT, Pickard AS *et al.* Psychometric properties of the EuroQol-5D and Short Form-6D in patients with systemic lupus erythematosus. *J Rheumatol* 2009;36:1209–16.
11. EuroQol Research Foundation. EQ-5D-3L user guide. 2018. <https://euroqol.org/publications/user-guides> (12 September 2023, date last accessed).
12. Athanassiou P, Athanassiou L. Current treatment approach, emerging therapies and new horizons in systemic lupus erythematosus. *Life* 2023;13:1496.
13. Lindblom J, Gomez A, Borg A *et al.* EQ-5D-3L full health state discriminates between drug and placebo in clinical trials of systemic lupus erythematosus. *Rheumatology (Oxford)* 2021;60:4703–16.
14. Navarra SV, Guzmán RM, Gallacher AE *et al.* Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet* 2011;377:721–31.
15. Furie R, Petri M, Zamani O *et al.* A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011;63:3918–30.
16. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
17. Petri M, Kim MY, Kalunian KC *et al.* Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med* 2005;353:2550–8.

18. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. *Arthritis Rheum* 1992;35:630–40.
19. Gladman DD, Ibanez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol* 2002;29:288–91.
20. Uribe AG, Vila LM, McGwin G Jr, *et al.* The Systemic Lupus Activity Measure-revised, the Mexican Systemic Lupus Erythematosus Disease Activity Index (SLEDAI), and a modified SLEDAI-2K are adequate instruments to measure disease activity in systemic lupus erythematosus. *J Rheumatol* 2004;31:1934–40.
21. Elisabetta C, Matteo P, Alberto F *et al.* Use of physician global assessment (PGA) in systemic lupus erythematosus: a systematic review of its psychometric properties. *Rheumatology (Oxford)* 2020;59:3622–32.
22. Gladman D, Ginzler E, Goldsmith C *et al.* The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus. *Arthritis Rheum* 1996;39:363–9.
23. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care* 2005;43:203–20.
24. Dramiński M, Koronacki J. rmcfs: an R package for Monte Carlo feature selection and interdependency discovery. *J Stat Softw* 2018;85:1–28.
25. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw* 2008;28:1–26.
26. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. *R J* 2014;6:79–89.
27. Robin X, Turck N, Hainard A *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
28. McKinney W. Data structures for statistical computing in python. *Proc 9th Python Sci Conf (scipy 2010)* 2010;56–61.
29. Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in python. *J Mach Learn Res* 2012;12:2825–30.
30. Harris CR, Millman KJ, van der Walt SJ *et al.* Array programming with NumPy. *Nature* 2020;585:357–62.
31. Virtanen P, Gommers R, Oliphant TE *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72.
32. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5.
33. Agapio P, Dimitris M. TableOne: an online web application and R package for summarising and visualising data. *Evid Based Mental Health* 2020;23:127.
34. Gu M, Cheng Q, Wang X *et al.* The impact of SLE on health-related quality of life assessed with SF-36: a systemic review and meta-analysis. *Lupus* 2019;28:371–82.
35. Gaballah NM, El-Najjar AR. Clinical characteristics and health related quality of life (HRQoL) in Egyptian patients with systemic lupus erythematosus. *Egypt Rheumatol* 2019;41:117–21.
36. Nyambane E, Genga E, Achieng L, Oyoo O, Orieno FC. P188 Assessment of disease activity and health related quality of life in patients with systemic lupus erythematosus at Kenyatta National Hospital. *Lupus Sci Med* 2020;7:A121–2.
37. Plantinga L, Lim SS, Bowling CB, Drenkard C. Association of age with health-related quality of life in a cohort of patients with systemic lupus erythematosus: the Georgians Organized Against Lupus study. *Lupus Sci Med* 2016;3:e000161.
38. Frederick W, Kaleb M, Tracy LI, Robert SK. EQ-5D and SF-36 quality of life measures in systemic lupus erythematosus: comparisons with rheumatoid arthritis, noninflammatory rheumatic disorders, and fibromyalgia. *J Rheumatol* 2010;37:296–304.
39. Schmeding A, Schneider M. Fatigue, health-related quality of life and other patient-reported outcomes in systemic lupus erythematosus. *Best Pract Res Clin Rheumatol* 2013;27:363–75.
40. Chaigne B, Chizzolini C, Perneger T *et al.* Impact of disease activity on health-related quality of life in systemic lupus erythematosus—a cross-sectional analysis of the Swiss Systemic Lupus Erythematosus Cohort Study (SSCS). *BMC Immunol* 2017;18:17.
41. Louthrenoo W, Kasitanon N, Morand E, Kandane-Rathnayake R. Associations between physicians' global assessment of disease activity and patient-reported outcomes in patients with systemic lupus erythematosus: a longitudinal study. *Lupus* 2021;30:1586–95.
42. Kiani A, Strand V, Fang H, Jaranilla J, Petri M. Predictors of self-reported health-related quality of life in systemic lupus erythematosus. *Rheumatology (Oxford)* 2013;52:1651–7.
43. Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact. *Nat Rev Rheumatol* 2016;12:605–20.
44. Elera-Fitzcarrald C, Fuentes A, González LA *et al.* Factors affecting quality of life in patients with systemic lupus erythematosus: important considerations and potential interventions. *Expert Rev Clin Immunol* 2018;14:915–31.
45. Daleboudt GM, Berger SP, Broadbent E, Kaptein AA. Health-related quality of life in patients with systemic lupus erythematosus and proliferative lupus nephritis. *Psychol Health Med* 2011;16:393–404.
46. Hua N, Gomez A, Lindblom J *et al.* Sensitivity analysis of EQ-5D-3L index scores in terms of discriminative and known-groups validity in SLE: introducing Adequate Health State. *Rheumatology (Oxford)* 2023;62:3916–23.
47. Lindblom J, Zetterberg S, Emamikia S *et al.* EQ-5D full health state after therapy heralds reduced hazard to accrue subsequent organ damage in systemic lupus erythematosus. *Front Med* 2022;9:1092325.
48. Wang C, Mayo NE, Fortin PR. The relationship between health related quality of life and disease activity and damage in systemic lupus erythematosus. *J Rheumatol* 2001;28:525–32.
49. van Vollenhoven RF, Mosca M, Bertsias G *et al.* Treat-to-target in systemic lupus erythematosus: recommendations from an international task force. *Ann Rheum Dis* 2014;73:958–67.
50. Machado Escobar MA, Yacuzzi MS, Martinez RN *et al.* Validation of an Argentine version of Lupus Quality of Life questionnaire. *Lupus* 2016;25:1615–22.