

DR. MICHELLE A PETRI (Orcid ID : 0000-0003-1441-5373)

PROF. ELOISA BONFA (Orcid ID : 0000-0002-0520-4681)

DR. SARFARAZ AHMED HASNI (Orcid ID : 0000-0002-4164-8837)

Article type : Full Length

### **Establishing Surrogate Kidney Endpoints for Lupus Nephritis Clinical Trials:**

#### **Development and Validation of a Novel Approach to Predict Future Kidney Outcomes**

Meggan Mackay MD, MS, Maria Dall’Era MD, Joanna Fishbein MPH, Kenneth Kalunian MD, Martin Lesser PhD, Jorge Sanchez Guerrero MD, MS, Deborah M. Levy MD, MS, Earl Silverman MD, Michelle Petri MD, MPH, Cristina Arriens MD, MSCS, Edmund J. Lewis MD, Stephen M. Korbet MD, Fabrizio Conti MD, PhD, Vladimir Tesar MD, PhD, Zdenka Hruskova MD, PhD, Eduardo F. Borba PhD, MD, Eloisa Bonfa PhD, MD, Tak Mao Chan MD, Manish Rathil MD, DM, K. L. Gupta MD, DM, Vivekanand Jha MD, DM, Sarfaraz Hasni MD, Melissa R. West, Elian Silverman, BA, Neil Solomons MD, Frederic A. Houssiau MD, PhD, Juanita Romera Diaz MD, MS, Juan Mejia-Vilet MD, MSc, and Brad H. Rovin MD

From the Feinstein Institute for Medical Research, Manhasset, NY (M.M, J.F., M.L.); Division of Rheumatology, University of California San Francisco, CA (M.D.); Division of Rheumatology, University of California San Diego, CA (K.K.); Division of Rheumatology, University of Toronto, Toronto, Canada (J.S.G); Division of Rheumatology, Hospital for Sick Children and University of Toronto, Toronto, Canada (D.M.L, E.S); Division of Rheumatology, Johns Hopkins University, Baltimore MD (M.P.); Oklahoma Medical Research Foundation, Oklahoma City (C.A.); Division of Nephrology, Rush University Medical Center, Chicago, IL (E.J.L., S.M.K.) and the Lupus Nephritis Collaborative Study Group\*; Rheumatology Unit, Department of Internal Medicine and Medical Specialties, Sapienza University of Rome, Rome, Italy (F.C.); Department of Medicine, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic (V.T., Z.H.); Rheumatology Division, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil (E.F.B., E.B.); Division of Nephrology, University of Hong Kong, Queen Mary Hospital, Hong Kong, China (T.M.C); Post-graduate Institute of Medical Education and Research, Chandigarh, India (M.R., K.L.G.); George Institute for Global Health India, New Delhi, India and University of Oxford, Oxford, UK, (V.J.); National Institute of Arthritis and Musculoskeletal and Skin

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.40724

This article is protected by copyright. All rights reserved.

Diseases, National Institutes of Health, Bethesda, MD (S.A.H.); Kidney Health Initiative, American Society of Nephrology, Washington D.C (MRW, ES); Aurinia Pharmaceuticals Inc, Victoria, British Columbia, Canada (N.S.); Cliniques Universitaires Saint-Luc, Pole de Pathologies Rhumatismales Inflammatoires et Systemiques, Institut de Recherche Experimentale et Clinique, Universite Catholique de Louvain, Belgium (F.A.H); Instituto Nacional de Ciencias Medicas and Nutricion Salvador Zubiran, Mexico City, Mexico (J.R.D., J.M-V.); Division of Nephrology, The Ohio State University Wexner Medical Center, Columbus, OH (B.H.R.)

\*A complete list of Collaborative Study Group Investigators is provided in the Supplementary Appendix

Address Correspondence To:

Brad H. Rovin, MD

Ohio State University Wexner Medical Center

395 West 12<sup>th</sup> Avenue

Ground Floor, Nephrology Division

Columbus, OH 43210

614.293.4997/614.293.3073 (fax)

Rovin.1@osu.edu

## **Abstract**

### **Objective**

Endpoints currently used in lupus nephritis (LN) clinical trials lack uniformity and questionably reflect long-term kidney survival. The objective of this investigation was to identify short-term endpoints that predict long-term kidney outcomes for use in clinical trials.

### **Methods**

A database of 944 LN patients was assembled from 3 clinical trials and 12 longitudinal cohorts. Variables from the first 12 months of treatment after diagnosis of active LN (prediction period) were assessed as potential predictors of long-term outcomes in a 36 month follow-up period. The long-term outcomes examined were new or progressive chronic kidney disease (CKD), severe kidney injury (SKI), and the need for permanent renal replacement therapy (RRT). Hazard Index Tools (HITs)

to predict risk for each outcome were derived using multivariable analysis with Cox proportional hazards regression.

## Results

Among 550 eligible subjects 54 CKD, 55 SKI and 22 RRT events occurred. Variables in the final CKD HIT were prediction period CKD status, 12-month proteinuria and 12-month serum creatinine (SCr). The SKI HIT included prediction period CKD status, ISN Class, 12-month proteinuria, 12-month SCr, race and an interaction between ISN Class and 12-month proteinuria. The RRT HIT included age at diagnosis, 12-month proteinuria and 12-month SCr. Each HIT validated well internally (c-indices 0.84-0.92) and in an independent LN cohort (c-indices 0.83-0.92).

## Conclusion

HITs, derived from short-term kidney responses to treatment correlate with long-term kidney outcomes, and now must be validated as surrogate endpoints for LN clinical trials.

## Introduction

Long-term preservation of kidney function is the goal in treating lupus nephritis (LN). In early LN clinical trials, improved renal outcomes were demonstrated by differences in hard endpoints such as doubling of serum creatinine concentration (SCr) or the development of end-stage kidney disease (ESKD) that were assessed prospectively over several years (1-4). In contrast, contemporary clinical trials are short, with 6-12 month complete and partial renal response endpoints (5, 6). There are many reasons for this change in trial duration, including the high cost of large, multicenter efforts and the challenges of retaining patients in a protocol over several years. Although all current definitions of renal response include an assessment of kidney function (SCr or estimated GFR, eGFR), proteinuria and urinalysis/urine sediment, no uniform response definition has been adopted, and the individual components are all assigned equal predictive value (Supplemental Table 1). The evolution of composite outcome measures to replace hard kidney endpoints reflects the practical need for tools to assess renal response in shorter trials. However composite outcome measures were often developed by expert opinion and consensus after review of available literature, leaving considerable room for interpretation and modifications when applied to individual clinical trials (7, 8), resulting in the lack of a single standard definition for renal response. This makes it difficult to compare results of LN trials or assess the efficacy of novel therapies. Furthermore, the evidence that these composite short-term endpoints predict preserved long-term kidney function is not definitive. The need for a consistent and validated endpoint for LN clinical trials is illustrated by the demonstration that a negative LN trial can become a positive LN trial if different renal response criteria are applied (6, 9). This study, in collaboration with the Kidney Health Initiative and the Lupus

Nephritis Trials Network, was undertaken to identify short-term clinical trial endpoints that predict risk of long-term adverse renal outcomes for use in future LN trials.

## **METHODS**

### **Subjects**

A database of 944 LN patients with extended, prospective follow-up from 12 international investigator-initiated lupus cohorts and 3 randomized controlled LN clinical trials (Table 1) was assembled. The investigator-initiated cohorts were not necessarily restricted to LN patients. The investigators were specifically asked to provide longitudinal LN patient data that was then entered into a uniform central database. Applicable data from the LN trials were also entered into the database. This process yielded the initial 944 LN patients. LN was treated according to local standards of care or per trial protocol. To avoid developing treatment-restricted endpoints we did not stratify by therapy, but all patients were treated with corticosteroids, and most received cyclophosphamide or MMF for induction.

After applying inclusion/exclusion criteria (Supplemental Figure 1) the sample size fell to 550 which included 31 pediatric patients. Baseline demographics and clinical characteristics are shown in Table 2. These patients were followed between 1981 and 2016. Follow-up visits were required to be 0.9-18 months apart to mitigate potential bias from extreme variability in time between follow-ups.

To verify the prediction models, an independent validation cohort of 275 LN patients followed prospectively and treated in accordance with local practices was analyzed (Table 2). These patients were followed from 2008-2016. This validation cohort was from Mexico City and more than 98% were of Hispanic ethnicity and Mestizo race, best characterized as having European, Native American and African admixture.

### **Study Design**

To identify short-term surrogate markers of long-term kidney outcomes, a prediction period was defined as the interval between baseline and 12±2 months. Clinical variables measured during this

period were assessed individually and as part of multivariable prediction models. Baseline was designated as the visit when the treating physician determined that a patient had active LN and started induction therapy. Kidney outcomes were assessed during the follow-up period that began immediately after the prediction period and ended at a maximum of 48 months from baseline.

### *Prediction Variables*

Potential predictors of kidney outcomes included sex, age at study entry, race, ethnicity, absolute values of proteinuria and SCr at baseline and 12 months, % proteinuria change from baseline to 12 months, presence or absence of urine RBCs (defined as  $\geq 5$  RBC/high-power-field) at baseline and 12 months, ISN/RPS Class, and the development or progression of chronic kidney disease (CKD) in the prediction period. Although serologies like anti-dsDNA antibody and complement levels, and kidney biopsy elements, such as tubulointerstitial disease have been implicated in outcomes, the potential predictors evaluated in this study were restricted by available data. Race was operationalized as a binary variable (Black versus others) to satisfy model assumptions and remain consistent with the CKD-EPI equation which handles race similarly. Ethnicity was defined as Hispanic, non-Hispanic, or not determined. Due to cohort-to-cohort variability in the measurement of proteinuria, total protein, urine protein-to-creatinine ratio (UPCR) from a 24-hour urine collection, or a random spot UPCR were all accepted as reflecting 24-hour proteinuria. Absolute values of SCr and proteinuria were analyzed after natural log (log) transformation to satisfy model assumptions of linearity, and expressed as  $\log(\text{value SCr})$  and  $\log(\text{value proteinuria})$ . A 2-fold increase in 12-month proteinuria corresponds to a change of approximately 0.693 on the natural log scale. Percent change for proteinuria was expressed as the  $\log\left(\frac{\text{measurement at 12 months}}{\text{measurement at baseline}}\right)$ . Development or progression of CKD in the prediction period was defined by an increase in SCr  $\geq 30 \pm 2\%$  sustained over the remainder of the prediction period leading to an eGFR  $< 60$  ml/min/1.73m<sup>2</sup>. ISN Class was operationalized as a binary covariate (Membranous vs Proliferative); subjects with class 5 or 2 & 5 were categorized as membranous, subjects with Classes 3, 4, 3 & 5, 4 & 5, or unknown were categorized as proliferative (the assumption being that those not biopsied were most likely to be proliferative).

### *Kidney Outcomes*

New or progressive CKD, severe acute kidney injury (SKI), and the need for renal replacement therapy (RRT) are recognized as outcomes that adversely impact renal and/or patient survival (10-13). The training cohort was used to develop predictors of risk for developing one or more of these poor outcomes in the follow-up period. Follow-up time was restricted to a maximum of 48 months

from baseline to reduce potential bias from the substantial variability in follow-up time between subjects in this cohort (range 34-268 months from baseline).

CKD was defined as a sustained decrease in eGFR, (CKD-EPI equation (11)), of at least  $30 \pm 2\%$  such that:

- the minimum 28% decrease in eGFR in the follow-up period was measured in relation to the highest eGFR reported in the 12-month prediction period and was sustained on a consecutive follow-up visit occurring at least 3 months ( $\pm 0.5$  months) later
- subjects requiring permanent RRT at any time during follow-up were included in the CKD outcome, as it was assumed that a patient had to develop CKD prior to reaching RRT
- the sustained decrease in eGFR had to reach an absolute value  $< 60 \text{ ml/min/1.73m}^2$

A 30% sustained decline in eGFR was considered the minimal fall in eGFR to qualify as CKD, based on a National Kidney Foundation and FDA workshop(14). The 2% variation was incorporated to allow variability in assays. CKD status was identified by a computerized algorithm and verified for accuracy through review of the raw data.

SKI was defined as an acute or sustained (no recovery) decline of 50% or more in eGFR in the follow-up period relative to the highest eGFR observed in the prediction period. Similar to the rationale for including RRT in the CKD outcome, all subjects who reached RRT during follow-up were also included in the SKI outcome. SKI was not included in the CKD or RRT outcomes.

RRT was defined as the need for permanent hemodialysis or peritoneal dialysis, or having received a kidney transplant at any time during follow-up (12-48 months).

## **Statistical Methods**

### *Model Development*

Descriptive statistics were calculated using proportions and frequencies for categorical variables and means and standard deviations for continuous variables for the training cohort. Each outcome (CKD, SKI, and RRT) was analyzed using standard methods of analysis for time to event data, such as survival analysis. Specifically, time to the endpoint of interest was measured in months and compared among categorical grouping variables via computation of the Kaplan-Meier product limit curves and log-rank tests. In cases where the “event” of interest was not observed during the follow-

up period, the number of months until last follow-up, up to a maximum of 36 months from end of the prediction period, was used and the subject's vital status was classified as censored.

Multivariable analysis was conducted using Cox proportional hazards regression to develop the three Hazard Index Tools (HITs) from clinical data collected in the prediction period that could predict risk for development of new or progressive CKD, SKI or need for RRT. Necessary model assumptions (e.g multicollinearity, proportional hazards, etc.) were addressed. Backward elimination with a 5% significance level was used to select final multivariable models. When applicable, appropriate data transformations were applied. The Likelihood Ratio Test was used to compare nested models and the Akaike's information criterion (AIC) was used to compare non-nested models whereby the candidate model with the lowest AIC was selected for each outcome. Results from these analyses were used to develop the formulae for the HITs to predict each adverse renal outcome. Results yielding a  $p$ -value $<0.05$  were considered statistically significant.

#### *Model Validation*

Internal validations of the predictive abilities for the CKD, SKI and RRT HITs were performed on the training cohort using Harrell's concordance index ( $c$ -index) (15), which measures discrimination, a prediction tool's ability to accurately rank subjects from high to low risk, and allows for the censoring in time-to-event survival analyses. The  $c$ -index ranges from 0.5 (no discrimination) to 1.0 (perfect discrimination) (16); a higher value indicates better discriminative ability. Corresponding 95% confidence intervals (CI) around the  $c$ -index for each prognostic model were estimated using 1,000 bootstrap resamples with replacement.

External validation of each final model was performed on an independent longitudinal LN cohort. (Table 1). Discrimination of each final model in the validation sample was evaluated by computing the  $c$ -index and corresponding 95% confidence intervals using 100 bootstrap resamples with replacement.

Analyses were performed with SAS Version 9.4 STAT 14.2 (Cary, NC).

## **Results**

### *Kidney Outcomes*

Among 550 subjects eligible for analysis in the training cohort (Supplemental Figure 1), 54 CKD, 55 SKI and 22 RRT events were identified. The median times to CKD, SKI and RRT for the entire cohort were not estimable based on the pattern of the censoring, event rate and follow-up. Times to events were known for individual patients. Three years after the prediction period, the estimated

proportion of subjects without: i) new or worsening CKD was 87.2% (95% CI: 83.3% to 90.2%); ii) SKI was 88.3% (95% CI: 84.9% to 91.0%); or iii) RRT was 94.6% (95% CI: 91.7% to 96.5%).

After backward elimination, urine RBC status at baseline and 12 months and ethnicity did not significantly contribute to any of the risk models. The final multivariable Cox proportional hazards regression model to predict CKD included proteinuria at 12 months, SCr at 12 months, and CKD status in the prediction period. The parameters and hazard ratios of these predictors are shown in Table 3 and the HIT formula to predict future CKD is given in Table 6.

Similarly, the final multivariable Cox proportional hazards regression model to predict SKI included 12-month proteinuria, 12-month SCr, and CKD status in the prediction period, as well as race and ISN class, with an interaction term between ISN Class and 12-month proteinuria. The parameters and hazard ratios of the predictors are shown in Table 4 and the HIT formula to predict SKI is given in Table 6. These results demonstrate that the relationship between 12-month proteinuria and risk of future SKI is contingent on ISN Class ( $p=0.0016$ ). Overall, patients with proliferative LN always have a greater risk of SKI than patients with membranous LN. The modeling is consistent with clinical and histologic observations of patients with proliferative and pure membranous LN. The model shows that in patients with proliferative LN a 1 unit increase in proteinuria (log scale) at 12 months (approximately 2.72 g/d) confers an additional 1.62-fold increased hazard of SKI (95% CI: 1.24, 2.09) after co-variate adjustment. This is likely because acute kidney injury in proliferative LN is driven mainly by intra-renal inflammation, and to a lesser extent by the level of proteinuria. In contrast, for membranous patients there is a 7.76-fold greater hazard for each unit increase in 12-month proteinuria (95% CI: 3.02, 19.93) after covariate adjustment. These findings indicate that in membranous LN the risk of developing SKI is more dependent on the level of proteinuria. This is consistent with membranous LN being non-inflammatory and acute kidney injury occurring mainly due to proteinuria-induced tubular damage.

Finally, the multivariable Cox proportional hazards regression model to predict RRT included proteinuria at 12 months, SCr at 12 months, and age at study entry. The parameters and hazard ratios of the predictors are shown in Table 5 and the HIT formula to predict RRT in LN patients is given in Table 6.

## *Validation*

Each model was internally validated for predictive ability in the training cohort. The RRT model yielded the greatest *c*-index of 0.92 (95% CI: 0.88, 0.97), followed by the CKD model with a *c*-index of 0.88 (95% CI: 0.84, 0.93) and the SKI model with a *c*-index of 0.84 (95% CI: 0.79, 0.90).

External validation using an independent LN cohort verified the predictive ability of each model. In the 275 subject validation cohort (Table 2) there were 56 CKD, 48 SKI and 6 RRT events within 36 months of entry. The RRT model again yielded the greatest *c*-index of 0.92 (95% CI: 0.81, 0.99), followed by the CKD and SKI models, both with a *c*-index of 0.89 (95% CI: (0.83, 0.92) and (0.83, 0.94), respectively). Of note, the training and external validation cohorts were racially disparate (Table 2). Despite differences in cohort composition, each model validated well.

## *Application of Hazard Index Tools to Lupus Nephritis*

The proposed use of HITs to assess the effects of a new therapy on the risk for developing adverse kidney outcomes is shown in Supplemental Figure 2. The proposed use of HITs to assess an individual patient's risk for adverse renal outcomes is shown in Supplemental Figure 3. The HIT equations can also be used to determine sample size for specific kidney outcomes in trial design. An example is shown in Supplemental Figure 4.

## **DISCUSSION**

Using real-world LN cohorts and multicenter randomized clinical trials, we developed risk models to predict future adverse kidney outcomes based on clinical data acquired during the first year of therapy after diagnosis of active LN. These models predict risk for new or progressive CKD, severe acute kidney injury, and the need for permanent RRT. The models validated well in the original training cohort and in an independent, ethnically different LN cohort, suggesting applicability to LN patients in general. Although novel LN therapies are currently evaluated by short-term, non-standardized criteria (6), the goal of LN treatment is to preserve long-term kidney function and prevent CKD, acute kidney failure, and need for RRT. We therefore suggest that modeling risk of poor long-term kidney outcomes using demographic and clinical data from the first 12 months of treatment may provide a superior way to assess new LN therapies compared to relying on short-term, arbitrarily defined renal responses.

To apply these risk models to clinical trials, at 1 year the HITs for CKD, SKI and RRT would be used to compute a risk score for each individual patient for each adverse outcome. Mean risk scores for the

Accepted Article

placebo and active treatment arms of a trial can be calculated for each renal outcome and compared for significance using appropriate statistical tests. The risk of developing a future adverse renal outcome is increased with a higher HIT score. In a successful trial the experimental treatment would be expected to reduce the risk of an adverse renal outcome by a pre-specified percentage based on clinical importance. The trial would be powered to achieve the pre-specified risk reduction using the parameters for each HIT model given in Tables 2-4 to calculate sample size.

Of the three risk models, HIT-RRT is the best candidate for future validation as a surrogate clinical trial endpoint because it directly measures an outcome that matters to patients (a patient-oriented outcome). Because RRT events occur infrequently it may be reasonable to develop HIT-CKD as a surrogate trial endpoint. CKD is a major risk factor for future need of RRT, so this outcome likely also matters to patients. Until it can be shown that treatment effects on HIT-RRT or CKD reproducibly predict treatment effects on the actual need for future RRT or development of CKD, these HITs can only be considered as a reasonably likely surrogates (17). Although HIT-RRT and HIT-CKD cannot determine absolute risk of RRT or CKD, the relative risk for placebo versus actively treated patients can be estimated by computing the natural antilog of the difference in the mean HIT scores. To consider a clinical trial successful, it will be necessary to pre-specify the level of RRT or CKD risk improvement expected for a novel therapy.

All of the HIT equations can also be used in daily practice to estimate the probability of future CKD, SKI or RRT for individual LN patients following 12 months of treatment for an LN flare. This information could be useful when discussing prognosis with patients. Additionally, a proteinuria level of 0.7-0.8 g/d after 12 months of treatment was recently identified as the best predictor of long-term kidney survival compared to SCr and urine RBCs in a *post hoc* analysis of the ELNT and MAINTAIN LN trials (18, 19). These results were verified in an independent, ethnically diverse LN cohort (20). Similarly, proteinuria is a key component of all of our risk models, whereas urine RBCs are not. However, SCr does contribute significantly to risk modeling in all three of our HITs, and Black race, younger age and development of CKD in the prediction period are also significantly associated with risk of poor outcomes. Nonwhite race has previously been identified as a risk factor for poor outcome in LN (21, 22). Older age has also been associated with poor LN outcomes in adults (23, 24). In our study a small, statistically significant protective effect of increased age on RRT risk was found, however this effect was not considered clinically meaningful because the 95% confidence interval around its hazard ratio was so close to 1.0. Additionally the small effect of age may reflect the progressive increase in age at the time of LN diagnosis observed over the last several decades (25).

Although the predictors in our models are not novel, it is reassuring that clinical tests we have used to guide therapies do appear to be verified by our modeling. Our models are unique in that they incorporate data on kidney function over the first year of induction therapy and have been shown to significantly correlate with long-term outcome. Furthermore, the models provide a quantitative way to predict future outcomes that can be applied to trials and clinical practice. This has not been done before in a large multiethnic global cohort with validation in a large independent cohort.

This study has several limitations. Because data were mainly from real-world, investigator-initiated LN cohorts, specific treatments, data collection and follow-up intervals were not standardized. This led to exclusion of many patients, mainly for missing data, which could have biased the results. Random spot UPCR was measured in place of 24-hour proteinuria in several individuals. To avoid excluding even more patients, we used random spot UPCR to represent daily proteinuria despite persisting questions surrounding its accuracy (26). Variability in length of follow-up also precluded analysis beyond 48 months. Other variables that may have contributed to the HIT models, such as urine cellular casts, complement levels, auto-antibody titers, biopsy histologic features (e.g. activity, chronicity indices), blood pressure, and incident versus recurrent disease were not evaluated because they were available for only a limited number of patients. In the training cohort race was examined as black versus other, but the validation cohort was mainly Hispanic. Additional validation in other racially diverse cohorts would be desirable. Nonetheless, the fact that all of the outcome models validated well in spite of these differences in racial and ethnic composition speaks to the robust nature of the models and suggests they are applicable to a general LN population. Finally, the total number of adverse kidney outcomes, such as RRT was small, potentially affecting the reliability of the models. While standardized data collected in a prospective registry may optimize risk model development, such a registry does not yet exist. Our models were developed using the best currently available, prospectively collected datasets.

### **Acknowledgements**

This work was supported in part by the Kidney Health Initiative (KHI) and the Lupus Nephritis Trials Network (LNTN). The KHI is a public–private partnership between the American Society of Nephrology, the U.S. Food and Drug Administration, and more than 90 member organizations and companies to enhance patient safety and foster innovation in kidney disease. The LNTN is composed of 200 investigators from 40 countries who are committed to collaborative research dedicated to improving outcomes for people with lupus nephritis. KHI and LNTN funds were used to defray costs incurred during the conduct of the project, including project management support. However, there was no honorarium or other financial support provided to KHI/LNTN workgroup members (MM, MD, KK, JF, ML, BHR). The KHI/LNTN workgroup had final review authority and are fully responsible for its

content. KHI and LNTN make every effort to avoid actual, potential, or perceived conflicts of interest that may arise as a result of industry relationships or personal interests among the members of the writing committee and workgroup. More information on KHI, the workgroup, or the conflict of interest policy can be found at [www.kidneyhealthinitiative.org](http://www.kidneyhealthinitiative.org). More information on the LNTN can be found at [www.lupusnephritis.org](http://www.lupusnephritis.org).

KHI and LNTN would like to thank the U.S Food and Drug Administration, in particular the following individuals who the KHI workgroup consulted during this effort- Larissa Lapteva, MD, Nikolay Nikolov, MD, MD, Aliza Thompson, MD, and Sarah Yim, MD.

Dr. Hasni was supported by the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the national Institutes of Health.

Dr. Petri was supported in part by NIH Grants AR 43737 and AR 69572

We would like to thank Amrutha Baskaran, MD, MSCR, former post-doc University of California, San Diego, Laura Straub, former Strategic Review and Planning manager for the Immune Tolerance Network, San Francisco, and Elian D. Silverman, Kidney Health Initiative, American Society of Nephrology, Washington, D.C. for help with data management, organization and logistic support for this project.

#### **Disclaimer**

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the official policies of any KHI member organization, the U.S. Department of Veterans Affairs, or the U.S. Department of Health and Human Services, nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government.

#### **References**

1. Austin HA, 3rd, Klippel JH, Balow JE, le Riche NG, Steinberg AD, Plotz PH, et al. Therapy of lupus nephritis. Controlled trial of prednisone and cytotoxic drugs. *The New England journal of medicine*. 1986;314(10):614-9.
2. Illei GG, Austin HA, Crane M, Collins L, Gourley MF, Yarboro CH, et al. Combination therapy with pulse cyclophosphamide plus pulse methylprednisolone improves long-term renal outcome without adding toxicity in patients with lupus nephritis. *Annals of internal medicine*. 2001;135(4):248-57.
3. Boumpas DT, Austin HA, 3rd, Vaughn EM, Klippel JH, Steinberg AD, Yarboro CH, et al. Controlled trial of pulse methylprednisolone versus two regimens of pulse cyclophosphamide in severe lupus nephritis. *Lancet*. 1992;340(8822):741-5.
4. Gourley MF, Austin HA, 3rd, Scott D, Yarboro CH, Vaughan EM, Muir J, et al. Methylprednisolone and cyclophosphamide, alone or in combination, in patients with lupus nephritis. A randomized, controlled trial. *Annals of internal medicine*. 1996;125(7):549-57.

5. Almaani S, Meara A, Rovin BH. Update on Lupus Nephritis. *Clinical journal of the American Society of Nephrology* : CJASN. 2017;12(5):825-35.
6. Wofsy D, Hillson JL, Diamond B. Comparison of alternative primary outcome measures for use in lupus nephritis clinical trials. *Arthritis and rheumatism*. 2013;65(6):1586-91.
7. Renal Disease Subcommittee of the American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response C. The American College of Rheumatology response criteria for proliferative and membranous renal disease in systemic lupus erythematosus clinical trials. *Arthritis and rheumatism*. 2006;54(2):421-32.
8. Boumpas DT, Balow JE. Outcome criteria for lupus nephritis trials: a critical overview. *Lupus*. 1998;7(9):622-9.
9. Wofsy D, Hillson JL, Diamond B. Abatacept for lupus nephritis: alternative definitions of complete response support conflicting conclusions. *Arthritis and rheumatism*. 2012;64(11):3660-5.
10. Mok CC, Kwok RC, Yip PS. Effect of renal disease on the standardized mortality ratio and life expectancy of patients with systemic lupus erythematosus. *Arthritis and rheumatism*. 2013;65(8):2154-60.
11. Kilbride HS, Stevens PE, Eaglestone G, Knight S, Carter JL, Delaney MP, et al. Accuracy of the MDRD (Modification of Diet in Renal Disease) study and CKD-EPI (CKD Epidemiology Collaboration) equations for estimation of GFR in the elderly. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2013;61(1):57-66.
12. Pannu N. Bidirectional relationships between acute kidney injury and chronic kidney disease. *Current opinion in nephrology and hypertension*. 2013;22(3):351-6.
13. Zhong J, Yang HC, Fogo AB. A perspective on chronic kidney disease progression. *American journal of physiology Renal physiology*. 2017;312(3):F375-F84.
14. Levey AS, Inker LA, Matsushita K, Greene T, Willis K, Lewis E, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *American journal of kidney diseases : the official journal of the National Kidney Foundation*. 2014;64(6):821-35.
15. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Jama*. 1982;247(18):2543-6.
16. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*. 1996;15(4):361-87.
17. Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Statistics in medicine*. 2012;31(25):2973-84.
18. Tamirou F, Lauwerys BR, Dall'Era M, Mackay M, Rovin B, Cervera R, et al. A proteinuria cut-off level of 0.7 g/day after 12 months of treatment best predicts long-term renal outcome in lupus nephritis: data from the MAINTAIN Nephritis Trial. *Lupus science & medicine*. 2015;2(1):e000123.
19. Dall'Era M, Cisternas MG, Smilek DE, Straub L, Houssiau FA, Cervera R, et al. Predictors of long-term renal outcome in lupus nephritis trials: lessons learned from the Euro-Lupus Nephritis cohort. *Arthritis & rheumatology*. 2015;67(5):1305-13.
20. Ugolini-Lopes MR, Seguro LPC, Castro MXF, Daffre D, Lopes AC, Borba EF, 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 science & medicine*. 2017;4.

21. Lewis MJ, Jawad AS. The effect of ethnicity and genetic ancestry on the epidemiology, clinical features and outcome of systemic lupus erythematosus. *Rheumatology*. 2017;56 (suppl 1):i67-i77.
22. Hanly JG, O'Keefe AG, Su L, Urowitz MB, Romero-Diaz J, Gordon C, et al. The frequency and outcome of lupus nephritis: results from an international inception cohort study. *Rheumatology*. 2016;55(2):252-62.
23. Hanly JG, Su L, Urowitz MB, Romero-Diaz J, Gordon C, Bae SC, et al. A Longitudinal Analysis of Outcomes of Lupus Nephritis in an International Inception Cohort Using a Multistate Model Approach. *Arthritis & rheumatology*. 2016;68(8):1932-44.
24. Kang JH, Park DJ, Lee KE, Lee JS, Choi YD, Lee SS. Comparison of clinical, serological, and prognostic differences among juvenile-, adult-, and late-onset lupus nephritis in Korean patients. *Clinical rheumatology*. 2017;36(6):1289-95.
25. Moroni G, Vercelloni PG, Quaglini S, Gatto M, Gianfreda D, Sacchi L, et al. Changing patterns in clinical-histological presentation and renal outcome over the last five decades in a cohort of 499 patients with lupus nephritis. *Annals of the rheumatic diseases*. 2018.
26. Birmingham DJ, Rovin BH, Shidham G, Nagaraja HN, Zou X, Bissell M, et al. Spot urine protein/creatinine ratios are unreliable estimates of 24 h proteinuria in most systemic lupus erythematosus nephritis flares. *Kidney international*. 2007;72(7):865-70.

## Tables

Table 1. Origin of the data sets used in this study

Cohort/Trial Name and Country of Origin	Contributing Investigators
Inception Cohort-Mexico	Jorge Sanchez Guerrero, Juanita Romero Diaz
Hopkins Lupus Cohort-United States	Michelle Petri
Ohio SLE Study-United States	Brad H. Rovin
Dallas UT Southwestern Lupus Cohort-United States	Cristina Arriens
Lupus Nephritis Collaborative Study Group-United States	Edmund J. Lewis, Stephen M. Korbet <sup>1</sup>
Sapienza Lupus Cohort-Italy	Fabrizio Conti
Prague Lupus Cohort-Czech Republic	Vladimir Tesar, Zdenka Hruskova
University of Sao Paulo Lupus Cohort-Brazil	Eduardo F. Borba, Eloisa Bonfa
SickKids Lupus Cohort-Canada	Deborah Levy, Earl Silverman
Hong Kong Lupus Cohort-China	Daniel T. M. Chan
India Lupus Cohort-India	Manish Rathi, K. L. Gupta, Vivekanand Jha
National Institutes of Health-United States	Sarfaraz Hasni
Aspreva Lupus Management Study-Multinational	Neil Solomons
Euro-Lupus Nephritis Trial-Multinational	Frederic Houssiau
MAINTAIN Trial-Multinational	Frederic Houssiau
VALIDATION COHORT: Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran Cohort-Mexico	Juan M. Mejia-Vilet

<sup>1</sup>The Lupus Nephritis Collaborative Study Group included the following: Rush-Presbyterian-St. Luke's Medical Center, Chicago - E.J. Lewis, J.L. Roberts (deceased), M.M. Schwartz, R.A. Rodby, and H.L. Corwin; George Washington University, Washington, D.C. - J.M. Lachin, S-P. Lan, P. Cleary; William Beaumont Hospital, Royal Oak Mich. - J. Bernstein (deceased), H. Shapiro, and B.F. Rosenberg; Cleveland Clinic, Cleveland - M.A. Pohl, J.

Clough, and G. Gephardt; University of Colorado, Denver - T. Berl; Henry Ford Hospital, Detroit - N. Levin; University of Iowa, Iowa City - L.G. Hunsicker, S. Bonsib; Evanston Hospital, Evanston, Ill. - N. Simon and H. Friederici; Northwestern University, Chicago - F. del Greco and F.A. Carone (deceased); Ohio State University, Columbus - L. Hebert and H.M. Sharma; University of Pennsylvania, Philadelphia - E. Nielson and J. Tomazewski; Tufts - New England Medical Center, Boston - A. Levey and A. Ucci; Medical College of Wisconsin, Milwaukee - J. Lemann (deceased), S.S. Blumenthal, and J. Garancis; New York Medical College, Valhalla - K. Shapiro and P. Chander; West Virginia University, Morgantown - F. Whittier, J.W. Graves, J. Bathon, and R. Riley. Pathology Committee: M.M. Schwartz (Chairman) Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL; J. Bernstein (deceased), William Beaumont Hospital, Royal Oak, MI (deceased); G.H. Hill (deceased), Francis Scott Key Medical Institution, a Johns Hopkins Medical Institution, Baltimore, MD; K. Holley, Mayo Clinic, Rochester, MI.

Table 2. Demographics and Baseline Clinical Data for the Training and Validation Cohorts

Attribute	Training Cohort (n=550)	Validation Cohort (n=275)
Female	484 (88.0%) <sup>1</sup>	244 (88.7%)
Age, in years, (Mean ±SD)	33.2±12.1	30.2±9.6
White	283 (51.5%)	4 (1.5%)
Black	85 (15.5%)	0 (0%)
Asian	132 (24.0%)	0 (0%)
Race-Other	50 (9.1%)	271 (98.5%)
Hispanic	96 (17.5%)	274 (99.6%)
Non-Hispanic	324 (58.9%)	1 (0.4%)
Ethnicity not determined	130 (23.6%)	0 (0%)
ISN/RPS <sup>2</sup> Class 2+5	5 (0.9%)	0 (0%)
ISN/RPS Class 3	85 (15.5%)	17 (6.2%)
ISN/RPS Class 3+5	20 (3.6%)	59 (21.5%)
ISN/RPS Class 4	306 (55.6%)	47 (17.1%)
ISN/RPS Class 4+5	34 (6.2%)	123 (44.7%)
ISN/RPS Class 5	82 (14.9%)	29 (10.5%)
Biopsy class not determined	18 (3.3%)	0 (0%)
Urine RBC <sup>3</sup> (≥5/hpf <sup>4</sup> ) at baseline and 12 months	128 (23.3%)	91 (33.1%)

Urine RBC (<5/hpf) at baseline and 12 months	104 (18.9%)	63 (22.9%)
Urine RBC (≥5/hpf) at only one visit	181 (32.9%)	121 (44.0%)
Urine RBC not evaluated at one or both visits	137 (24.9%)	0 (0%)
Duration of follow-up, in months (median(IQR))	42.0(36.0-48.8)	52.0(36.0-72.0)
Serum creatinine (mg/dl) at baseline (Median (IQR))	1.0 (0.8-1.3)	1.0 (0.7-1.6)
eGFR <sup>5</sup> (ml/min/1.73 m <sup>2</sup> ) <sup>-1</sup> at baseline (median(IQR))	81.5 (54.1-110.3)	83.3 (45.2-117.6)
Proteinuria (g/d)at baseline (median(IQR))	3.0 (1.6-5.5)	3.4 (2.1-5.8)
Serum creatinine (mg/dl) at 12 months (median(IQR))	0.8 (0.7-1.0)	0.8 (0.6-1.1)
eGFR (ml/min/1.73 m <sup>2</sup> ) <sup>-1</sup> at 12 months (median(IQR))	95.6 (74.5-117.3)	104.8 (70.5 – 122.2)
Proteinuria (g/d) at 12 months (median(IQR))	0.5 (0.1-1.3)	0.6 (0.2-2.0)

<sup>1</sup>Unless otherwise state all data presented as N (%); <sup>2</sup>International Society of Nephrology/Renal Pathology Society; <sup>3</sup>Red blood cells; <sup>4</sup>High power field; <sup>5</sup>Estimated glomerular filtration rate

Table 3. Final multivariable Cox regression model for predicting chronic kidney disease in the lupus nephritis training cohort (n=550)

Factor	Parameter Estimate	Standard Error	p	Hazard Ratio	95% CI
log <sup>1</sup> proteinuria at 12 months <sup>2</sup>	0.432 <sup>3</sup>	0.115	0.0002	1.54	1.23, 1.93
log SCr at 12 months	2.180 <sup>3</sup>	0.304	<0.0001	8.84	4.88, 16.03
CKD status in prediction period (Yes vs No)	1.451 <sup>4</sup>	0.415	0.0005	4.27	1.89, 9.63

<sup>1</sup>A 1-unit increase on the natural logarithmic (log) scale equates to an approximate 2.72-fold increase on the raw scale

<sup>2</sup>A “start” of 0.01 was added to the proteinuria value to avoid taking logarithm of a zero value

<sup>3</sup>Parameter estimate reflects the magnitude of risk change per unit change in the predictor on the log scale

<sup>4</sup>Parameter estimate reflects the magnitude of risk change for a subject who acquired CKD in prediction period compared to one who did not

Table 4. Final multivariable Cox regression model for predicting severe kidney injury in the lupus nephritis training cohort (n=550)

Factor	Parameter Estimate	Standard Error	p	Hazard Ratio	95% CI
log <sup>1</sup> proteinuria at 12 months <sup>2</sup>	2.05 <sup>3</sup>	0.481	<0.0001	Refer to interaction term below	
ISN Class Proliferative vs Membranous <sup>4</sup>	2.307 <sup>5</sup>	0.837	0.0059	Refer to interaction term below	
log SCr at 12 months	1.457 <sup>3</sup>	0.303	<0.0001	4.29	2.37, 7.77
CKD status in prediction period (Yes vs no)	1.485 <sup>5</sup>	0.423	0.0005	4.41	1.93, 10.12
Black race	0.648 <sup>5</sup>	0.308	0.0354	1.912	1.05, 3.50
Interaction Term: ISN Class x log proteinuria at 12 months			0.0016		
Membranous <sup>6</sup>	2.049 <sup>5</sup>	0.481		7.76	3.02, 19.93
Proliferative <sup>6</sup>	0.477 <sup>5</sup>	0.132		1.61	1.24, 2.09

<sup>1</sup>A 1-unit increase on the natural logarithmic (log) scale equates to an approximate 2.72-fold increase on the raw scale

<sup>2</sup>A “start” of 0.01 was added to the proteinuria value to avoid taking log of a zero value

<sup>3</sup>Parameter estimate reflects the magnitude of risk change per unit change in the predictor on the log scale

<sup>4</sup>Mixed proliferative and membranous LN were included in as proliferative, along with cases without a biopsy

<sup>5</sup>Parameter estimate reflects the magnitude of risk change for a subject in one category compared to the reference category

<sup>6</sup>Hazard Ratios for each 1-unit increase in the log proteinuria (at 12 months), are presented at specific ISN Class designations

Table 5. Final multivariable Cox regression model for predicting the need for renal replacement therapy in the lupus nephritis training cohort (n=550)

Factor	Parameter Estimate <sup>1</sup>	Standard Error	p	Hazard Ratio	95% CI
log <sup>2</sup> proteinuria at 12 months <sup>3</sup>	0.717	0.212	0.0007	2.05	1.35, 3.10
log SCr at 12 months	2.318	0.369	<0.0001	10.16	4.92, 20.95
Age at study entry	-0.059	0.022	0.0070	0.94	0.90, 0.98

<sup>1</sup> Parameter estimates reflect the magnitude of risk change per unit change in the predictor on the log scale

<sup>2</sup> A 1-unit increase on the natural logarithmic (log) scale equates to an approximate 2.72-fold increase on the raw scale

<sup>3</sup> A "start" of 0.01 was added to the proteinuria value to avoid taking log of a zero value

Table 6. Risk Models for Adverse Kidney Outcomes

$$\text{Hazard Index for CKD}_i = 0.43231 * X_1 + 2.17960 * X_2 + 1.45107 * X_3$$

Where: i represents the i<sup>th</sup> subject in the data set, and for each subject:

- $X_1 = \log^1 (\text{proteinuria at 12 months} + 0.01)^2$
- $X_2 = \log \text{SCr at 12 months}$
- $X_3 = 1$  if reached CKD in prediction period; 0 if otherwise

$$\text{Hazard Index for SKI}_i = 2.04929 * X_1 + 1.45709 * X_2 + 0.64806 * X_3 + 1.48460 * X_4 + 2.30656 * X_5 - 1.57160 * X_1 * X_6$$

Where: i represents the i<sup>th</sup> subject in the data set, and for each subject:

- $X_1 = \log (\text{proteinuria at 12 months} + 0.01)$
- $X_2 = \log \text{SCr at 12 months}$
- $X_3 = 1$  if Black; 0 if Other
- $X_4 = 1$  if reached CKD in prediction period; 0 if otherwise
- $X_5 = 1$  if ISN Class is Membranous (5 or 2&5); 0 if otherwise

**Hazard Index for  $RRT_i = 0.71703 * X_1 + 2.31807 * X_2 - 0.05865 * X_3$**

Where:  $i$  represents the  $i^{th}$  subject in the data set, and for each subject:

- $X_1 = \log(\text{proteinuria at 12 months} + 0.01)$
- $X_2 = \log \text{SCr at 12 months}$
- $X_3 = \text{Age (in years) at study entry}$

<sup>1</sup>Log represents natural logarithm

<sup>2</sup>A “start” of 0.01 was added to the proteinuria value to avoid taking log of a zero value