Development of a lupus nephritis suboptimal response prediction tool using renal histopathological and clinical laboratory variables at the time of diagnosis

Lindsay N Helget,[1,2] David J Dillon,[1] Bethany Wolf,[3] Laura P Parks,[1] Sally E Self,[4] Evelyn T Bruner,[4] Evan E Oates,[5] Jim C Oates ![ORCID][1,6]

For numbered affiliations see end of article.

Correspondence to
Dr Jim C Oates; oatesjc@musc.edu

ABSTRACT

Objective Lupus nephritis (LN) is an immune complex-mediated glomerular and tubulointerstitial disease in patients with SLE. Prediction of outcomes at the onset of LN diagnosis can guide decisions regarding intensity of monitoring and therapy for treatment success. Currently, no machine learning model of outcomes exists. Several outcomes modelling works have used univariate or linear modelling but were limited by the disease heterogeneity. We hypothesised that a combination of renal pathology results and routine clinical laboratory data could be used to develop and to cross-validate a clinically meaningful machine learning early decision support tool that predicts LN outcomes at approximately 1 year.

Methods To address this hypothesis, patients with LN from a prospective longitudinal registry at the Medical University of South Carolina enrolled between 2003 and 2017 were identified if they had renal biopsies with International Society of Nephrology/Renal Pathology Society pathological classification. Clinical laboratory values at the time of diagnosis and outcome variables at approximately 1 year were recorded. Machine learning models were developed and cross-validated to predict suboptimal response.

Results Five machine learning models predicted suboptimal response status in 10 times cross-validation with receiver operating characteristics area under the curve values >0.78. The most predictive variables were interstitial inflammation, interstitial fibrosis, activity score and chronicity score from renal pathology and urine protein-to-creatinine ratio, white blood cell count and haemoglobin from the clinical laboratories. A web-based tool was created for clinicians to enter these baseline clinical laboratory and histopathology variables to produce a probability score of suboptimal response.

Conclusion Given the heterogeneity of disease presentation in LN, it is important that risk prediction models incorporate several data elements. This report provides for the first time a clinical proof-of-concept tool that uses the five most predictive models and simplifies understanding of them through a web-based application.

INTRODUCTION

Lupus nephritis (LN) is an immune complex-mediated glomerular and tubulointerstitial

Key messages

What is already known about this subject?
► Individual elements of the International Society of Nephrology/Renal Pathology Society renal histopathological activity and chronicity indices associate with lupus nephritis outcomes; however, individually, each element has poor predictive performance.

What does this study add?
► This report demonstrates proof of concept that machine learning techniques, including histopathological and laboratory elements, can account for the heterogeneity of disease in lupus nephritis and provide good prediction of approximately 1 year lupus nephritis suboptimal response in a largely African-American population.

How might this impact on clinical practice or future developments?
► The web-based clinical tool accompanying this report can be used by clinicians to initiate discussions with patients and care teams about lupus nephritis care coordination and monitoring; however, it needs to be tested in a prospective fashion to determine if making therapeutic decisions based on the predictions can improve outcomes.

disease in patients with SLE. Approximately 50% of patients with SLE develop kidney-related complications, including LN, and up to 48% of those with diffuse proliferative disease can progress to end-stage renal disease within 5 years of diagnosis among African-Americans.[1] The American College of Rheumatology (ACR) currently recommends changing therapy when patients with LN are deemed non-responders after for 6 months of induction therapy.[2] Clinicians use a variety of serum markers, including C3, C4, antidouble stranded DNA (anti-dsDNA) and creatinine, as well as urine protein-to-creatinine

Table 1 Baseline characteristics by response status at approximately 1 year

	Non-responder (n=42)	Responder (n=41)	P value
Demographics (n (%) unless stated)			
Age at biopsy, years, mean (SD)	32 (11)	31 (9.6)	0.455
Sex, female	39 (91)	39 (89)	1.000
Race, black	41 (95)	38 (86)	0.266
Baseline lab measures (median, IQR unless stated)			
Urine protein-to-creatinine ratio (g/g)	2.7 (4.8)	1.31 (3.7)	0.018
eGFR	92 (76)	106 (77)	0.534
Serum C3 (mg/dL), mean (SD)	65 (30)	71 (29)	0.554
C4, mean (SD)	15 (8.6)	16 (7.3)	0.761
Anti-dsDNA antibodies	176 (162)	119 (189)	0.256
Serum creatinine	1.0 (1.3)	0.90 (0.70)	0.269
White blood cell count	5.8 (4.4)	6.5 (6.1)	0.504
Platelets	251 (123)	247 (113)	0.745
Haemoglobin, mean (SD)	10 (1.4)	10 (1.4)	0.656
Albumin, mean (SD)	2.4 (0.8)	2.6 (0.6)	0.149
Baseline pathology (# patients >0, median (IQR) or # of patients (%))			
Activity score	38, 5 (5)	35, 3 (6)	0.091
Chronicity score	36, 4 (5)	34, 2 (2)	0.004
Interstitial fibrosis	34, 2 (1.)	24, 1 (1)	0.001
Interstitial inflammation	30, 1 (1)	29, 1 (1)	0.978
Glomeruli	42, 18 (13)	41, 19 (11)	0.539
Crescents	15, 0 (2)	20, 0 (2)	0.382
Crescent:Glomeruli ratio	15, 0 (0.2)	20, 0 (0.08)	0.261
Necrosis	1 (1)	2 (1)	0.703
Proliferative disease, yes	37 (86)	32 (73)	0.125
Membranous disease, yes	23 (53)	20 (46)	0.454
Mesangial disease, yes	3 (7)	6 (14)	0.484
Thrombotic microangiopathy (#, %)	2 (5)	1 (2)	1.000
Induction medication use (n (%))			
Prednisone, yes	36 (86)	38 (93)	0.307
Hydroxychloroquine, yes	34 (81)	34 (83)	0.815
Mycophenolate/Mycophenolic acid, yes	28 (67)	32 (78)	0.367
Cyclophosphamide, yes	14 (33)	9 (22)	0.361
Rituximab, yes	6 (14)	1 (2.4)	0.109
Azathioprine, yes	7 (17)	6 (15)	0.799
# immunosuppressants at baseline*			
0	0 (0)	0 (0)	
1–2	12 (29)	10 (24)	0.403
≥3	30 (71)	31 (76)	

dsDNA, double stranded DNA; eGFR, estimated glomerular filtration rate.

ratio and sediment to monitor response to therapy, but response to therapy is not defined in the guidelines.[2] While in this 6-month trial of induction therapy, patients who do not respond can develop additional irreversible renal damage. A decision support tool based on machine learning models could be useful in determining the baseline characteristics of patients who are less likely to respond to induction therapy. Currently, no clinically useful machine learning model of 1-year outcomes has been developed. Univariate or linear modelling has not predicted outcomes well in this heterogeneous disease.

We hypothesised that a combination of renal pathology results and routine clinical laboratory data could be used to develop and to cross-validate an early LN decision

Table 2 Univariate cvAUC for the subset of seven predictors selected for inclusion in the models

Variable	cvAUC
Activity score	0.591
Chronicity score	0.660
Interstitial fibrosis	0.643
Interstitial inflammation	0.482
Urine protein-to-creatinine ratio	0.503
White blood cell	0.511
Haemoglobin	0.524

AUC, area under the curve; cvAUC, cross-validated AUC.

support tool predictive of suboptimal response at approximately 1 year in LN.

METHODS
Patient population
Patient data were derived from the Medical University of South Carolina (MUSC) Core Center for Clinical Research (CCCR) prospective longitudinal cohort. The CCCR database is sponsored by the NIH under the P30 mechanism and is longitudinal registry and biorepository of predominately African-American patients with SLE. Patients selected met at the ACR or SLE International Cooperating Clinics criteria for SLE,[3 4] were evaluated by a rheumatology provider at MUSC, and agreed to be enrolled in the cohort. The prospective longitudinal cohort database was created in 2003. Disease criteria and renal biopsy data were largely entered prospectively, but some were added retrospectively through chart review.

Inclusion criteria
Patients were also selected if they had revised International Society of Nephrology/Renal Pathology Society (ISN/RPS) active class I, II, III, IV or V nephritis by histopathology either at initial presentation of LN or with a worsening of LN indicating repeat biopsy between 2003 and 2017.[5] Only patients with complete laboratory data available at time of renal biopsy and approximately 12 months (7–24 months) after renal biopsy were included in the study.

Laboratory analysis
Laboratory values were obtained as part of standard patient care. Tests for 24-hour urine protein, protein-to-creatinine ratio, serum creatinine, albumin, haemoglobin, white blood cell count, platelet count, anti-dsDNA, C3 and C4 were performed by Clinical Laboratory Improvement Amendments-certified central laboratories at MUSC, LabCorp or external hospital laboratories. The estimated glomerular filtration rate (eGFR) was determined by using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.[6]

Pathology analysis
Renal biopsies were read by one of two renal pathologists at MUSC (SES or ETB) using the 2018 revised ISN/RPS activity (0–24) and chronicity (0–12) index elements (each scored 0–3 for 0%, <25%, 25%–50% and >50% involvement).[5] Included in the activity index are endocapillary hypercellularity, karyorrhexis, fibrinoid necrosis, hyaline deposits, cellular or fibrocellular crescents and interstitial inflammation. Scores were doubled for crescents and necrosis. Included in the chronicity index were total glomerulosclerosis score, fibrous crescents, tubular atrophy and interstitial fibrosis.

Data handling
Data for the CCCR database were directly adjudicated into the REDCap database using the Data Transfer Service[7] to eliminate transcription error. Some laboratory values were extracted by chart review of the electronic medical record and scanned lab reports and were entered manually into the database. Five per cent double entry of abstracted data was performed to ensure transcription error was <5%. Pathology reports were manually abstracted and entered into the registry if not already present.

Statistical analysis
The outcome variable was the failure to completely respond to therapy at approximately 1 year. This time point was chosen, as variables of response at 1 year are predictive of long-term response years in the MAINTAIN and Euro-Lupus Nephritis Trials.[8] Response was defined by a modification of the ACR response criteria.[9] The modifications to these

Table 3 Prediction performance and variables selected for each model for the five models with a cvAUC >0.75

Model	cvAUC	Activity score	Chronicity score	Interstitial fibrosis	Interstitial inflammation	Urine protein-to-creatinine ratio	WBC	Hgb	# of predictors
LR	0.780	X		X	X		X	X	5
RF	0.800		X		X	X	X		4
SVML	0.783	X		X	X			X	4
SVMR	0.790	X	X	X	X		X	X	6
ANN	0.775	X	X	X	X	X			5

ANN, artificial neural networks; AUC, area under the curve; cvAUC, cross-validated AUC; Hgb, haemoglobin; LR, logistic regression; RF, random forest; WBC, white blood cell.

Table 4 Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for different cut-offs for probability of non-response at 1 year based on the predicted probability of non-response for each model and the average predicted probability across the five models

Model	Cut-off	Sensitivity	Specificity	PPV	NPV
	0.3	0.86	0.48	0.62	0.80
LR	0.5	0.65	0.62	0.67	0.67
	0.7	0.37	0.91	0.80	0.60
	0.3	0.91	0.36	0.58	0.80
RF	0.5	0.74	0.61	0.65	0.71
	0.7	0.42	0.86	0.75	0.60
	0.3	0.93	0.27	0.56	0.80
SVML	0.5	0.67	0.68	0.67	0.68
	0.7	0.19	0.93	0.73	0.54
	0.3	0.93	0.30	0.56	0.81
SVMR	0.5	0.70	0.61	0.64	0.68
	0.7	0.37	0.95	0.89	0.61
	0.3	0.81	0.48	0.60	0.72
ANN	0.5	0.65	0.68	0.67	0.67
	0.7	0.33	0.84	0.67	0.56
Average	0.3	0.93	0.30	0.56	0.85
prediction across	0.5	0.70	0.73	0.71	0.71
the five models	0.7	0.30	0.95	0.87	0.58

ANN, artificial neural networks; LR, logistic regression; RF, random forest.

criteria were described previously by Wofsy et al.[10] Briefly, this modified complete response includes attaining a urine protein-to-creatinine ratio of <0.5 at approximately 1 year and achieving an eGFR of 90 or an improvement of at least 15% from baseline. The suboptimal response outcome was defined by lack of achieving complete response as defined above. Thus, the outcome includes non-responders and partial responders. Variables collected in the data included patient sex, age at time of biopsy, proliferative disease (ISN/RPS classes III or IV, Y/N), mesangial disease (ISN/RPS class I or II, Y/N), membranous disease (ISN/RPS class V, Y/N), activity score (0–3), chronicity score (0–3), interstitial fibrosis (0–3), interstitial inflammation (0–3), number of glomeruli evaluated, crescents (number), crescent-to-glomeruli ratio (0–3×2), necrosis (0–3×2), urine protein-to-creatinine ratio, eGFR by the CKD-EPI formula (eGFR, mL/min/1.73 m²),[6] serum creatinine (mg/dL), dsDNA (IU), C3 (mg/dL), C4 (mg/dL), white blood cells count (k/µL), platelet count (k/µL), haemoglobin (g/dL), serum albumin (mg/dL), prednisone (Y/N), hydroxychloroquine (Y/N), mycophenolate mofetil/mycophenolic acid (Y/N), cyclophosphamide (Y/N), rituximab (Y/N), azathioprine (Y/N) and number of medications. Since the data were retrospective and not prospectively randomised, immunosuppressants used for induction are subject to bias by indication and were not considered predictive. They were excluded from consideration during model development for clinical use, as their presence might imply that choice of induction therapy based in the modelling might

affect the outcomes. Descriptive statistics were calculated for all participant characteristics by treatment response category. Univariate associations between all baseline characteristics and treatment response were evaluated using a series of logistic regression models.

The goal for this study was to identify a parsimonious subset of predictors from patient demographics and baseline laboratory and biopsy data that yielded good prediction performance characteristics for a set of multivariable prediction models of suboptimal response at approximately 1 year. Multivariable classification models considered in this study included logistic regression (LR), classification and regression trees (CART), random forest (RF), support vector machines with linear, polynomial and Gaussian kernels (SVML, SVMP and SVMR, respectively), naïve Bayes (NB) and artificial neural networks (ANN). RF models were fit using the 'randomForest' package; LR models were fit using the 'stats' packages; SVMs and NB models were fit using the 'e1071' package; ANNs models were fit using the 'nnet' available in R.[11 12] Tuning parameters for the different models considered were selected before developing the models. An initial exhaustive search of all combinations of up to 20 variables was considered. However, the results from this search found the best average performance occurred when models included eight variables (online supplemental figure 2). Thus, variable selection was conducted using an exhaustive examination of all subsets of eight or fewer predictors. This threshold number was considered for all modelling approaches to make the
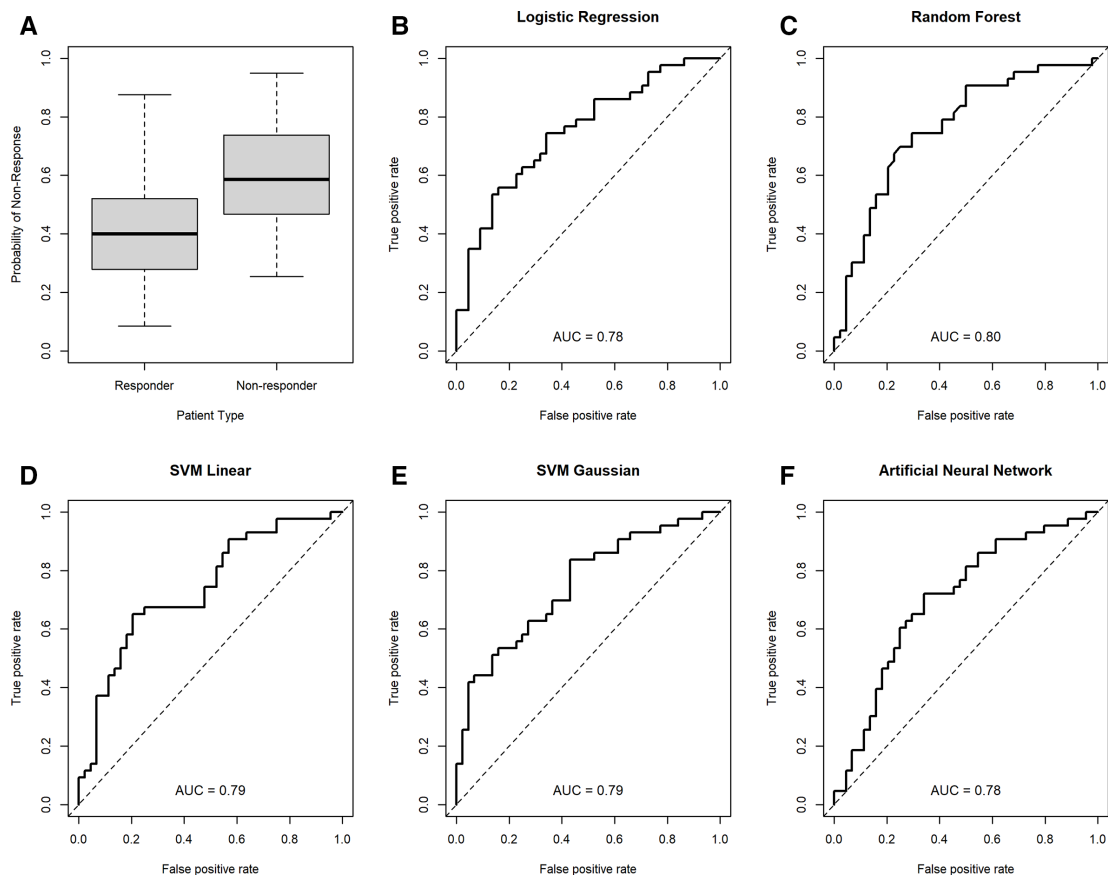
Figure 1 Cross-validation area under the curve (cvAUCs) for each of the final machine-learning models. A summary of probability scores from all models in responders and non-responders (A); cvAUCs depicted for logistic regression (B), random forest, (C) SVM linear, (D) SVM Gaussian (E) and artificial neural network (F) models.

models more useful in a busy clinical setting. Specifically, prediction performance for each model's subset of predictors was conducted using a 10-fold cross-validation (CV) approach. Ten-fold CV divides the data into 10 subsets. Models are trained using 9/10 of the data and tested on the remaining 1/10 of the data, and this is repeated for each subset. The cross-validated area under the curve (cvAUC) is the average AUC calculated for each subset of 1/10 of the data excluded during model development and has been shown to be more robust than use of a single training-test set approach.[13] The goal was to identify a small subset of predictors with good prediction performance across the models.[14] Prediction performance was measured by 10-fold cvAUC and the best subset of eight variables was selected as the subset that resulted in the highest average cvAUC across all models. Sensitivity, specificity, positive predictive value and negative predictive values were determined for select thresholds for the predicted probability of non-response returned by each model. All analyses were conducted in R V.4.0.2. An R-Shiny, web-based tool was created based on the resulting models that were selected.

RESULTS

Within the registry, 149 patients had renal biopsy information available between 2003 and 2017. Of these, 83 patients with LN had baseline and approximately 1-year follow-up renal response data between 7.5 and 24 months (online supplemental figure 1). Three with follow-up <7 months and four with follow-up >2 years were excluded. Approximately half of the participants were classified as subtoptimal responders at 1 year. Participant characteristics by treatment response are reported in table 1.

The subset of eight or fewer predictors yielding the best prediction performance across the different models included activity, chronicity, interstitial fibrosis and interstitial inflammation scores and baseline laboratory values for urine protein-to-creatinine ratio, white blood cell count and haemoglobin. The univariate cvAUCs for these seven variables are shown in table 2. The cvAUCs for the best models selected for each modelling approach range from 0.62 to 0.80 with the random forest model yielding the best cvAUC. Five of the eight models considered had a cvAUC >0.75 and included the LR, RF, SVML, SVMR and the ANN models. Table 3 shows the cvAUCs and predictors subset for the five models with high cvAUC. Table 4 shows the sensitivity and specificity of each model and based on the average prediction across all five models at three thresholds each. Interstitial inflammation was the most consistent predictor and was included in all five models. Activity score and interstitial fibrosis were also relatively consistent and were included in all but the RF model. The receiver operating characteristics (ROC)
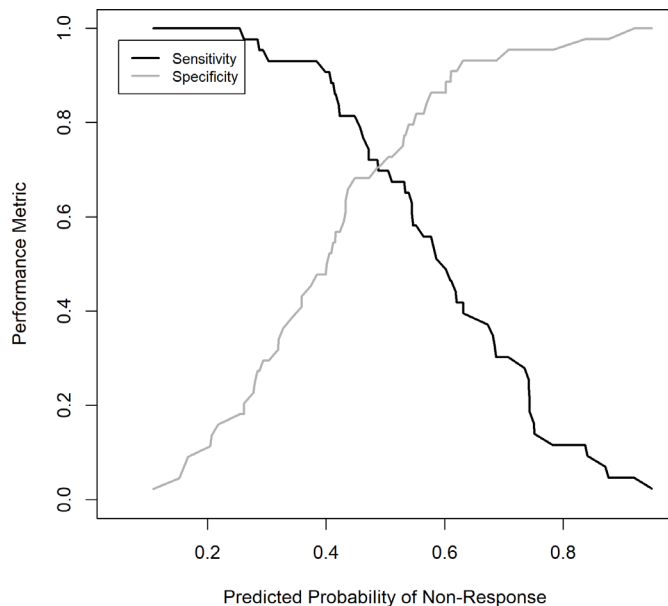
Figure 2 Mean model sensitivity and specificity based on chosen prediction threshold. The mean of all model predictions was used to determine the performance of the model at select thresholds. The sensitivity (black line) and specificity (grey line) are depicted on the y-axis for each threshold (reported on the x-axis).

curves based on the 10-fold CV prediction from each of the five selected models is shown in Figure 1. Figure 2 depicts the performance (sensitivity and specificity) of the mean of all the models based on the prediction threshold chosen. The CART, SVMP and NB models all had cvAUCs <0.7 and were excluded from further consideration.

Using the seven most consistent variables, a web-based application was created with R shiny to serve as a clinical tool and can be found here: (https://histologyapp.shinyapps.io/LN_histology_prediction_tool/). In the application, an example of patient's data is displayed by default. To evaluate a new patient, users can enter histology and clinical laboratory values obtained at the time of biopsy. The resulting graph depicts, as coloured dots, the predicted probability of non-response of each model at a threshold of 0.5. The size of each dot reflects the ROC AUC in the validation set. The box plots on the overview page represent the median (line), IQR (box) and the 25th and 75th percentile±1.5 times the IQR (whiskers) of probability scores for suboptimal response in individual patients in the responder and non-responder groups in the validation sets (also depicted in figure 2A). Each coloured dot represents a single model prediction of the patient viewed in context of the performance of the model in the validation set. The black 'X' represents the mean predicted probability of non-response of all five models.

DISCUSSION
Our study determined that machine learning can be used to develop cross-validated models with good prediction

of suboptimal response to therapy in patients with LN, predominantly of African descent. These models were developed using readily available clinical laboratory and histopathological elements at the time of diagnosis. The utility of individual histopathological features to predict outcomes in diffuse proliferative lupus nephritis is well described. Previous work has demonstrated the predictive strength of composite activity index score in progression to renal failure, with individual histological features of activity such as cellular crescents and fibrinoid necrosis showing positive associations with renal failure.[15] The composite of chronicity index as well as individual features such as interstitial fibrosis, glomerular sclerosis and fibrous crescents were also predictive, with tubular atrophic change determined to be of particularly high predictive value with respect to progression to renal failure.[15] The addition of clinical data elements, particularly serum creatinine, haematocrit and race to histological features improved prediction in subsequent work, while other studies showed at use of stand-alone activity or chronicity indices was insufficient in predicting response.[16–19] Recent machine learning approaches have used multilinear regression and random forest modelling to predict of pathological classification, activity and chronicity from clinical laboratory values and have shown promise.[20] This study expands on this approach by using pathological variables to predict a outcome at approximately 1 year.

As demonstrated in previous studies, the addition of novel biomarkers obtained at time of diagnosis can be used to develop a robust model to predict the 1-year outcome.[21] However, no measures of these novel biomarkers are in clinical use. In this study, readily available histological and clinical laboratory values were used to predict LN outcomes with an ROC AUC of >0.75 in five models. This study is unique in that individual activity and chronicity score elements were used in machine learning models. This is supported by prior literature describing their use individually.[15]

The results of this prediction modelling should be interpreted in the proper clinical context. The presence of high interstitial fibrosis with a low activity score is likely indicative chronic injury for which immunosuppression would be used to prevent further worsening rather than to restore renal function. However, those with high activity scores might be approached with combination immunosuppressive therapies or monitored for addition of second-line therapies sooner to improve or preserve renal function. Given that these data were retrospective, we do not know whether using the predictions to guide decisions on induction therapy will change outcomes. As is often the case in managing patients with autoimmune disease, it would not be appropriate for a clinician to make decisions about choice of therapy based on this prediction score without consideration of the clinical context. However, a prediction of non-response could be used to change the frequency of monitoring for response and care coordination to ensure medication adherence.

It could also be considered in decisions to add on or change therapy for lack of response at earlier time points than the recommended 6 months according to the ACR guidelines.

This study has several limitations. The information used to model outcomes came mostly from patients of African descent. Therefore, use in clinical practice may be limited to this demographic. While the longitudinal data in this study were collected prospectively, the analysis was retrospective. While thrombotic microangiopathy has been associated with outcomes in prior studies,[22] the number of biopsies with this finding (five) limited its significance in the models. The decision aid presented here has not been used prospectively to determine if altering treatment strategy based on prediction scores improves outcomes. It has not been evaluated in the prediction of long-term outcomes and should therefore not be used to predict > 1-year outcomes. However, 1-year outcomes have been associated with long-term outcomes in the MAINTAIN and Euro-Lupus Nephritis Trials.[8] Selection of patients for inclusion in this study could bias outcomes. For instance, those with more rapidly progressive renal disease are more likely to receive a biopsy. Those with baseline and approximately 1-year follow-up available in the medical record may be biased to more favourable outcomes. We could not rigorously study the effect of medications on outcomes, as most patients received mycophenolate or mycophenolic acid.

Given the heterogeneity of disease presentation in LN, it is important that risk prediction models incorporate several data elements. This report provides for the first time a clinical tool that uses the five most predictive models and simplifies understanding of them through a web-based application. With the prediction models proposed here, we present a proof of concept for a tool that can inform both the frequency of monitoring and facilitate discussions with patients about choice of therapy. More closely monitoring or deploying care coordination for those who are predicted to have a suboptimal response to therapy has the potential to improve outcomes but should be tested prospectively.

Author affiliations

[1]Department of Medicine, Medical University of South Carolina, Charleston, South Carolina, USA

[2]Department of Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

[3]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

[4]Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, South Carolina, USA

[5]Vanderbilt University, Nashville, Tennessee, USA

[6]Medical Service, Ralph H Johnson VA Medical Center, Charleston, South Carolina, USA

ORCID iD

Jim C Oates http://orcid.org/0000-0002-5426-1359

REFERENCES

1 Dooley MA, Hogan S, Jennette C, *et al*. Cyclophosphamide therapy for lupus nephritis: poor renal survival in black Americans. glomerular disease collaborative network. *Kidney Int* 1997;51:1188–95.

2 Hahn BH, McMahon MA, Wilkinson A, *et al*. American College of rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res* 2012;64:797–808.

3 Hochberg MC. Updating the American College of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.

4 Isenberg D, Wallace DJ, Nived O. Derivation and validation of systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012;64.

5 Bajema IM, Wilhelmus S, Alpers CE, *et al*. Revision of the International Society of Nephrology/Renal pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of health activity and chronicity indices. *Kidney Int* 2018;93:789–96.

6 Pei X, Yang W, Wang S, *et al*. Using mathematical algorithms to modify glomerular filtration rate estimation equations. *PLoS One* 2013;8:e57852.

7 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.

8 Ugolini-Lopes MR, Seguro LPC, Castro MXF, *et al*. Early proteinuria response: a valid real-life situation predictor of long-term lupus renal outcome in an ethnically diverse group with severe biopsy-proven nephritis? *Lupus Sci Med* 2017;4:e000213.

9 Renal Disease Subcommittee of the American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria. The American College

of rheumatology response criteria for proliferative and membranous renal disease in systemic lupus erythematosus clinical trials. *Arthritis Rheum* 2006;54:10.1002/art.21625:421–32.

10 Wofsy D, Hillson JL, Diamond B. Comparison of alternative primary outcome measures for use in lupus nephritis clinical trials. *Arthritis Rheum* 2013;65:1586–91.

11 Team RC. A language and environment for statistical computing, 2020. Available: https://www.R-project.org

12 Liaw A, Wiener M. Classification and regression by random forest. *R News* 2002;2:18–22.

13 Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. New York: Springer, 2001.

14 LeDell E, Petersen M, van der Laan M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron J Stat* 2015;9:10.1214/15-EJS1035:1583–607.

15 Austin HA, Muenz LR, Joyce KM, *et al*. Diffuse proliferative lupus nephritis: identification of specific pathologic features affecting renal outcome. *Kidney Int* 1984;25:689–95.

16 Austin HA, Boumpas DT, Vaughan EM, *et al*. High-Risk features of lupus nephritis: importance of race and clinical and histological factors in 166 patients. *Nephrol Dial Transplant* 1995;10:1620–8.

17 Austin HA, Boumpas DT, Vaughan EM, *et al*. Predicting renal outcomes in severe lupus nephritis: contributions of clinical and histologic data. *Kidney Int* 1994;45:544–50.

18 Schwartz MM, Bernstein J, Hill GS, *et al*. Predictive value of renal pathology in diffuse proliferative lupus glomerulonephritis. lupus nephritis collaborative Study Group. *Kidney Int* 1989;36:891–6.

19 Rijnink EC, Teng YKO, Wilhelmus S, *et al*. Clinical and histopathologic characteristics associated with renal outcomes in lupus nephritis. *Clin J Am Soc Nephrol* 2017;12:734–43.

20 Tang Y, Zhang W, Zhu M, *et al*. Lupus nephritis pathology prediction with clinical indices. *Sci Rep* 2018;8:10231.

21 Wolf BJ, Spainhour JC, Arthur JM, *et al*. Development of biomarker models to predict outcomes in lupus nephritis. *Arthritis Rheumatol* 2016;68:1955–63.

22 Song D, Wu L-hua, Wang F-mei, *et al*. The spectrum of renal thrombotic microangiopathy in lupus nephritis. *Arthritis Res Ther* 2013;15:R12.