External Validation of International Risk-Prediction Models of IgA Nephropathy in an Asian-Caucasian Cohort

Yuemiao Zhang1,2,3,4,8, Ling Guo1,2,3,4,8, Zi Wang1,2,3,4, Jinwei Wang1,2,3,4, Lee Er5, Sean J. Barbour5,6, Hernan Trimarchi7, Jicheng Lv1,2,3,4 and Hong Zhang1,2,3,4

1Renal Division, Department of Medicine, Peking University First Hospital, Beijing China; Institute of Nephrology, Peking University, Beijing, China; 2Key Laboratory of Renal Disease, Ministry of Health of China, Beijing, China; 3Key Laboratory of Chronic Kidney Disease Prevention and Treatment of Peking University, Ministry of Education, Beijing, China; 4Research Units of Diagnosis and Treatment of Immune-Mediated Kidney Diseases, Chinese Academy of Medical Sciences, Beijing, China; 5BC Renal, Vancouver, British Columbia, Canada; 6Division of Nephrology, University of British Columbia, Vancouver, British Columbia, Canada; and 7Nephrology Service, Hospital Británico de Buenos Aires, Argentina

Introduction: Two prediction models for IgA nephropathy (IgAN) using clinical variables and the Oxford MEST scores were developed and validated in 2 multiethnic cohorts. Additional external validation is required.

Methods: Biopsy-proven Chinese and Argentinian patients with IgAN were included. The primary outcome was defined as a 50% decline in estimated glomerular filtration rate (eGFR) or end-stage renal disease. C-statistics and stratified analyses were used for model discrimination, coefficient of determination ($R^2_D$) for model fit, and calibration plots for model calibration. Baseline survival function was also evaluated.

Results: A total of 1275 patients were enrolled, with a mean age of 34 (interquartile range: 27–42) years, 50% of whom (638 of 1275) were men. Use of renin-angiotensin system blockers was higher than in previously reported cohorts, whereas other variables were comparable. The C-statistic of the models was 0.81, and $R^2_D$ was higher than reported. Survival curves in the subgroups (<16th, ~16th to <50th, ~50th to <84th, and ≥84th percentiles of linear predictor) were well separated. Most of the predictor variables, including hazard ratio, predicted 5-year risk, and eGFR decline slope, were worse with risk increasing. The baseline survival function was comparable in our cohort and the reported cohorts. The calibration was acceptable for the full model without race. However, the risk probability over 3 years was overestimated in the full model with race included.

Conclusion: The prediction models showed good performance on personalized risk assessment, which may be used as drug-specific, precision-medicine approaches to treatment decisionmaking.

Kidney Int Rep (2020) 5, 1753–1763; https://doi.org/10.1016/j.ekir.2020.07.036

KEYWORDS: calibration; discrimination; external validation; IgA nephropathy; prediction models © 2020 International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

IgA nephropathy (IgAN) is the most common form of primary glomerulonephritis worldwide and is characterized by clinical heterogeneity and ethnic variation.1,2 Nearly 10% and >20% of patients progress to end-stage renal disease (ESRD) within 5 and 20 years after diagnosis, respectively.3,4 Although a number of clinical variables, such as eGFR, proteinuria, and blood pressure, have been suggested as reliable prognostic factors of IgAN, none has accurately discriminated the disease risk individually or jointly.5,6 The diagnosis of IgAN depends on renal biopsy. The Oxford MEST (M = mesangial hypercellularity, E = endothelial hypercellularity, S = segmental sclerosis, T = tubular atrophy and interstitial fibrosis) histologic scores for IgAN were derived from a multiethnic population7 with high reproducibility8 and can provide prognosis evaluation.9,10 Thus, a prediction model combining clinical features and histologic scores would be valuable to better discriminate patients with a progressive disease course and permit assessment of targeted

Correspondence: Jicheng Lv, Renal Division, Department of Medicine, Peking University First Hospital, Peking University Institute of Nephrology, No. 8, Xishiku Street, Xicheng District, Beijing 100034, China. E-mail: jichenglv75@gmail.com; or Hernan Trimarchi, Nephrology Service, Hospital Británico de Buenos Aires, Perdriel 74 (1280), Buenos Aires, Argentina. E-mail: htrimarchi@hotmail.com

8YMZ and LG contributed equally to this research.

Received 7 July 2020; accepted 28 July 2020; published online 7 August 2020
therapies in clinical trials and risk stratification in clinical practice.

Recently, the International IgA Nephropathy Network took advantage of the existing large multi-ethnic cohorts with long-term follow-up to develop and validate 2 full prediction models without and with race included.11 The 2 full models have the advantage of simplicity, being derived from factors routinely available, and have shown sufficient discrimination and calibration. The authors also provided a mobile app (QxMD) and a web-based calculator (https://qxmd.com/calculate-by-qxmd) for clinical use. We believe the development of such models in patients from diverse ethnic backgrounds will aid clinicians in patient stratification, treatment decisionmaking, clinical trial recruitment, and biomarker validation.

As emphasized in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement,12 for all the prediction models, development and validation is only a first step. For any prediction model to be widely applicable, it must be validated in cohorts outside that in which it was derived but similar to the target population. Clearly, more validation studies detecting the generalizability of prediction models will be required.13,14 Although the full models performed well in the reported validation cohort, additional external validation would provide further evidence of prediction model performance in target populations. Notably, as it was stated that the study cohorts included patients from an “old era” (the 1980s), during which only about 30% of patients received renin-angiotensin system blockers (RASBs) at biopsy and a further 30% to 50% had these drugs added during follow-up. Because RASBs have greatly changed the progression of IgAN, in current clinical practice nearly all patients with IgAN routinely receive RASBs at or soon after biopsy, as recommended by the 2012 Kidney Disease International Global Outcomes (KDIGO) guideline.15 In recent large, randomized, controlled trials, including the Supportive Versus Immunosuppressive Therapy for the Treatment of Progressive IgA Nephropathy (STOP-IgAN),16 Targeted-release Budesonide Versus Placebo in Patients with IgA Nephropathy (NEFIGAN),17 and Therapeutic Evaluation of Steroids in IgA Nephropathy Global (TESTING)18 studies, approximately 90% patients were given RASBs. Recognizing that these models may be used in more current populations, the prediction performance of this model needs to be validated in cohorts from a “new era,” especially those consistent with the current KDIGO guideline.19 This will help clinicians to predict the risk of failure for routine treatment and to determine the level of treatment with immunosuppressants. Risk stratification is essential for the care of patients with IgAN to avoid unnecessary exposure to toxic therapies while reducing the risk of chronic kidney disease progression. It is also of vital importance for patients and physicians in making optimal health-related and life decisions.

Thus, in this study, we further evaluated the performance of 2 full models without and with race in a
large external cohort of IgAN patients from northern China and Argentina.

**METHODS**

**Patients**
To validate the full model without or with race, we enrolled an Asian-Caucasian cohort, including 1360 Chinese patients who were registered and with long-term follow-up in the Peking University First Hospital IgAN database (www.renal-online.org) since January 2003 and 116 patients with a long-term follow-up who were diagnosed and treated at the Hospital Británico de Buenos Aires since 1995. All patients were diagnosed by biopsy and those with <8 glomeruli per biopsy section were excluded. Our study was approved by the ethics committee of Peking University First Hospital and by the institutional review board of the Hospital Británico de Buenos Aires. Written informed consent was provided by all participants.

**Variable Definitions**
Baseline characteristics, including proteinuria, systolic blood pressure (SBP), diastolic blood pressure (DBP), eGFR (calculated using the Chronic Kidney Disease Epidemiology Collaboration formula\(^{20}\)), age, and previous use of RASBs and/or immunosuppression, were collected at the time of biopsy. MEST histologic scores were evaluated according to the Oxford classification system\(^{21}\) by 3 pathologists independently, who were blinded to clinical data. Mean arterial pressure (MAP) was calculated as \(1/3 \times \text{SBP} + 2/3 \times \text{DBP}\). The primary outcome was a composite of the first occurrence of either ESRD (eGFR <15 ml/min per 1.73 m\(^2\), dialysis, or kidney transplantation) or a 50% decline in eGFR from the value at biopsy. The decline slope of eGFR was calculated using a mixed-effects model as reported elsewhere.\(^{11}\) For validation, each covariate and outcome were defined exactly according to the original publication using the same measurement units (Table 1).

**Prediction Models for External Validation**
The published prediction models for validation were derived as follows\(^{11}\):

\[
\text{Predicted risk (time} \ t = \ 1 - S_0(t)^{\text{Exp(linear predictor)}}
\]

(1) For the full model without race:

\[
\begin{align*}
\text{Linear predictor} &= -0.320 \times ([\text{sqt(eGFR)}] - 8.8) \\
&+ 0.002 \times (\text{MAP} - 97) - 0.035 \\
&\times [\log(\text{proteinuria} - 0.09)] \\
&+ [(\text{MAP} \times \log(\text{proteinuria}))] \\
&- 8.73 + 0.201 \times M1 - 0.035 \\
&\times E1 + 0.084 \times S1 + 0.700 \\
&\times T1 + 1.237 \times T2 + 0.101 \\
&\times T1 \times \log(\text{proteinuria}) - 0.017 \\
&\times (\text{age} - 38) + 0.118 \times \text{RASB} \\
&+ 0.166 \times \text{RASB} \times \log(\text{proteinuria}) \\
&- 0.266 \times \text{immunosuppression}
\end{align*}
\]

\[
S_0(t) = 1.0003754 - 0.1131641 \\
\times (t + 0.1)^2 + 0.0964763 \\
\times (t + 0.1)^2 \times \log((t + 0.1) / 100).
\]

(2) For the full model with race:

\[
\begin{align*}
\text{Linear predictor} &= -0.351 \times [\text{sqt(eGFR)}] - 8.8 \\
&+ 0.002 \times (\text{MAP} - 97) - 0.093 \\
&\times [\log(\text{proteinuria} - 0.09)] \\
&+ [(\text{MAP} \times \log(\text{proteinuria}))] \\
&- 8.73 + 0.201 \times M1 - 0.131 \\
&\times E1 + 0.097 \times S1 + 0.607 \\
&\times T1 + 1.189 \\
&\times T2 + 0.109 \times T1 \\
&\times \log(\text{proteinuria}) - 0.339 \\
&\times T2 \times \log(\text{proteinuria}) - 0.016 \\
&\times (\text{age} - 38) + 0.246 \times \text{RASB} \\
&+ 0.166 \times \text{RASB} \times \log(\text{proteinuria}) \\
&- 0.225 \times \text{immunosuppression} \\
&- 0.396 \text{(if } t \leq \text{36 months)} \\
&+ 0.818 \text{(if } t > \text{36 months)}.
\end{align*}
\]

\[
S_0(t) = 1.9964303 + 0.04392517 \\
\times ([t + 0.1] / 100)^{0.5} - 0.1257002 \\
\times ([t + 0.1] / 100)
\]

where log is the natural log function.

**Statistical Analysis**
For model validation, we initially calculate the linear predictor for each patient in the current cohort based on the exact predictors and coefficient values as
mentioned. We then assessed the model performance of discrimination and calibration according to Royston and Altman.\textsuperscript{22}

For discrimination, we first estimated the regression coefficient on the linear predictor coefficient by fitting a Cox proportional hazards model for the full model without race and an interval format Cox proportional hazards model\textsuperscript{23} for the full model with race in our data set. If $\beta_{\text{linear predictor}} \geq 1$, then discrimination of the models would be acceptable. Second, the C-statistic was used to determine how well the model could distinguish those with an endpoint from those without an endpoint. Considering the censoring issue, the AUC.cd, according to Chambless and Diao, was calculated accordingly.\textsuperscript{24} Coefficient of determination ($R^2_D$) was calculated according to a method evaluating model fit performance.\textsuperscript{25} Third, we divided patients into risk groups, including $<16$th (low risk), $\sim 16$th to $<50$th (intermediate risk), $\sim 50$th to $<84$th (higher risk), and $\geq 84$th (the highest risk) percentiles of the linear predictor from the full model without or with race. Subgroup analyses were performed and survival curves derived. As suggested, in contrast to $P$ values for comparing risk groups, the hazard ratios were suggested to be a sensible verification of model discrimination.\textsuperscript{22} Thus, hazard ratios were evaluated by fitting a Cox model with a dummy variable representing each risk group referring to the lowest risk group. When survival curves are more widely separated, the hazard ratio tends to be greater.

For calibration, because it is preferred to have patients with similar baseline risks on average, we first investigated the accuracy of the baseline survival function itself. The reported baseline survival function was obtained directly from the publication. A Kaplan-Meier–like estimate of the baseline survival function in our data was obtained by standard methods after fitting a Cox model with no covariates other than the linear predictor, with regression coefficient constrained to 1. We then applied the averaging method to obtain predicted mean survival curves in our cohort and compared them with the Kaplan-Meier survival curves in the risk groups. Finally, we assessed calibration graphically by

\textbf{Figure 1.} (a) Enrollment flowchart and (b) cumulative incidence of the primary outcome in the current cohort. Overall, 1275 of the original 1476 patients remained in the final cohort, including 1169 Chinese patients and 106 Argentinian patients. Among the excluded patients, 92 had other forms of glomerulopathy, 59 were $<18$ years old, 12 had end-stage renal disease at the time of renal biopsy, 22 were without available MEST score, and 16 lacked medication information. ESRD, end-stage renal disease; MEST, $M =$ mesangial hypercellularity, $E =$ endothelial hypercellularity, $S =$ segmental sclerosis, $T =$ tubular atrophy and interstitial fibrosis; RASB, renin-angiotensin system blocker.
comparing the predicted survival probability with the observed percentage stratified by deciles and calculated Kaplan-Meier statistics using val.surv embedded in the “rms” package in R software. All statistical analyses were performed using R version 3.3.0 (https://www.R-project.org/).

**RESULTS**

**Baseline Characteristics and Outcomes**

The flowchart for patient enrollment is shown in Figure 1a. There were 1275 patients enrolled in the analysis, including 1169 Chinese and 106 Argentinians. In our cohort, the percentage of combined outcomes was 14.2%. Among these, 13.5% achieved halving of their eGFR and 8.6% reached ESRD during the median 3.8-year follow-up. All rates were lower than the reported derivation and reported validation cohorts due to our relatively shorter follow-up period ($P < 0.05$, considering the general clinical slow progressive course of IgAN) (Figure 1b). The patient characteristics are displayed in Table 2. As stated earlier, the rates of RASB application ranged from 30.0% to 32.4% at biopsy and 66.4% to 86.7% after biopsy in the previously reported cohorts, but were 72.6% at biopsy and reached 91.3% after biopsy in our cohort. Similarly, rates of immunosuppressant use ranged from 7.1% to 9.1% in the earlier cohorts and

| Characteristics | Reported derivation cohort | Reported validation cohort | Our validation cohort |
|-----------------|---------------------------|---------------------------|----------------------|
| Number of patients | 2781 | 1146 | 1275 |
| Follow-up time, median (IQR), yr | 4.8 (3.0–7.6) | 5.8 (3.4–8.5) | 3.8 (2.1–6.9) |
| Year of biopsy, median (IQR) | 2006 (2004–2008) | 1998 (1993–2003) | 2010 (2006–2013) |
| Age, median (IQR), yr | 35.6 (28.2–45.4) | 34.8 (26.9–45.0) | 34 (27–42) |
| Male, n (%) | 1608 (57.8) | 565 (49.3) | 638 (50.0) |
| Race, n (%) | 1167 (42.0) | 176 (15.5) | 106 (8.3) |
| Caucasian | 1021 (36.7) | 292 (25.8) | 1169 (91.7) |
| Chinese | 569 (20.5) | 616 (54.4) | — |
| Other | 22 (0.8) | 49 (4.3) | — |
| SCr at biopsy, median (IQR), μmol/l | 92.0 (70.7–123.8) | 84.0 (66.2–111.4) | 90.20 (71.3–120.2) |
| eGFR at biopsy, median (IQR), ml/min per 1.73 m² | 83.0 (56.7–108.0) | 89.7 (65.3–112.7) | 82.8 (59.9–104.6) |
| <30, n (%) | 142 (5.1) | 37 (3.2) | 44 (3.5) |
| ~30–60, n (%) | 657 (23.6) | 191 (16.7) | 276 (21.6) |
| ~60–90, n (%) | 800 (28.0) | 350 (30.5) | 417 (32.7) |
| ≥90, n (%) | 1182 (42.5) | 568 (49.6) | 538 (42.2) |
| MAP at biopsy, median (IQR), mm Hg | 96.7 (88.7–106.3) | 93.3 (85.0–103.3) | 93.3 (86.7–100.0) |
| Proteinuria at biopsy, median (IQR), g/d | 1.2 (0.7–2.2) | 1.3 (0.6–2.4) | 1.2 (0.7–2.3) |
| <0.5, n (%) | 383 (13.9) | 221 (19.4) | 217 (17.0) |
| ~0.5–1, n (%) | 772 (28.1) | 209 (18.3) | 306 (24.0) |
| ~1–2, n (%) | 817 (29.7) | 352 (30.6) | 370 (29.0) |
| ~2–3, n (%) | 360 (13.1) | 145 (12.7) | 158 (12.4) |
| ≥3, n (%) | 415 (15.1) | 218 (18.8) | 224 (17.6) |
| MEST histologic score, n (%) | 1064 (38.0) | 481 (42.0) | 570 (44.7) |
| M1 | 478 (17.3) | 478 (41.5) | 385 (30.2) |
| E1 | 2137 (77.0) | 912 (79.6) | 768 (60.2) |
| S1 | 686 (24.7) | 207 (18.1) | 306 (24.0) |
| T1 | 128 (4.8) | 122 (10.6) | 112 (8.8) |
| RASB use, n (%) | 862 (32.4) | 320 (30.0) | 926 (72.6) |
| At biopsy | 2400 (86.7) | 708 (66.4) | 1164 (91.3) |
| Immunosuppressant use, n (%) | 252 (9.1) | 81 (7.1) | 142 (11.1) |
| At biopsy | 1209 (43.5) | 359 (31.3) | 432 (33.9) |
| Primary outcome*, n (%) | 420 (15.1) | 210 (18.3) | 173 (13.5) |
| 50% decline in eGFR | 372 (13.4) | 155 (13.5) | 110 (8.6) |
| ESRD | 492 (17.7) | 213 (18.6) | 181 (14.2) |
| Total primary outcomes | 50% decline in eGFR and 8.6% reached ESRD during the median 3.8-year follow-up. All rates were lower than the reported derivation and reported validation cohorts due to our relatively shorter follow-up period ($P < 0.05$, considering the general clinical slow progressive course of IgAN) (Figure 1b). The patient characteristics are displayed in Table 2. As stated earlier, the rates of RASB application ranged from 30.0% to 32.4% at biopsy and 66.4% to 86.7% after biopsy in the previously reported cohorts, but were 72.6% at biopsy and reached 91.3% after biopsy in our cohort. Similarly, rates of immunosuppressant use ranged from 7.1% to 9.1% in the earlier cohorts and

---

eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; IQR, interquantile range; MAP, mean arterial pressure; MEST, M = mesangial hypercellularity, E = endothelial hypercellularity, S = segmental sclerosis, T = tubular atrophy or interstitial fibrosis; RASB, renin-angiotensin system blocker; SCr, serum creatinine.

*Total primary outcomes defined as the first event of either 50% decline in eGFR or ESRD.
was 11.1% in our cohort at biopsy. The distributions of other clinical parameters, including age, male ratio, baseline eGFR, blood pressure, proteinuria, and Oxford MEST histologic scores, were broadly similar between the present and previously reported cohorts. The baseline features of Chinese and Argentinian patients in our cohort are presented separately in Supplementary Table S1.

Regression on Linear Predictor in Validation Data
The calibration slopes of linear prediction were 0.87 and 0.89 for the full model without and with race, respectively. Although the slopes were smaller than those in the other cohorts, they were close to 1 and not significantly different from 1 ($P > 0.05$; Table 3), so the discrimination appeared to be preserved.

| Group                                      | Regression slope on linear prediction | C-statistic      | $R^2_D$ (%) |
|--------------------------------------------|--------------------------------------|-----------------|-------------|
| Full model without race                    | 0.81 (0.80–0.81)                     | 0.81 (0.80–0.81) | 25.3        |
| Reported derivation cohort                 | 0.82 (0.81–0.82)                     | 0.82 (0.81–0.82) | 26.3        |
| Reported validation cohort                 | 1.12                                 | 0.82 (0.81–0.83) | 35.3        |
| Our validation cohort                      | 0.89 (0.75–1.00)                     | 0.81 (0.81–0.82) | 42.2        |

Measures of Discrimination and Model Fit
When the reported models were applied directly to our current cohort, the C-statistic was 0.81 for both full models without and with race, with similar values in the reported cohorts. In addition, $R^2_D$ values were 37.6% and 42.2% for the full model without and with race in our cohort, indicating an increase compared with ~25% in the reported cohorts, suggesting good performance of the model fit (Table 3).

Comparison of Risk Groups
Figure 2 shows Kaplan-Meier curves according to risk groups based on the percentiles of the linear predictor ($<16$th for low-risk group [red], $\sim 16$th to $<50$th for intermediate-risk group [green], $\sim 50$th to $<84$th for higher risk group [blue], and $\geq 84$th for the highest risk group [purple]). We found that the Kaplan-Meier curves in the risk groups were well separated, especially those for the highest risk group.

Figure 2. Kaplan-Meier curves for survival probability of primary outcome in 4 risk groups based on percentile of the linear predictor. Full model without race (a). Full model with race (b). The 4 risk groups were defined as $<16$th (low risk), $\sim 16$th to $<50$th (intermediate risk), $\sim 50$th to $<84$th (higher risk), and $\geq 84$th (the highest risk) percentiles of the linear predictor from the full model without and with race, respectively.
group of the 2 full models, confirming our earlier conclusion that the models have preserved discrimination. By visual comparison of our validation results with the original publication, the discrimination was broadly similar, but the full model with race seemed less able to distinguish between the 2 lowest risk groups in our validation cohort. Accordingly, the hazard ratios between risk groups were well-maintained, confirming the impression in Figure 2. The predicted 5-year risks for patients in the 4 groups defined in our cohort were 2.51%, 5.37%, 15.57%, and 48.13% for the full model without race, and 5.22%, 10.25%, 25.81%, and 61.22% for the full model with race, respectively. Similarly, the eGFR decline slopes in the 4 groups were $-0.67$, $-1.56$, $-2.27$, and $-3.32$ for the full model with race, and $-1.67$, $-1.87$, $-1.95$, and $-3.72$ for the full model without race (Table 4). The demographic characteristics of the patients in the 4 risk groups based on the full model without/with race are presented in Supplementary Tables S2 and S3, respectively. In addition, we found that all predictor variables were worse with increasing risk, such as more proteinuria, more Oxford MEST lesions, and lower eGFR.

Model Calibration

As shown in Figure 3, the Kaplan-Meier–like estimates of baseline survival function estimated after fitting the model were similar to the reported ones. Model calibration performance was obtained by comparing observed and predicted risks of primary outcome over the duration of follow-up. For the full model without race (Figure 4a), the predicted curves (red) showed an acceptable fit with the observed curves (black). For the full model with race, however, calibration was better earlier on, whereas the predicted risk curves displayed a creep at 3 years in all risk groups (Figure 4b). Consistent with this, a comparison of 5-year predicted and observed survival probability of primary outcome is shown in Figure 5. Results at 1, 2, 3, and 4 years are presented in Supplementary Figures S1 and S2. Differences between observed and predicted survival probabilities

---

### Table 4. Hazard ratios, mean predicted 5-year risk of primary outcome, and rate of kidney function decline in subgroups based on linear predictor

| Measure                  | Hazard ratio | Mean predicted 5-yr risk, % | eGFR decline slope |
|--------------------------|--------------|-----------------------------|--------------------|
| Full model without race  |              |                             |                    |
| Low-risk group           | Reference    | 2.51                        | −1.67              |
| Intermediate-risk group  | 2.57 (1.16–5.72) | 5.37                        | −1.87              |
| Higher risk group        | 3.85 (1.75–8.46) | 15.57                       | −1.95              |
| Highest risk group       | 20.35 (9.29–44.57) | 48.13                       | −3.72              |
| Full model with race     |              |                             |                    |
| Low-risk group           | Reference    | 5.22                        | −0.67              |
| Intermediate-risk group  | 1.33 (0.55–3.23) | 10.25                       | −1.56              |
| Higher risk group        | 2.97 (1.28–6.87) | 25.81                       | −2.27              |
| Highest risk group       | 8.29 (3.84–20.71) | 61.22                       | −3.32              |

eGFR, estimated glomerular filtration rate. Subgroups were <16th (low risk), 16th to <50th (intermediate risk), 50th to <84th (higher risk), and ≥84th (the highest risk) percentiles of the linear predictor from the full models without and with race, respectively.

---

Figure 3. Estimates of the baseline survival function in the reported and the current data sets. Full model without race (a). Full model with race (b). Baseline survival function of the reported data set is shown in red, and the baseline survival function of the current data set is shown in blue.
of primary outcome over 3 years were apparent for the full model with race.

**DISCUSSION**

The 2 reported full prognostic models without and with race\(^1\) provided us with a useful prediction tool for clinical IgAN progression. In this study, we further evaluated model performance using an external cohort from China and Argentina. For the full models without race and with race, good discrimination (C-statistic >0.80) was observed and the models fit well. The survival curves of patients stratified by percentiles of linear predictor were well separated. Our external validation provides further evidence that the clinical and pathologic variables used in the model appeared to be sufficient for patient discrimination. The full model without race showed acceptable calibration. Although the full model with race showed a similar regression slope on linear predictor (0.89), it seemed to overestimate the prognostic risk over 3 years.

Compared with the vigorous progress in model development, only about 25% (32 of 127) of them have been validated and few have been used in clinical practice.\(^2\) Although there is an increasing number of clinical prediction models for IgAN, most were performed in single ethnicity and few were approved to accurately identify high-risk patients.\(^25-30\) Considering that the diagnosis of IgAN depends on renal biopsy, a prediction model with histologic variants would help increase the model accuracy. However, although there were some prediction models developed with pathologic variables, these models either developed with a relatively earlier cohort or in a single population. Moreover, some of these models used different pathologic scoring systems that are not widely used. For example, Goto et al derived a prediction model based on the Japanese population, with both clinical variables and pathologic variables that were not internationally committed.\(^11,12\) Chen et al used the XGBoost system and stepwise Cox regression to develop a prediction model, based on Chinese patients at different centers.\(^13\) The predictors included demographic, clinical, and pathologic variables. However, this model was based on single ethnic population, and external validation based on non-Asian population is not available. In this context, 2 full models integrating clinical variables and Oxford MEST histologic scores were derived and validated in 2 multiple ethnic cohorts.\(^11\) The well-established factors for disease progression of IgAN, including eGFR, proteinuria, blood pressure, Oxford MEST histologic scores, age, and use of RASB/immunosuppressant, could be “easy” and consistently obtained in clinical practice, demonstrating its potential in clinical practice.

In this study, we have further performed external validation of the full models in a Chinese-Argentinian cohort from a relatively “new era.” Although it was shown that 86% patients received RASBs after

![Figure 4](image-url)
diagnosis in the reported derivation cohort, which is not that different from the 91% in our current cohort, the rate of immunosuppressant use between the reported and our current cohort were also similar after biopsy. Compared with the reported cohorts, there was a significantly higher rate of RASB initiation before diagnosis (30% vs. 70%). This strengthens the analysis because our current cohort is much more representative of patients receiving the current treatment regimens. We found that the prediction models consistently performed well for discrimination in our validation cohort. Our subgroup analysis suggests that survival curves of different risk groups were quite well-separated in both models. Accordingly, the eGFR decline slope was relatively larger with risk increasing. This was consistent with the worse predictor variables, such as more proteinuria, more Oxford MEST lesions, and lower eGFR, across the risk groups. In this way, the “simple, robust” models validated in our cohort could be applied to improve risk prediction, which was approved to both increase treatment allocation to patients at high risk of disease progression and avoid treatment in patients with nonprogressive disease. Moreover, for the full model without race, the calibration was acceptable. However, for the full model with race, it seemed to overestimate the prediction risk over 3 years in our cohort. Considering that, for a given high-risk clinical decision, a well-calibrated model providing a wider risk stratification is likely to have greater clinical utility, we suggest using the full model without race for further assessment of setting the thresholds, which is used to estimate the benefits and costs of specific interventions.

The strength of this study is that we did external validation based on a Chinese/Argentinian population from a relatively “new era,” which enabled us to validate the performance of the full models in the populations with treatment under current guidelines. However, there are also some limitations. First, our final cohort excluded patients who did not have a renal biopsy performed or for whom the Oxford MEST histologic scores were not available, meaning that we may have missed some very high-risk patients because the Oxford MEST histologic scores were hard to evaluate due to few glomeruli in the biopsy. Second, as there was a large proportion of Chinese patients in our validation cohort, the model performance in other new ethnic populations and recalibration of the full model with race are to be evaluated in the future. Third, a limitation of the prediction model is that, taking into account that IgAN is an entity (not a disease) with a long-term evolution, the model offers only short-term prognosis, up to 8 years at the most.

In summary, we externally validated the full prediction models to risk stratify patients after an initial diagnosis of IgAN. The prediction models showed good performance on personalized risk assessment, which will help allocate immunosuppression to those patients at high risk of disease progression and avoid treatment in those with nonprogressing disease.
DISCLOSURE

All the authors declared no competing interests.

ACKNOWLEDGMENTS

This study was supported by grants from the National Natural Science Foundation of China (81800636), the Training Program of the Major Research Plan of the National Natural Science Foundation of China (91642120), the Natural Science Foundation for Innovation Research Group of China (81621092), the University of Michigan Health System–Peking University Health Science Center Joint Institute for Translational and Clinical Research (BMU2017JI007), and the CAMS Innovation Fund for Medical Sciences (2019-I2M-5-046).

AUTHOR CONTRIBUTIONS

YMZ and JCL designed the study; LG, ZW, and HT acquired and cleaned the data; YMZ, JWW, LEM, and SJB analyzed and interpreted the data; and YMZ and LG drafted the manuscript. All authors assisted in revision of the work and approved the final version submitted for publication.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Table S1. Description and comparison of the Chinese and Argentinian patients in our cohort.

Table S2. Demographic features in subgroups based on percentile of the linear predictor of the full model without race.

Table S3. Demographic features in subgroups based on percentile of the linear predictor of the full model with race.

Figure S1. Calibration curves depicting the predicted-vs.-observed survival probability of the 1- to 4-year primary outcome for the full model without race.

Figure S2. Calibration curves depicting the predicted-vs.-observed survival probability of the 1- to 4-year primary outcome for the full model with race.

REFERENCES

1. Lai KN, Tang SC, Schena FP, et al. IgA nephropathy. Nat Rev Dis Primers. 2016;2:16001.
2. Yeo S, Goh S, Barratt J. Is immunoglobulin A nephropathy different in different ethnic populations? Nephrology. 2019;24:885–895.
3. Li X, Liu Y, Lv J, et al. Progression of IgA Nephropathy under current therapy regimen in a Chinese population. Clin J Am Soc Nephrol. 2014;9:484–489.
4. Magistroni R, D’Agati VD, Appel GB, et al. New developments in the genetics, pathogenesis, and therapy of IgA nephropathy. Kidney Int. 2015;88:974–989.
5. Lv J, Zhang H, Zhou Y, et al. Natural history of immunoglobulin A nephropathy and predictive factors of prognosis: a long-term follow up of 204 cases in China. Nephrology (Carlton). 2008;13:242–246.
6. Barbour SJ, Reich HN. Risk stratification of patients with IgA nephropathy. Am J Kidney Dis. 2012;59:865–873.
7. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society, Cattran DC, Coppo R, et al. The Oxford classification of IgA nephropathy: rationale, clinicopathological correlations, and classification. Kidney Int. 2009;76:534–545.
8. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society, Roberts IS, Cook HT, Troyanov S, et al. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. Kidney Int. 2009;76:546–556.
9. Barbour SJ, Espino-Hernandez G, Reich HN, et al. The MEST score provides earlier risk prediction in IgA nephropathy. Kidney Int. 2016;89:167–175.
10. Schimpf JI, Klein T, Fitzner C, et al. Renal outcomes of STOP-IgAN trial patients in relation to baseline histology (MEST-c scores). BMC Nephrol. 2018;19:328.
11. Barbour SJ, Coppo R, Zhang H, et al. Evaluating a new international risk-prediction tool in IgA nephropathy. JAMA Intern Med. 2019;179:942–952.
12. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
13. Siontis GC, Tsoulaki I, Castaldi PJ, et al. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015;68:25–34.
14. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1–W73.
15. Kidney Disease: Improving Global Outcomes (KDIGO) Glomerulonephritis Work Group. KDIGO clinical practice guideline for Glomerulonephritis. Kidney Int Suppl. 2012;2:139–274.
16. Rauen T, Eitner F, Fitzner C, et al. Intensive supportive care plus immunosuppression in IgA nephropathy. N Engl J Med. 2015;373:2225–2236.
17. Fellström BC, Barratt J, Cook H, et al. Targeted-release budesonide versus placebo in patients with IgA nephropathy (NEFIGAN): a double-blind, randomised, placebo-controlled phase 2b trial. Lancet. 2017;389:2117–2127.
18. Lv J, Zhang H, Wong MG, et al. Effect of oral methylprednisolone on clinical outcomes in patients with IgA nephropathy. JAMA. 2017;318:432–442.
19. Floege J. A new tool to predict the risk of progression in IgA nephropathy. Kidney Int. 2019;96:808–809.
20. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150:604–612.
21. Trimmer HC, Barratt J, Cattran DC, et al. Oxford classification of IgA nephropathy 2016: an update from the IgA Nephropathy Classification Working Group. Kidney Int. 2017;91:1014–1021.
22. Royston P, Altman D. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013;13:33.
23. Thomas L, Reyes E. Tutorial: Survival estimation for Cox regression models with time-varying coefficients using SAS and R. J Stat Softw. 2014;61(Code Snippet 1).
24. Chambless L, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med*. 2006;25:3474–3486.

25. Berthoux F, Mohey H, Laurent B, et al. Predicting the risk for dialysis or death in IgA nephropathy. *J Am Soc Nephrol*. 2011;22:752–761.

26. Knoop T, Vågane AM, Vikse BE, et al. Addition of eGFR and age improves the prognostic absolute renal risk-model in 1,134 Norwegian patients with IgA nephropathy. *Am J Nephrol*. 2015;41:210–219.

27. Pesce F, Diciolla M, Binetti G, et al. Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. *Nephrol Dial Transplant*. 2016;31:80–86.

28. Liu J, Duan S, Chen P, et al. Development and validation of a prognostic nomogram for IgA nephropathy. *Oncotarget*. 2017;8:94371–94381.

29. Bartosik LP, Lajoie G, Sugar L, et al. Predicting progression in IgA nephropathy. *Am J Kidney Dis*. 2001;38:728–735.

30. Xie J, Lv J, Wang W, et al. Kidney failure risk prediction equations in IgA nephropathy: a multicenter risk assessment study in Chinese patients. *Am J Kidney Dis*. 2018;72:371–380.

31. Goto M, Kawamura T, Wakai K, et al. Risk stratification for progression of IgA nephropathy using a decision tree induction algorithm. *Nephrol Dial Transplant*. 2008;24:1242–1247.

32. Wakai K, Kawamura T, Endoh M, et al. A scoring system to predict renal outcome in IgA nephropathy: from a nationwide prospective study. *Nephrol Dial Transplant*. 2006;21:2800–2808.

33. Chen T, Li X, Li Y, et al. Prediction and risk stratification of kidney outcomes in IgA nephropathy. *Am J Kidney Dis*. 2019;74:300–309.

34. Barbour SJ, Canney M, Coppo R, et al. Improving treatment decisions using personalized risk assessment from the International IgA Nephropathy Prediction Tool. *Kidney Int*. 2020;98:1009–1019.