CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA

PORTO

# MACHINE LEARNING ALGORITHMS FOR PREDICTING EQ-5D FULL HEALTH STATE IN SYSTEMIC LUPUS ERYTHEMATOSUS

by

João Ricardo da Costa Monteiro Botto

December 2023

CATÓLICA
ESCOLA SUPERIOR DE BIOTECNOLOGIA
PORTO

**MACHINE LEARNING ALGORITHMS FOR PREDICTING EQ-5D FULL HEALTH STATE IN SYSTEMIC LUPUS ERYTHEMATOSUS**

Thesis presented to *Escola Superior de Biotecnologia* of the *Universidade Católica Portuguesa* to fulfill the requirements of Master of Science degree in Biomedical Engineering

by

João Ricardo da Costa Monteiro Botto

Supervisor: Prof. Ioannis Parodis, MD, PhD

Center for Molecular Medicine, Karolinska Institutet

Tutor: Prof. Ana Leite Oliveira, PhD

Escola Superior de Biotecnologia, Universidade Católica Portuguesa

December 2023

# Abstract

**Objectives:** To determine factors associated with EQ-5D full health state (FHS) in systemic lupus erythematosus (SLE) before and after a trial intervention, resorting to machine learning algorithms.

**Methods:** We conducted a post-hoc analysis of data from two phase III clinical trials of belimumab (BLISS-52, BLISS-76). Demographic, laboratory, and clinical features were retrieved, and the Monte Carlo Feature Selection algorithm was employed, further refined upon consideration of collinearity and clinical relevance/expertise. The models used were support vector machine with radial basis function kernel (SVMRadial), least absolute shrinkage and selection operator (LASSO), neural network (NNet), and logistic regression (LR).

**Results:** In a cohort of 1642 SLE patients, 12.9% reported FHS at baseline and 23.1% at week 52. Selected features were age, sex, Asian ancestry, baseline clinical Systemic Lupus Erythematosus Disease Activity Index-2000, Safety of Estrogens in Lupus National Assessment-SLEDAI Physician Global Assessment, and urine protein/creatinine ratio (UPCR), and baseline EQ-5D utility index score (week-52 models only). The models predicting FHS demonstrated comparable performance at baseline and week 52, where for baseline a maximum area under the curve of 0.73 was seen for the LASSO and LR models, versus a 0.77 maximum for the week-52 LASSO and NNet models. Particularly high for all models was the negative predictive value (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. Baseline EQ-5D utility index constituted the most informative feature for predicting FHS experience at week 52.



**Keywords:** systemic lupus erythematosus; quality of life; EQ-5D; machine learning.


**Key messages:**

- Older age, female sex, non-Asian ancestry, high disease activity, and low UPCR predicted non-FHS experience.
- Baseline EQ-5D utility index score was the main predictor of FHS experience at week 52.

# Resumo

**Objetivos:** Determinar fatores associados ao estado de saúde perfeita (ESP) do EQ-5D no lúpus eritematoso sistémico (LES) antes e depois da intervenção num ensaio clínico, recorrendo a algoritmos de *machine learning*.

**Métodos:** Realizámos uma análise *post-hoc* de dados de dois ensaios clínicos de fase III do belimumab (BLISS-52, BLISS-76). Foram recolhidas variáveis demográficas, laboratoriais e clínicas, tendo sido utilizado o algoritmo *Monte Carlo Feature Selection*, posteriormente refinado considerando colinearidade e relevância/perícia clínica. Os seguintes modelos foram usados: *support vector machine with radial basis function kernel* (SVMRadial), *least absolute shrinkage and selection operator* (LASSO), *neural network* (NNet), e *logistic regression* (LR).

**Resultados:** Numa coorte de 1642 doentes com LES, 12.9% reportaram FHS na *baseline* e 23.1% na semana 52. As variáveis selecionadas foram idade, sexo, ascendência Asiática, e *clinical Systemic Lupus Erythematosus Disease Activity Index-2000*, *Safety of Estrogens in Lupus National Assessment-SLEDAI Physician Global Assessment*, e rácio proteína/creatinina urinário (RPCU) na *baseline*. Exclusivamente para os modelos da semana 52, o índice do EQ-5D na *baseline* também foi selecionado. Os modelos preditores do ESP demonstraram um desempenho comparável na *baseline* e na semana 52. Na *baseline*, a máxima área sob a curva foi vista nos modelos LASSO e LR (0.73), enquanto na semana 52 foi vista nos modelos LASSO e NNet (0.77). O valor preditivo negativo foi particularmente alto para todos os modelos (0.88–0.94). A calibração mostrou uma ligeira melhoria nos modelos da semana 52.

**Conclusão:** Idade avançada, sexo feminino, ascendência não-Asiática, alta atividade de doença, e baixo RPCU foram associados, através de *machine learning*, à não-experiência de um ESP em doentes com LES na *baseline* e na semana 52. O índice do EQ-5D na *baseline* constituiu a variável mais informativa para prever a experiência de um ESP na semana 52.

**Palavras-chave:** lúpus eritematoso sistémico; qualidade de vida; EQ-5D; *machine learning*.

**Mensagens-chave:**

- Idade avançada, sexo feminino, ascendência não-Asiática, alta atividade de doença, e baixo RPCU previram a não-experiência de um ESP.
- O índice do EQ-5D na *baseline* foi o principal preditor da experiência de um ESP na semana 52.

# Acknowledgements

To **Professor Ioannis Parodis**,

For the opportunity to develop my thesis at *Karolinska Institutet*,

For the immensurable and unconditional support and guidance throughout all this time,

For the kindness;

To **Professor Ana Leite Oliveira**,

For all the support, encouragement, and kindness shown throughout the whole Master's;

To **Nurşen**,

For all the unconditional help, encouragement, support, kindness, and friendship,

For being who she is – amazing;

To **Jesper**,

For being my rock, and making everything so much more bearable,

For making me smile even when I don't feel like smiling,

For being an incredible boyfriend;

To my **Family**, to my **Friends**, to my **Colleagues**,

For being my lever, and/or having made me become who I am today;

My deepest and honest

THANK YOU!

*“Para ser grande, sê inteiro: nada*

*Teu exagera ou exclui.*

*Sê todo em cada coisa. Põe quanto és*

*No mínimo que fazes.*

*Assim em cada lago a lua toda*

*Brilha, porque alta vive.”*

Fernando Pessoa, in *Odes de Ricardo Reis*

# Contents

# List of Figures

# List of Tables

## List of Abbreviations

| | |
|---|---|
| Acc. | Accuracy |
| AUC | Area under the curve |
| BMI | Body mass index |
| CI | Confidence interval |
| cSLEDAI-2K | Clinical Systemic Lupus Erythematosus Disease Activity Index-2000 |
| CV | Cross-validation |
| EQ-5D-3L | EQ-5D 3-level version |
| FHS | Full health state |
| HRQoL | Health-related quality of life |
| IQR | Interquartile range |
| LASSO | Least absolute shrinkage and selection operator |
| LR | Logistic regression |
| MCFS | Monte Carlo Feature Selection |
| NNet | Neural network |
| NPV | Negative predictive value |
| PGA | Physician Global Assessment |
| PPV | Positive predictive value |
| ROC | Receiver operating characteristic |
| SD | Standard deviation |
| SDI | Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index |
| SELENA-SLEDAI | Safety of Estrogens in Lupus National Assessment – Systemic Lupus Erythematosus Disease Activity Index |
| SLE | Systemic lupus erythematosus |
| SLEDAI | Systemic Lupus Erythematosus Disease Activity Index |
| SLEDAI-2K | Systemic Lupus Erythematosus Disease Activity Index-2000 |
| ST | Standard therapy |
| SVMRadial | Support vector machine with radial basis function kernel |
| UPCR | Urine protein to creatinine ratio |
| VAS | Visual analogue scale |

# Machine learning algorithms for predicting EQ-5D full health state in systemic lupus erythematosus

**João Botto[1,2], Nursen Cetrez[1,2], Julius Lindblom[1,2], Ioannis Parodis[1,2,3,*]**

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

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

[3]Department of Rheumatology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

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

CHAPTER 1

# INTRODUCTION

Systemic lupus erythematosus (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, the patients' health-related quality of life (HRQoL) constitutes an important parameter, which is highly relevant to the patients. Patients with SLE often experience significantly lower HRQoL compared to healthy individuals [3], and to those with other chronic diseases [4]. Thus, researchers and clinicians have strived over the years to understand the impact of SLE on this parameter, and the factors that are associated with its variations [3, 4]. This understanding is crucial, in the sense that one of the ultimate goals in providing care is to improve the patients' HRQoL, and 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 EQ-5D health questionnaire is a widely used tool for measuring HRQoL [9]. Its psychometric properties have been validated for patients with SLE [10]. EQ-5D consists of a visual analogue scale (VAS), where the patients mark their self-assessed health state on a scale ranging from 0 (worst possible health state) to 100 (best possible health state), and a descriptive health evaluation part stratified into five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each of these dimensions is represented by one question, and respondents rate their health state in each question using three levels of severity (in the EQ-5D-3L version): no problems (score of 1), some problems (score of 2), and extreme problems (score of 3). The five scores are put together to form an individual health profile, and are also converted into a single utility index score (reflecting an individual's overall health status), through means of a population-specific scoring algorithm. This overall score can range from <0 (where 0 represents experience equal to death, yet allowing for health states being experienced as worse than death) to 1. When a patient scores “no problem” (score of 1) in all five dimensions (hence, yielding a utility index score of 1), the experience corresponds to an idealised state of complete health, which is termed full health state (FHS) and serves as a reference point against which the impact of disease and interventions can be evaluated [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 EQ-5D FHS. Identification of variables with the ability to predict attainment of EQ-5D FHS in SLE patients could provide clinicians with more realistic expectations, and with valuable insights for individualised patient care and optimisation of the allocation of healthcare resources.

In a recent study, we reported an EQ-5D FHS prevalence of 23.0% among SLE patients from a clinical trial setting at week 52 from baseline, and disclosed factors such as female sex, increasing body mass index (BMI), and non-Asian ethnicity to be negatively associated with the achievement of EQ-5D FHS at week 52 [13]. The present study aimed to determine factors that are associated with 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 the time of commencement of the trial intervention. This paper pioneers the application of machine learning algorithms for addressing this question, which constitutes a novel approach in this context.

CHAPTER 2

# MATERIALS AND METHODS

## 2.1 Study population

Data analysis was conducted using information from two phase III 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 belonging to the tumour necrosis factor family (BAFF), in patients with active SLE, i.e., 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 American College of Rheumatology (ACR) [16], were adults, and had an antinuclear antibody (ANA) titre of ≥1:80 and/or a serum anti-double stranded (ds)DNA antibody level of ≥30 IU/mL during the screening phase. Additionally, they exhibited a Safety of Estrogens in Lupus National Assessment – Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI) [17] score of ≥6 during the screening phase. All participants had been on a stable non-biological standard therapy (ST) for at least 30 days prior to the trial baseline. Patients were not eligible for participation if they had an active severe lupus nephritis or central nervous system involvement.

The study participants were randomly assigned to one of three arms: belimumab 1 mg/kg per infusion, belimumab 10 mg/kg per infusion, or placebo, administered intravenously, on top of ST. 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 GlaxoSmithKline (Uxbridge, UK) through the Clinical Study Data Request consortium. Owing to missing data for variables of interest for the present investigation, 42 patients were excluded, resulting in a study population of 1642 SLE patients.

## 2.2 Ethics

Before enrolment, all study participants provided written informed consent. The trial protocols underwent thorough evaluation and received approval from regional ethics review boards across all participating centres. The protocol for 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.

## 2.3 Clinical definitions

The Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) evaluates overall SLE activity using 24 descriptors, each assigned a specific weight based on its relative significance [18]. A modified iteration of SLEDAI is SLEDAI-2000 (SLEDAI-2K), which 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 levels) [20], was the one employed for assessing clinical disease activity in the present investigation.

The SELENA-SLEDAI Physician Global Assessment (PGA) gauges SLE overall activity using a VAS, which ranges from 0 (none) to 3 (severe) [21].

Lastly, the Systemic Lupus International Collaborating Clinics/American College of Rheumatology (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. SDI comprises 39 items, grouped into 12 organ systems, with a theoretical maximum score of 47 points [22].

## 2.4 Study outcome

The outcome in this study was patient-reported FHS (versus non-FHS) assessed using the EQ-5D-3L instrument at baseline and at week 52.

## 2.5 Data collection

Demographic parameters, baseline clinical and laboratory features, as well as concomitant medications, were retrieved. EQ-5D data were retrieved from the baseline and week-52 visits. EQ-5D utility index scores were calculated based on the United States (US) population scoring algorithm.

## 2.6 Feature selection and dataset split

Machine learning models were developed with the goal of predicting patient-reported EQ-5D FHS at two timepoints, i.e., 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. 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 utility index scores in the algorithm for the latter models.

With MCFS, a prioritised 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 [23]. Next, further refinement of the feature selection was made based on collinearity, clinical expertise and/or clinical relevance, only selecting features that could be used both for baseline and week-52 models (with the exception mentioned above).

Patients with incomplete data on the selected features and outcome variables were removed from the datasets during subsequent analysis. Upon finalisation 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 [24]. 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 [25].

## 2.7 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). Predictions were classified as positive (i.e., EQ-5D FHS experience) if their respective probabilities were $\geq 0.5$. 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 [24]. 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 characteristic (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 [26], while the others were assessed through the caret R package [24]. 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) [27], scikit-learn (version 1.2.1) calibration module [28], NumPy (version 1.23.5) [29], SciPy (version 1.10.0) interpolate module [30], and Matplotlib (version 3.7.0) pyplot module [31].

## 2.8 Statistical analysis

For normal distributions, descriptive statistics are indicated as the number (percentage) or the mean (standard deviation), while for non-normal distributions, data are reported as the median (interquartile range). Descriptive statistics and comparisons between patients who experienced EQ-5D FHS and patients who did not were conducted using the tableone R package, version 0.13.2 [32]. Differences yielding $p$-values <0.05 were considered statistically significant.

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

CHAPTER 3

# RESULTS

## 3.1 Patient characteristics and subgroup comparisons

**Table 3.1** details demographic and baseline clinical characteristics for the overall study population, as well as comparisons between patients who reported EQ-5D FHS and patients who did not at baseline and at week 52. 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 at week 52) was more enriched in patients of Asian ancestry, while the non-FHS subgroup was more enriched in women and patients of older age with higher baseline BMI, SELENA-SLEDAI PGA, SDI and cSLEDAI-2K scores, and lower baseline urine protein to creatinine ratio (UPCR). Patients in the FHS subgroup at week 52 had reported greater EQ-5D utility index scores at baseline.

## 3.2 Feature selection

Results from the MCFS algorithm for baseline and week 52 are displayed in **Supplementary Figure A.1**. 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 assessments. 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 latter models.

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 the features mentioned above, EQ-5D utility index scores were also included (**Supplementary Table A.1**).

**Table 3.1.** Patient demographic and baseline clinical features, comparing the EQ-5D FHS and non-FHS subgroups at baseline and at week 52.

| | Overall | | Baseline FHS | | Baseline Non-FHS | | | Week 52 FHS | | Week 52 Non-FHS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (*n*=1642) | | (*n*=212) | | (*n*=1430) | | | (*n*=380) | | (*n*=1262) | | |
| | ***n*** | **(%)** | ***n*** | **(%)** | ***n*** | **(%)** | ***p*-value** | ***n*** | **(%)** | ***n*** | **(%)** | ***p*-value** |
| **Female sex** | 1544 | (94.0) | 192 | (90.6) | 1352 | (94.5) | **0.033** | 347 | (91.3) | 1197 | (94.8) | **0.015** |
| **Ancestry** | | | | | | | | | | | | |
| **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** | 694 | (42.3) | 70 | (33.0) | 624 | (43.7) | **0.004** | 111 | (29.2) | 583 | (46.2) | **<0.001** |
| | **mean** | **(SD)** | **mean** | **(SD)** | **mean** | **(SD)** | | **mean** | **(SD)** | **mean** | **(SD)** | |
| **Age**, years | 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** | 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 utility index score** | 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** | 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] | 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** | 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** | 7.31 | (3.63) | 6.21 | (3.65) | 7.47 | (3.60) | **<0.001** | 6.66 | (3.84) | 7.51 | (3.54) | **<0.001** |
| | **median** | **(IQR)** | **median** | **(IQR)** | **median** | **(IQR)** | | **mean** | **(SD)** | **mean** | **(SD)** | |
| **SLE disease duration**, years | 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 | 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** |

The bold characters in the *p*-value column refer to statistically significant differences ($p<0.05$).

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

BMI: body mass index; cSLEDAI-2K: clinical Systemic Lupus Erythematosus Disease Activity Index 2000; FHS: full health state; IQR: interquartile range; NA: not applicable; SELENA-SLEDAI PGA: Safety of Estrogens in Lupus National Assessment – Systemic Lupus Erythematosus Disease Activity Index Physician Global Assessment; SD: standard deviation; SDI: Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) Damage Index; SLE: Systemic Lupus Erythematosus; SLEDAI-2K: Systemic Lupus Erythematosus Disease Activity Index 2000; UPCR: urine protein to creatinine ratio.

## 3.3 Training and test sets

Each training set contained 80% ($n$=1314) of the total dataset, while 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 timepoint (**Supplementary Table A.1**). This suggests a similar feature distribution, and indicates that the models were trained on datasets that are representative of the data encountered during testing.

## 3.4 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.

Concerning the test sets, the performance metrics for the baseline and week-52 models are presented in **Table 3.2**. 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 versus 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 ranges: 0.52–0.71 for baseline and 0.67–0.72 for week 52; PPV ranges: 0.21–0.25 for baseline and 0.39–0.44 for week 52). Regarding NPVs, these were particularly high for all models (range: 0.88–0.94). The ROC curves from the models at both timepoints are shown in **Figure 3.1**, and the confusion matrices from all individual models are presented in **Supplementary Tables A.2–A.9**. Calibration curves, displayed in **Figure 3.1**, show that all models exhibited poor calibration results at both timepoints, overestimating their predictions. Nevertheless, the week-52 models were slightly better calibrated, especially concerning greater predicted probabilities.

**Table 3.2.** Performance of baseline and week-52 models on the test sets.

| | Baseline | | | | Week 52 | | | |
|---|---|---|---|---|---|---|---|---|
| | **SVMRadial** | **LASSO** | **NNet** | **LR** | **SVMRadial** | **LASSO** | **NNet** | **LR** |
| **Acc. [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 |

Acc.: accuracy; AUC: area under the curve; CI: confidence interval; LASSO: least absolute shrinkage and selection operator; LR: logistic regression; NNet: neural network; PPV: positive predictive value; NPV: negative predictive value; SVMRadial: support vector machine with radial basis function kernel.

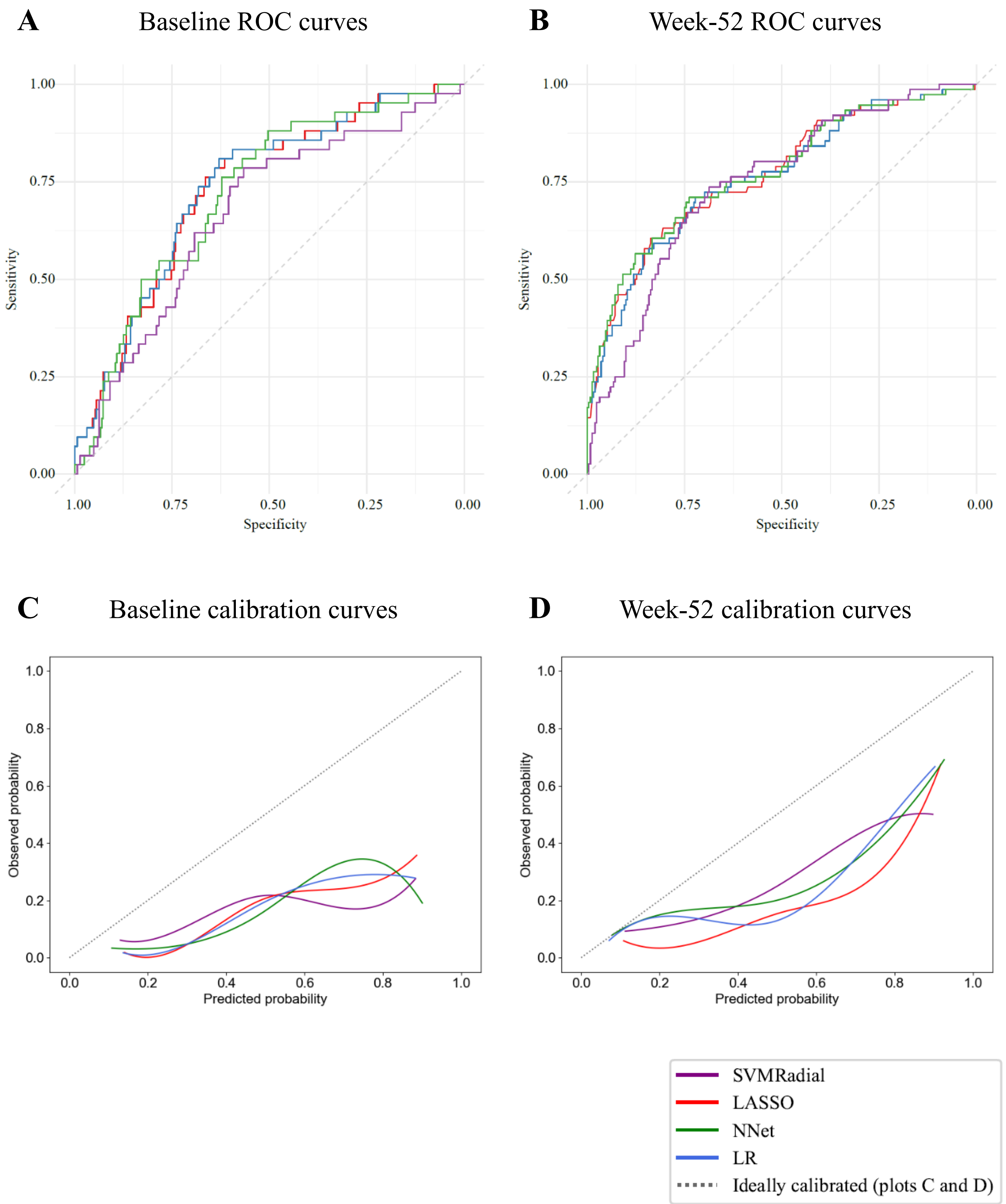



**Figure 3.1.** Receiver operating characteristic (ROC) curves on the test datasets from the baseline (**A**) and week-52 (**B**) models, and smoothed calibration curves on the same datasets, from the baseline (**C**) and week-52 (**D**) models.

LASSO: least absolute shrinkage and selection operator; LR: logistic regression; NNet: neural network; SVMRadial: support vector machine with radial basis function kernel.

CHAPTER 4

# DISCUSSION

The main purpose of this post-hoc analysis of the BLISS-52 and BLISS-76 clinical trials was to identify factors that are associated with EQ-5D FHS experience or lack of FHS experience at baseline and 52 weeks after commencement of a therapeutic intervention with belimumab or placebo on top of ST, using machine learning algorithms. Following feature selection by means of the MCFS algorithm, mitigation was attempted based on collinearity, clinical relevance, and expertise. 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 EQ-5D FHS experience both at baseline and after the trial intervention, while low baseline EQ-5D utility index scores constituted the most important determinant of failure to experience EQ-5D FHS at week 52.

Relating to sex, our findings are in line with previous studies demonstrating women with SLE to report worse HRQoL compared with men with SLE [33-35]. The relationships between age or disease activity and HRQoL have not been consistent in previous SLE literature [33-38]; we herein found that advancing age and greater clinical disease activity were negatively associated with EQ-5D FHS experience. Discrepancies in literature may be multifactorial, spanning from differences in the relationships across different patient populations to the large variation in the instruments used to measure disease activity, as well as in the instruments used to estimate HRQoL. The most common instrument for estimating HRQoL has probably been the Medical Outcomes Study Short-Form 36 (SF-36) [38], with only few studies hitherto having used the EQ-5D in SLE populations [13, 37]. Despite the large interrater variability observed for PGA scoring, advancing PGA scores have consistently been negatively associated with HRQoL in patients with SLE [39-41], which highlights that PGA captures features of disease activity beyond the span that is captured by the SLEDAI.

Although patients with SLE of Asian origin are known to have a more severe disease burden compared with White/Caucasian patients [42], we herein demonstrated that Asian patients more frequently reported EQ-5D FHS compared to the non-Asian population. In fact, this is in line with previous literature that has documented a tendency of Asian people with SLE to report better HRQoL [43].

There is limited information in previous studies regarding the relationship between renal involvement and SLE patients' HRQoL. Daleboudt et al. reported a moderate inverse correlation between proteinuria and HRQoL in a study population of proliferative lupus nephritis, in which substantial reductions in proteinuria levels after successful treatment were reported [44]. We herein report that greater UPCR levels were associated with EQ-5D FHS experience, both at baseline and at week 52, which contrasts with that previous finding by Daleboudt et al. and seemingly constitutes a counterintuitive finding. Importantly, the BLISS trials excluded patients with active severe lupus nephritis during the screening phase, resulting in relatively low proteinuria levels in the overall study population, thus limiting the generalisability of our findings to SLE populations with more active renal disease. Nonetheless, our observation contrasts with what one would consider an expected finding and has therefore 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 the renal manifestations in people with SLE with frequent urinalysis and creatinine levels.

Last but not least, baseline EQ-5D utility index scores constituted in our study the most important feature in the week-52 models, with greater scores predicting 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 ultimate goal of machine learning models is to identify features with predictive value that eventually could be applied in routine clinical practice. To this end, refinement of the feature selection was needed accounting for collinearity and clinical relevance. Well-performing and easy-to-apply models for predicting EQ-5D FHS experience in patients with SLE would be useful, especially since EQ-5D assessment has shown ability to distinguish responders from non-responders to therapy [13, 45], as well as predict long-term outcome, with FHS experience being linked to less and/or slower organ damage accrual [46]. Furthermore, EQ-5D has demonstrated favourable psychometric properties in terms of validity and reliability [47], it is easy to fill in by the patients (yielding a high rate of acceptability and completion), and it is simple to implement in clinical practice and research [47, 48]. For all these reasons, HRQoL reported through EQ-5D may be considered a useful tool among currently used patient-reported outcome measures in SLE [45, 46], acknowledging that it provides important complemental information to current definitions of remission and low disease activity which 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 the test sets, indicating absence of overfitting issues. Moreover, AUCs from all models on the training sets were comprised of reasonable values, indicating absence of underfitting as well. It is important to underscore that our models yielded very high NPVs, around 90%, ensuring robust predictions of which patients are not likely to experience EQ-5D FHS. By contrast, the downside of the models lies on the relatively low PPVs and the calibration curves, where a likely overestimation of predictions is noticed, with the week-52 models performing however slightly better. This could be explained by how we addressed the low number of positive training examples, an issue known as class skew or class imbalance. By employing a common technique known as oversampling, we inevitably introduced some disadvantages. 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, which 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, socioeconomic status, and daily physical activity, features that are likely to contribute to SLE patients' HRQoL experience. 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 generalisability of our findings is hindered by the exclusion of patients with active severe lupus nephritis or neuropsychiatric lupus. Nonetheless, the study also has some important strengths, including the large, and diverse in ancestral terms, 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, in particular EQ-5D FHS experience.

CHAPTER 5

# CONCLUSIONS AND FUTURE WORK

Through a machine learning approach, we determined factors associated with 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 (by both cSLEDAI-2K and SELENA-SLEDAI PGA), and low UPCR levels to be negatively associated with EQ-5D FHS experience. Baseline EQ-5D utility index scores constituted the most important predictor of EQ-5D FHS experience at week 52. Our study contributes insights towards person-centred care, acknowledging patient-reported health experience as an integral part of disease and treatment evaluation procedures.

Machado Escobar et al. suggested that an EQ-5D index score of ≥0.739 is indicative of good HRQoL among SLE patients [50], and a recent study by Hua et al. advocated that an EQ-5D index score of ≥0.800 distinguishes responders to treatment from non-responders [45]. Thus, an analysis focusing on predictive factors for achieving EQ-5D health states below FHS could offer valuable insights. Application of machine learning models in such a study would potentially translate to a greater number of positive training examples, which could increase the performance of the models.

CHAPTER 6

# STATEMENTS

## 6.1 Acknowledgements

The authors wish to express their gratitude to GlaxoSmithKline (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.

## 6.2 Authors contributions

Conception and design of the work: J.B., N.C., I.P. Acquisition of data: N.C., J.L., I.P. Statistical analysis and interpretation of data: J.B., N.C., I.P. Original draft: J.B., I.P. Critical revision of the manuscript for important intellectual content: all authors. All authors reviewed and approved the final version of the manuscript prior to submission, and agree to be accountable for all aspects of the work.

## 6.3 Funding

This work was supported by the GlaxoSmithKline Investigator-Sponsored Studies (ISS) programme 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, analyses or interpretation of data, the writing of the manuscript or the decision to submit it for publication.

## 6.4 Conflicts of interest

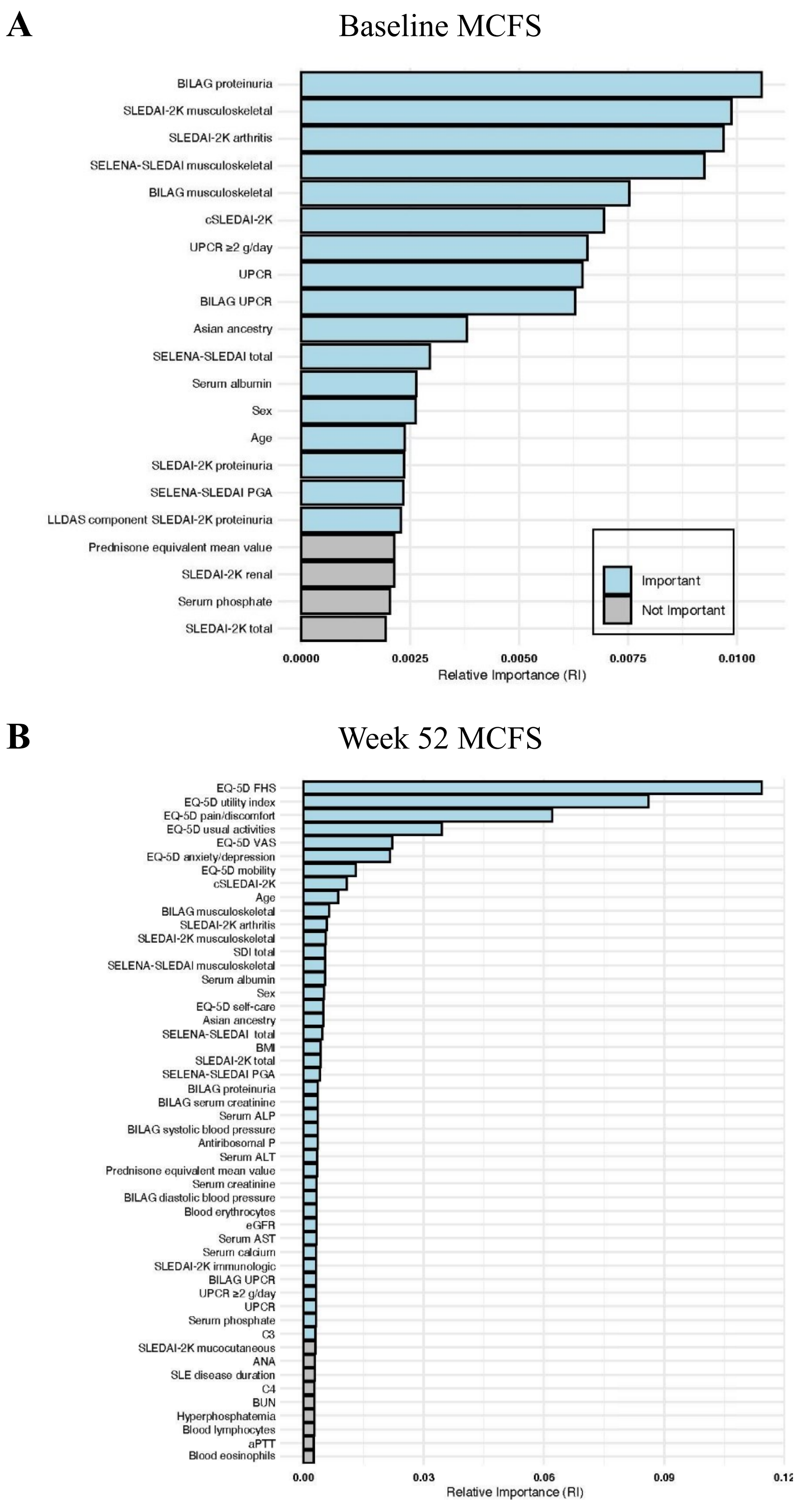
I.P. has received research funding and/or honoraria from the EuroQol Research Foundation, Amgen, AstraZeneca, Aurinia, Bristol-Myers Squibb, Elli Lilly, Gilead, GlaxoSmithKline, Janssen, Novartis, Otsuka and Roche. The other authors declare no conflicts of interest.

## 6.5 Data availability

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

APPENDIX A

# SUPPLEMENTARY FIGURE AND SUPPLEMENTARY TABLES




**Supplementary Figure A.1.** Monte Carlo Feature Selection (MCFS) plots for baseline (**A**) and week 52 (**B**).

ALP: alkaline phosphatase; ALT: alanine transaminase; ANA: antinuclear antibody; aPTT: activated partial thromboplastin time; AST: aspartate transferase; BILAG: British Isles Lupus Assessment Group; BMI: body mass index; BUN: blood urea nitrogen; C3: complement component 3; C4: complement component 4; cSLEDAI-2K: clinical Systemic Lupus Erythematosus Disease Activity Index 2000; eGFR: estimated glomerular filtration rate; FHS: full health state; LLDAS: Lupus Low Disease Activity State; PGA: Physician Global Assessment; SDI: Systemic Lupus International Collaborating Clinics (SLICC)/American College of Rheumatology (ACR) Damage Index; SELENA-SLEDAI: Safety of Estrogens in Lupus National Assessment – Systemic Lupus Erythematosus Disease Activity Index; SLE: Systemic Lupus Erythematosus; SLEDAI-2K: Systemic Lupus Erythematosus Disease Activity Index 2000; UPCR: urinary protein to creatinine ratio; VAS: visual analogue scale.

**Supplementary Table A.1.** Final features in the training and test sets (for the baseline and week-52 models).

| | Overall set (*n*=1642) | | Baseline: Training set (*n*=1314) | | Baseline: Test set (*n*=328) | | Baseline | Week 52: Training set (*n*=1314) | | Week 52: Test set (*n*=328) | | Week 52 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***n*** | **(%)** | ***n*** | **(%)** | ***n*** | **(%)** | ***p*-value** | ***n*** | **(%)** | ***n*** | **(%)** | ***p*-value** |
| **Female sex** | 1544 | (94.0) | 1238 | (94.2) | 306 | (93.3) | 0.616 | 1242 | (94.5) | 302 | (92.1) | 0.123 |
| **Asian ancestry** | 332 | (20.2) | 259 | (19.7) | 73 | (22.3) | 0.342 | 261 | (19.9) | 71 | (21.6) | 0.521 |
| | **mean** | **(SD)** | **mean** | **(SD)** | **mean** | **(SD)** | | **mean** | **(SD)** | **mean** | **(SD)** | |
| **Age**, years | 37.78 | (11.52) | 37.97 | (11.47) | 36.99 | (11.70) | 0.168 | 37.59 | (11.45) | 38.51 | (11.81) | 0.196 |
| **EQ-5D utility index score** | 0.74 | (0.19) | *NA* | *NA* | *NA* | *NA* | *NA* | 0.74 | (0.19) | 0.75 | (0.19) | 0.355 |
| **SELENA-SLEDAI PGA** | 1.42 | (0.48) | 1.43 | (0.48) | 1.40 | (0.49) | 0.398 | 1.42 | (0.49) | 1.44 | (0.47) | 0.440 |
| **cSLEDAI-2K scores** | 7.31 | (3.63) | 7.30 | (3.66) | 7.35 | (3.51) | 0.839 | 7.30 | (3.53) | 7.35 | (4.01) | 0.826 |
| | **median** | **(IQR)** | **median** | **(IQR)** | **median** | **(IQR)** | | **median** | **(IQR)** | **median** | **(IQR)** | |
| **UPCR**, mg/mg | 0.15 | (0.09–0.36) | 0.14 | (0.09–0.36) | 0.16 | (0.09–0.40) | 0.438 | 0.15 | (0.09–0.38) | 0.15 | (0.09–0.32) | 0.934 |

cSLEDAI-2K: clinical Systemic Lupus Erythematosus Disease Activity Index 2000; IQR: interquartile range; NA: not applicable; PGA: Physician Global Assessment; SD: standard deviation; UPCR: urine protein to creatinine ratio.

**Supplementary Table A.2.** Confusion matrix for the baseline SVMRadial model.

| | | Observed number of patients | |
|---|---|---|---|
| | | FHS | Non-FHS |
| Predicted number of patients | FHS | 22 | 81 |
| | Non-FHS | 20 | 205 |

FHS: full health state; SVMRadial: support vector machine with radial basis function kernel.

**Supplementary Table A.3.** Confusion matrix for the baseline LASSO model.

| | | Observed number of patients | |
|---|---|---|---|
| | | FHS | Non-FHS |
| Predicted number of patients | FHS | 30 | 90 |
| | Non-FHS | 12 | 196 |

FHS: full health state; LASSO: least absolute shrinkage and selection operator.

**Supplementary Table A.4.** Confusion matrix for the baseline NNet model.

| | | Observed number of patients | |
|---|---|---|---|
| | | FHS | Non-FHS |
| Predicted number of patients | FHS | 23 | 84 |
| | Non-FHS | 19 | 202 |

FHS: full health state; NNet: neural network.

**Supplementary Table A.5.** Confusion matrix for the baseline LR model.

| | | Observed number of patients | |
|---|---|---|---|
| | | FHS | Non-FHS |
| Predicted number of patients | FHS | 29 | 89 |
| | Non-FHS | 13 | 197 |

FHS: full health state; LR: logistic regression.

**Supplementary Table A.6.** Confusion matrix for the week-52 SVMRadial model.

| | | Observed number of patients | |
|---|---|---|---|
| | | **FHS** | **Non-FHS** |
| **Predicted number of patients** | **FHS** | 51 | 67 |
| | **Non-FHS** | 25 | 185 |

FHS: full health state; SVMRadial: support vector machine with radial basis function kernel.

**Supplementary Table A.7.** Confusion matrix for the week-52 LASSO model.

| | | Observed number of patients | |
|---|---|---|---|
| | | **FHS** | **Non-FHS** |
| **Predicted number of patients** | **FHS** | 55 | 87 |
| | **Non-FHS** | 21 | 165 |

FHS: full health state; LASSO: least absolute shrinkage and selection operator.

**Supplementary Table A.8.** Confusion matrix for the week-52 NNet model.

| | | Observed number of patients | |
|---|---|---|---|
| | | **FHS** | **Non-FHS** |
| **Predicted number of patients** | **FHS** | 54 | 69 |
| | **Non-FHS** | 22 | 183 |

FHS: full health state; NNet: neural network.

**Supplementary Table A.9.** Confusion matrix for the week-52 LR model.

| | | Observed number of patients | |
|---|---|---|---|
| | | **FHS** | **Non-FHS** |
| **Predicted number of patients** | **FHS** | 55 | 82 |
| | **Non-FHS** | 21 | 170 |

FHS: full health state; LR: logistic regression.

# References


[1] Barber MRW, Drenkard C, Falasinnu T, Hoi A, Mak A, Kow NY, et al. Global epidemiology of systemic lupus erythematosus. Nature Reviews Rheumatology. 2021;17(9):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 Science & Medicine. 2022;9(1):e000640.

[3] McElhone K, Abbott J, Teh LS. A review of health related quality of life in systemic lupus erythematosus. Lupus. 2006;15(10):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(9):1706-8.

[5] Gomez A, Qiu V, Cederlund A, Borg A, Lindblom J, Emamikia S, et al. Adverse Health-Related Quality of Life Outcome Despite Adequate Clinical Response to Treatment in Systemic Lupus Erythematosus. Frontiers in Medicine. 2021;8.

[6] Annapureddy N, Devilliers H, Jolly M. Patient-reported outcomes in lupus clinical trials with biologics. Lupus. 2016;25(10):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(8):660-70.

[8] Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Endpoints: consensus recommendations from OMERACT IV. Outcome Measures in Rheumatology. Lupus. 2000;9(5):322-7.

[9] Gusi N, Olivares PR, Rajendram R. The EQ-5D Health-Related Quality of Life Questionnaire. In: Preedy VR, Watson RR, editors. Handbook of Disease Burdens and Quality of Life Measures. New York, NY: Springer New York; 2010. p. 87-99.

[10] Aggarwal R, Wilke CT, Pickard AS, Vats V, Mikolaitis R, Fogg L, et al. Psychometric properties of the EuroQol-5D and Short Form-6D in patients with systemic lupus erythematosus. J Rheumatol. 2009;36(6):1209-16.

[11] Van Reenen M, Oppe M, Secnik Boye K, Herdman M, Kennedy-Martin M, Kennedy-Martin T. EuroQol research foundation. EQ-5D-3L User Guide. 2018. Available online at: https://euroqol.org/publications/user-guides (accessed September 12, 2023).

[12] Athanassiou P, Athanassiou L. Current Treatment Approach, Emerging Therapies and New Horizons in Systemic Lupus Erythematosus. Life [Internet]. 2023; 13(7).

[13] Lindblom J, Gomez A, Borg A, Emamikia S, Ladakis D, Matilla J, et al. EQ-5D-3L full health state discriminates between drug and placebo in clinical trials of systemic lupus erythematosus. Rheumatology (Oxford). 2021;60(10):4703-16.

[14] Navarra SV, Guzmán RM, Gallacher AE, Hall S, Levy RA, Jimenez RE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. Lancet. 2011;377(9767):721-31.

[15] Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzová D, 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(12):3918-30.

[16] Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40(9):1725.

[17] Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med. 2005;353(24):2550-8.

[18] Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang CH. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum. 1992;35(6):630-40.

[19] Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol. 2002;29(2):288-91.

[20] Uribe AG, Vilá LM, McGwin G, Jr., Sanchez ML, Reveille JD, Alarcón GS. 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(10):1934-40.

[21] Elisabetta C, Matteo P, Alberto F, Hervé D, Alberto C, Laurent A. Use of Physician Global Assessment (PGA) in Systemic lupus erythematosus: a systematic review of its psychometric properties. medRxiv. 2020:2020.04.14.20064683.

[22] Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Sanchez-Guerrero J, 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 & Rheumatism. 1996;39(3):363-9.

[23] Dramiński M, Koronacki J. rmcfs: An R Package for Monte Carlo Feature Selection and Interdependency Discovery. Journal of Statistical Software. 2018;85(12):1 - 28.

[24] Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software. 2008;28(5):1 - 26.

[25] Lunardon N, Menardi G, Torelli N. ROSE: a Package for Binary Imbalanced Learning. R Journal. 2014;6:79-89.

[26] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC bioinformatics. 2011;12:77.

[27] McKinney W. Data Structures for Statistical Computing in Python2010. 56-61 p.

[28] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2012;12.

[29] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357-62.

[30] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17(3):261-72.

[31] Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9(3):90-5.

[32] Agapios P, Dimitris M. TableOne: an online web application and R package for summarising and visualising data. Evidence Based Mental Health. 2020;23(3):127.

[33] Gu M, Cheng Q, Wang X, Yuan F, Sam NB, Pan H, 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(3):371-82.

[34] Gaballah NM, El-Najjar AR. Clinical characteristics and health related quality of life (HRQoL) in Egyptian patients with systemic lupus erythematosus. The Egyptian Rheumatologist. 2019;41(2):117-21.

[35] Nyambane E, Genga E, Achieng L, Oyoo O, Otieno FC. P188 Assessment of disease activity and health related quality of life in patients with systemic lupus erythematosus at Kenyatta national hospital. Lupus Science & Medicine. 2020;7(Suppl 1):A121-A2.

[36] 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(1):e000161.

[37] Frederick W, Kaleb M, Tracy LI, Robert S K. EQ-5D and SF-36 Quality of Life Measures in Systemic Lupus Erythematosus: Comparisons with Rheumatoid Arthritis, Noninflammatory Rheumatic Disorders, and Fibromyalgia. The Journal of Rheumatology. 2010;37(2):296.

[38] Schmeding A, Schneider M. Fatigue, health-related quality of life and other patient-reported outcomes in systemic lupus erythematosus. Best Practice & Research Clinical Rheumatology. 2013;27(3):363-75.

[39] Chaigne B, Chizzolini C, Perneger T, Trendelenburg M, Huynh-Do U, Dayer E, 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(1):17.

[40] 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(10):1586-95.

[41] 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, England). 2013;52.

[42] Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact. Nature Reviews Rheumatology. 2016;12(10):605-20.

[43] Elera-Fitzcarrald C, Fuentes A, González LA, Burgos PI, Alarcón GS, Ugarte-Gil MF. Factors affecting quality of life in patients with systemic lupus erythematosus: important considerations and potential interventions. Expert Review of Clinical Immunology. 2018;14(11):915-31.

[44] 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(4):393-404.

[45] Hua N, Gomez A, Lindblom J, Emamikia S, Enman Y, Grannas D, 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. 2023.

[46] Lindblom J, Zetterberg S, Emamikia S, Borg A, von Perner G, Enman Y, et al. EQ-5D full health state after therapy heralds reduced hazard to accrue subsequent organ damage in systemic lupus erythematosus. Frontiers in Medicine. 2022;9.

[47] Rohit A, Caitlyn T W, A. Simon P, Vikrant V, Rachel M, Lewis F, et al. Psychometric Properties of the EuroQol-5D and Short Form-6D in Patients with Systemic Lupus Erythematosus. The Journal of Rheumatology. 2009;36(6):1209.

[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(3):525-32.

[49] van Vollenhoven RF, Mosca M, Bertsias G, Isenberg D, Kuhn A, Lerstrøm K, et al. Treat-to-target in systemic lupus erythematosus: recommendations from an international task force. Ann Rheum Dis. 2014;73(6):958-67.

[50] Machado Escobar MA, Yacuzzi MS, Martinez RN, González Lucero L, Bellomio VI, Santana M, et al. Validation of an Argentine version of Lupus Quality of Life questionnaire. Lupus. 2016;25(14):1615-22.