External Validation of the International IgA Nephropathy Prediction Tool

Junjun Zhang,1 Bo Huang,1,2 Zhangsuo Liu,3,4,5 Xutong Wang,1 Minhua Xie,1 Ruxue Guo,1 Yongli Wang,1 Dan Yu,1 Panpei Wang,1 Yuze Zhu,1 and Jingjing Ren1

Abstract

Background and objectives The International IgA Nephropathy Network recently developed and externally validated two models to predict the risk of progression of IgA nephropathy: full models without and with race. This study sought to externally validate the International IgA Nephropathy Prediction Tool in a large, independent, and contemporary cohort in China.

Design, setting, participants, & measurements We included 1373 patients with biopsy-confirmed primary IgA nephropathy from The First Affiliated Hospital of Zhengzhou University from January 2012 to May 2018 and calculated predicted risks for each patient. The outcomes of interest were a 50% decline in eGFR or kidney failure. We assessed the performance of both models using discrimination (concordance statistics and Kaplan–Meier curves between subgroups), calibration (calibration plots), reclassification (net reclassification improvement and integrated discrimination improvement), and clinical utility (decision curve analysis).

Results The median follow-up was 29 months (interquartile range, 21–43 months; range, 1–95 months), and 186 (14%) patients reached the kidney outcomes of interest. Both models showed excellent discrimination (concordance statistics >0.85 and well separated survival curves). Overall, the full model without race generally underestimated the risk of primary outcome, whereas the full model with race was well calibrated for predicting 5-year risk. Compared with the full model without race, the full model with race had significant improvement in reclassification, as assessed by the net reclassification improvement (0.49; 95% confidence interval, 0.41 to 0.59) and integrated discrimination improvement (0.06; 95% confidence interval, 0.04 to 0.08). Decision curve analysis showed that both full models had a higher net benefit than default strategies, and the model with race performed better.

Conclusions In this study, both full models demonstrated remarkable discrimination, acceptable calibration, and satisfactory clinical utility. The relatively short follow-up time may have limited the validation of these models. CJASN 15: 1112–1120, 2020. doi: https://doi.org/10.2215/CJN.16021219

Introduction

IgA nephropathy is the most common form of GN globally (1). A recent study involving 29 centers reported that IgA nephropathy accounts for 40% of all glomerular diseases diagnosed by kidney biopsy in Asia, 22% in Europe, 12% in the United States/Canada, and 6% in Latin America (2). The clinical course of IgA nephropathy varies from persistent asymptomatic urinary sediment abnormalities to rapid progression to kidney failure. About 30%–40% of patients with IgA nephropathy will develop kidney failure within 20 years (3).

Reliable risk assessment and stratification can help clinicians make appropriate clinical decisions for patients with IgA nephropathy by identifying high-risk patients early for interventions that delay disease progression and by reducing unnecessary exposure to the toxicity of immunosuppressive drugs in low-risk patients. Using a pooled cohort involving 3927 patients from seven cohorts across Europe, North America/South America, China, and Japan, the International IgA Nephropathy Network developed and externally validated two Cox proportional hazards models (full models with and without race) to provide risk prediction of kidney outcomes in patients with IgA nephropathy (4). Both models demonstrated excellent discrimination and calibration in the original study, and they were made into a calculator available online (https://qxmd.com/calculate-by-qxmd) and a mobile app (QxMD). This was the first international, multi-ethnic prediction tool for patients with IgA nephropathy that used different races as one of the predictors and integrated both mesangial and endocapillary hypercellularity, segmental glomerulosclerosis, and tubular atrophy/interstitial fibrosis (MEST) score and routinely available clinical predictors, which can provide risk estimates for individual patients at the time of kidney biopsy (5).

Extensive external validation is required before any risk prediction model can be widely adopted for clinical practice. Although the original study comprised two
cohorts from eastern China, both of those cohorts were collected before the release of the Kidney Disease Improving Global Outcomes (KDIGO) guideline in 2012 (6), which had changed some treatment principles for patients with IgA nephropathy. The purpose of this study was to externally validate the International IgA Nephropathy Prediction Tool using a large, independent, and contemporary cohort from central China.

Materials and Methods

Source of Data and Study Population

We performed a retrospective cohort study to externally validate the International IgA Nephropathy Prediction Tool. A total of 1373 patients with biopsy-confirmed primary IgA nephropathy were recruited from The First Affiliated Hospital of Zhengzhou University from January 2012 to May 2018 (Figure 1). Patient records were collected from the Hospital Information System of The First Affiliated Hospital of Zhengzhou University. We included patients who had biopsy-proven primary IgA nephropathy and had follow-up data for >1 year or developed kidney failure or a permanent reduction in eGFR of >50% within 1 year. We excluded patients who were younger than 18 years old, had kidney failure at the time of kidney biopsy, had no available MEST scores, or had missing baseline eGFR.

This study was approved by the medical ethics committee of The First Affiliated Hospital of Zhengzhou University. The data are anonymous, and therefore, the requirement for informed consent was waived. Results have been presented following the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement (7,8).

Predictors and Outcome

All of the predictors for this external validation study were defined and collected using the same methods as in the original cohort. Both full models included the following predictors: age, eGFR, mean arterial pressure, proteinuria, MEST score, and the use of renin-angiotensin-aldosterone system (RAAS) blockers and immunosuppression. Proteinuria, age, mean arterial pressure, and eGFR (according to the Chronic Kidney Disease Epidemiology Collaboration equation [9]) were gathered at the time of kidney biopsy. When data regarding 24-hour urinary protein excretion were missing, proteinuria was estimated using spot urine protein-creatinine ratio (10). Mean arterial pressure was defined as diastolic pressure plus one third of the pulse pressure. The use of RAAS blockers and immunosuppression was defined as any exposure at or prior to kidney biopsy. Kidney biopsy specimens were graded independently by two experienced pathologists blinded to clinical data according to the Oxford classification score of IgA nephropathy (11-13). The full model with race included the following additional race information: Chinese, Japanese, white, or other. The primary outcome was a combined event including either a permanent >50% reduction in eGFR or kidney failure (eGFR<15 ml/min per 1.73 m² or KRT), whichever occurred first.

Sample Size and Missing Data

There are no reliable sample size recommendations for studies that validate prognostic models, but at least 100 events are recommended (14). Only 4% of patients in our study had missing predictors (proteinuria; n=55). Therefore, to ensure the reliability of data, patients with missing predictors were excluded (15).

Calculation of Predicted Risk and Risk Groups

To calculate the linear predictor and prediction probability of the primary outcome for each patient, we used the formulas for both full models given by Barbour et al. (4). In the full model with race, the Chinese race was provided a slope of the linear predictor, by 

Statistical Analyses

We assessed the performance of models using discrimination, calibration, reclassification, and clinical utility.

Discrimination was firstly evaluated using concordance statistic (C statistic): Harrell c index (17) and time-dependent receiver operating characteristic (ROC) curve (18). We then calculated the calibration slope, the coefficient of the linear predictor, by fitting a Cox model to regress the primary outcome with the linear predictor as the only variable. A slope value greater than one indicates greater discrimination. Additionally, Kaplan–Meier curves

---

Figure 1. | Flow chart of cohort selection. MEST, mesangial and endocapillary hypercellularity, segmental glomerulosclerosis, and tubular atrophy/interstitial fibrosis.
were compared within the risk groups (16), with separation of the survival curves indicating that the models had good discrimination. The log-rank test was used to analyze the differences between survival curves. Hazard ratios between risk groups were calculated using the lowest-risk group as reference. The rate of eGFR decline for each risk group was the mean of individual slopes calculated using a linear mixed effects model.

Reclassification of clinical risk was estimated using continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI). For both continuous NRI and IDI, a value greater than zero indicates positive improvement, a value less than one indicates negative improvement, and a value equal to zero indicates no improvement; 95% confidence intervals (95% CIs) for continuous NRI and IDI were constructed using 100 bootstrap samples.

Calibration was evaluated using a graphic comparison of observed and predicted risks of the primary outcome over the follow-up period. Time-specific calibration was reported using calibration plots with predicted versus observed 5-year risks by tenths of predicted risk and risk groups according to linear predictors. Predicted risks were mean predicted risks of all individuals or groups, and observed risks were estimated using the Kaplan–Meier method.

Decision curve analysis was used to assess the clinical utility of prediction models (19). For a specified high-risk threshold, the net benefit of a model is the difference between the true positive rate and the false positive rate. Within a reasonable range of probability thresholds, the prediction model is considered of clinical use only if the net benefits are higher than treating either all patients or no patients. When comparing different models from the same dataset, it is possible to make an intuitive judgment using the decision curve analysis curves—at a certain threshold, the model with the higher net benefit is better (20). Software for decision curve analysis is publicly available on www.decisioncurveanalysis.org.

Statistical analyses were conducted using R Version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) and SAS Version 9.4 (SAS Institute, Cary, NC).

### Table 1. Characteristics of participants in the external validation cohort, the original derivation cohort, and the original validation cohort

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>External Validation Cohort</th>
<th>Original Derivation Cohort</th>
<th>Original Validation Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, no.</td>
<td>1373</td>
<td>2781</td>
<td>1146</td>
</tr>
<tr>
<td>Follow-up, median (IQR), yr</td>
<td>2.4 (1.8–3.6)</td>
<td>4.8 (3.0–7.6)</td>
<td>5.8 (3.4–8.5)</td>
</tr>
<tr>
<td>Age, median (IQR), yr</td>
<td>35 (27–45)</td>
<td>36 (28–45)</td>
<td>35 (27–45)</td>
</tr>
<tr>
<td>Men</td>
<td>766 (56%)</td>
<td>1608 (58%)</td>
<td>565 (49%)</td>
</tr>
<tr>
<td>Creatinine at biopsy, median (IQR), mg/dl</td>
<td>1.0 (0.8–1.4)</td>
<td>1.0 (0.8–1.4)</td>
<td>1.0 (0.8–1.3)</td>
</tr>
<tr>
<td>eGFR at biopsy, median (IQR), ml/min per 1.73 m²</td>
<td>86 (54–110)</td>
<td>83 (57–108)</td>
<td>90 (65–113)</td>
</tr>
<tr>
<td>Proteinuria at biopsy, median (IQR), g/d</td>
<td>1.7 (0.9–3.3)</td>
<td>1.2 (0.7–2.2)</td>
<td>1.3 (0.6–2.4)</td>
</tr>
<tr>
<td>Mean arterial pressure at biopsy, median (IQR), mm Hg</td>
<td>102 (94–111)</td>
<td>97 (89–106)</td>
<td>93 (85–103)</td>
</tr>
<tr>
<td>MEST histologic score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>251 (18%)</td>
<td>1054 (38%)</td>
<td>481 (42%)</td>
</tr>
<tr>
<td>E1</td>
<td>341 (25%)</td>
<td>478 (17%)</td>
<td>476 (42%)</td>
</tr>
<tr>
<td>S1</td>
<td>864 (63%)</td>
<td>2137 (77%)</td>
<td>912 (80%)</td>
</tr>
<tr>
<td>T1</td>
<td>253 (18%)</td>
<td>686 (25%)</td>
<td>207 (18%)</td>
</tr>
<tr>
<td>T2</td>
<td>237 (17%)</td>
<td>128 (5%)</td>
<td>122 (11%)</td>
</tr>
<tr>
<td>Crescents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>505 (37%)</td>
<td>953 (34%)</td>
<td>642 (56%)</td>
</tr>
<tr>
<td>C2</td>
<td>464 (34%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>RAAS blocker use at or prior to biopsy</td>
<td>716 (52%)</td>
<td>862 (32%)</td>
<td>320 (30%)</td>
</tr>
<tr>
<td>Immunosuppression use at or prior to biopsy</td>
<td>203 (15%)</td>
<td>252 (9%)</td>
<td>81 (7%)</td>
</tr>
</tbody>
</table>

IQR, interquartile range; MEST, mesangial and endocapillary hypercellularity, segmental glomerulosclerosis, and tubular atrophy/interstitial fibrosis; M, mesangial hypercellularity; E, endocapillary hypercellularity; S, segmental glomerulosclerosis; T, tubular atrophy/interstitial fibrosis; N/A, not applicable; RAAS, renin-angiotensin-aldosterone system.
Results

Study Population Characteristics

The comparison of participant characteristics between this external validation cohort and the original cohorts is presented in Table 1. The total follow-up of our cohort was 45,442 person-months. Among the 1373 participants in our cohort, a total of 186 (14%) experienced the combined primary outcome (50% reduction in eGFR or kidney failure), including 33 (2%) patients who experienced the primary outcome within 1 year. The median follow-up time was 29 months (interquartile range, 21–43; range, 1–95), and 133 patients (10%) were followed for 5 years or more. The overall survival rate from the primary outcome was 72% (95% CI, 67% to 77%) at 5 years.

Compared with the two original cohorts, our cohort was more contemporary (the median years of biopsies were 2016 versus 2006 and 1998, respectively), the duration of follow-up was shorter (2.4 years versus 4.8 and 5.8 years, respectively), and the proportion of patients with an eGFR<30 ml/min per 1.73 m² at biopsy was higher (9% versus 5% and 3%, respectively). A lower proportion of M1 lesions (18% versus 38% and 42%, respectively) and a higher proportion of T2 lesions (17% versus 5% and 11%, respectively) were observed in our cohort. Furthermore, more patients in our cohort were treated with RAAS blockers (52% versus 32% and 30%, respectively) and immunosuppression (15% versus 9% and 7%, respectively) at or prior to biopsy.

Performance of the International IgA Nephropathy Prediction Tool

Discrimination and Reclassification. In this external validation cohort, both full models showed excellent discrimination (Table 2). The Harrell c index was 0.88 (95% CI, 0.86 to 0.91) for both full models without and with race, and the area under the ROC curve values at 5 years were 0.862 and 0.861, respectively (Supplemental Figure 2). The calibration slopes were 1.30 (95% CI, 1.15 to 1.45) and 1.26 (95% CI, 1.12 to 1.40), respectively, proving that the discrimination was better than in the original derivation cohort. Compared with the full model without race, the full model with race demonstrated significant improvement in risk reclassification for predicting 5-year risk, with NRI and IDI of 0.49 (95% CI, 0.41 to 0.59) and 0.06 (95% CI, 0.04 to 0.08), respectively. Kaplan–Meier curves and hazard ratios between risk subgroups are presented in Figure 2 and Table 3. The well separated survival curves further confirmed the excellent distinguishing ability of both full models. Both full models were less able to distinguish between the two lower-risk (low-risk and intermediate-risk) groups, but they were able to differentiate between the two lower-risk groups, the higher-risk group, and the highest-risk group. Although the other three groups had similar rates of eGFR decline, the highest-risk group had the fastest rate in both models.

Calibration. Figure 3A shows the mean predicted risk of the primary outcome against the observed risk over the follow-up period. Overall, the full model without race modestly underestimated the risk, whereas the full model with race more obviously underestimated the risk within 3 years, modestly overestimated the risk in 3–4 years, and accurately estimated the risk in 4–7 years. The observed versus predicted 5-year risks according to tenths of predicted risk and risk groups are presented in Figure 3, B–E. For the 5-year predicted risk, there was an underestimation in the full model without race in all subgroups. The full model with race was calibrated well, with a mild overestimation in the low-, intermediate-, and higher-risk groups and mild underestimation in the highest-risk group.

Clinical Utility. The decision curve analysis is presented in Figure 4, and both models showed a positive net benefit for 5-year predicted risk over the entire range of threshold probabilities. The full model with race had better performance, with the net benefit higher than that of the full model without race despite a few overlaps and higher than the strategies of treating all patients or no patients.

Discussion

In this study, we assessed the performance of the International IgA Nephropathy Prediction Tool by external validation in a large, independent, and contemporary cohort of 1373 patients from The First Affiliated Hospital of Zhengzhou University. We found that both full models showed remarkable discrimination and performed well concerning clinical utility, with the full model with race calibrated well for predicting 5-year risk of the primary outcome. On the basis of discrimination, calibration, reclassification, and clinical utility, the full model with

<table>
<thead>
<tr>
<th>Variable</th>
<th>Full Model without Race</th>
<th>Full Model with Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harrell c index</td>
<td>0.88 (0.86 to 0.91)</td>
<td>0.88 (0.86 to 0.91)</td>
</tr>
<tr>
<td>AUC at 5 yr</td>
<td>0.862</td>
<td>0.861</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>1.30 (1.15 to 1.45)</td>
<td>1.26 (1.12 to 1.40)</td>
</tr>
<tr>
<td>5-yr performance compared with the full model without race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRI</td>
<td>0.49 (0.41 to 0.59)</td>
<td>-0.44 (-0.49 to -0.39)</td>
</tr>
<tr>
<td>NRI (events)</td>
<td>0.92 (0.87 to 1.00)</td>
<td>0.06 (0.04 to 0.08)</td>
</tr>
<tr>
<td>NRI (nonevents)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values in parentheses represent 95% confidence intervals. For NRI and IDI, statistically significant improvement is indicated by a 95% confidence interval that does not include zero. AUC, area under the receiver operating characteristic curve; NRI, net reclassification improvement; IDI, integrated discrimination improvement.
race had better overall performance for predicting 5-year risk in our cohort.

Accurate risk evaluation is essential for patients with IgA nephropathy because of the heterogeneity of kidney disease progression (21,22). Several prediction models have been proposed to evaluate the progression risk of patients with IgA nephropathy (23–35), but many of these models have several limitations. These limitations include (1) small sample sizes (23,24,27); (2) included predictors not being available at baseline (23); (3) no inclusion of pathologic indicators (23,27,30); (4) pathologic indicators needed to be reassessed or had not been extensively validated (24–26,29,31); (5) only inclusion of single-ethnic participants (23–30,32); and (6) lack of external validation (25,27,29,30). By using a large multiethnic collaboration cohort, the International IgA Nephropathy Prediction Tool, by contrast, was developed and externally validated integrating the most validated clinical and histology risk factors. Additionally, all of these predictors are readily available at the time of biopsy, enabling personal risk prediction at diagnosis, which is of great value for clinical decisions, clinical trial recruitment, and biomarker validation. Although the models can provide prediction probability of kidney outcomes at any time point, the authors suggested using the models for predicting risks at 5 years because this is the median follow-up time of the original derivation cohort.

In principle, extensive external validation of prediction models is necessary before widespread adaptation in clinical practice (36,37). There are three types of external validation studies: geographic, temporal, and domain validations (37). The Chinese population in the original study was from two cohorts in eastern China (Beijing and Nanjing), and the median years of biopsy in these original derivation and validation cohorts were 2006 and 1998, respectively. Therefore, the majority of participants in these original cohorts were recruited 10–20 years ago. In contrast, the participants of our cohort were almost exclusively from central China, and the time of biopsy for the participants recruited was between 2012 and 2018, when the KDIGO guidelines (6) had been issued and widely used to guide the treatment of patients with IgA nephropathy. Additionally, the proportion of the Chinese population in the original validation cohort was only 22% (Beijing cohort n = 247), and therefore, the performance of the model between different cohorts may be “neutralized.” Therefore, it was necessary to externally validate the International IgA Nephropathy Prediction Tool using an independent and more contemporary IgA nephropathy population from central China.

Discrimination partially depends on the variability of patient characteristics in the validation dataset. Both full models showed outstanding discrimination ability in our cohort, with C statistics >0.85 and well separated survival curves, and they performed even better than the original study (C statistics = 0.81 and 0.82). In addition to proving that the tool has good discrimination ability, it can be speculated that the patients within our cohort have greater heterogeneity, with a slower rate of eGFR decline in the
low-risk group (−1.07 and −1.15 ml/min per 1.73 m² per year versus −1.24 and −1.64 ml/min per 1.73 m² per year in the original study) and faster rates of eGFR decline in the highest-risk group (−4.34 and −4.19 ml/min per 1.73 m² per year versus −3.43 and −3.54 ml/min per 1.73 m² per year in the original study). We also included more patients with higher risk, with 44%–64% of participants having 5-year predicted risks higher than 10% (compared with 36%–44% in the original study), which may have contributed to the greater heterogeneity of our population.

The fast rate of decline in eGFR in the highest-risk group indicated that the models could identify the patients with rapid clinical progression. According to the KDIGO guideline (6), after appropriate supportive care for 3–6 months, patients with IgA nephropathy, persistent proteinuria >1 g/d, and GFR>50 ml/min per 1.73 m² are recommended to receive corticosteroid therapy. The Therapeutic Evaluation of Steroids in IgA Nephropathy Global study showed that glucocorticoids could delay the rate of GFR decline in patients with IgA nephropathy (38). It is worth considering whether there is a possibility that the use of glucocorticoids slowed down the rate of eGFR decline in patients in the intermediate- and higher-risk groups (characteristics of patients in each subgroup are presented in Supplemental Tables 2 and 3).

It is not sufficient to evaluate the prediction capability of a model using discrimination alone. The full models behaved differently in terms of calibration. Because Asians with IgA nephropathy are at higher risk for progression to kidney failure (39), we were not surprised that the full model without race underestimated the prediction risks generally. The full model with race demonstrated under- or overestimation within 4 years, and it accurately estimated the risk in 4–7 years, which was further confirmed by the well calibrated 5-year predicted risk. Reclassification (assessed by continuous NRI and IDI) and clinical utility (assessed by decision curve analysis) also indicated that the full model with race demonstrated better performance in predicting 5-year risk.

There are several limitations to this study that should be acknowledged. First, this was a retrospective study of a single Chinese center, and we did not have access to data on the use of RAAS blockers and immunosuppression after kidney biopsy. Therefore, we were unable to compare drug use between different risk subgroups as well as between our cohort and the original cohorts to infer whether the use of RAAS blockers or immunosuppression affected the rates of eGFR decline. Second, the patients in our cohort may have had relatively more severe conditions than all patients diagnosed with IgA nephropathy in our hospital because patients with milder conditions are less likely to follow up regularly in a large tertiary care center. This may also explain why our patients had such a high rate of T2 lesions. Third, there were some missing data in this study, and we performed analysis using only the complete dataset to ensure reliability. However, the proportion of the missing data was <5%, and therefore, the potential effect of ignoring the missing data could be considered negligible (15). Finally, the authors of the International IgA Nephropathy Prediction Tool originally suggested that the tool can be used to predict 5-year risk of kidney outcomes; however, the median follow-up of our cohort was only 29 months. Although the results of the time-dependent area under the ROC curve proved that both full models had excellent discrimination at different time points (Supplemental Figure 2), the short follow-up period may have affected the accuracy of this external validation study. We plan to extend follow-up and revalidate the full models in the future for more reliable results.

In conclusion, we externally validated the International IgA Nephropathy Prediction Tool using a large cohort from a different institution and a different period from the original study. Both full models demonstrated excellent discrimination and satisfactory clinical utility, and the full model with race was well calibrated in predicting 5-year risk and showed the best overall performance. This tool can be easily applied in clinical practice to provide risk estimates for individual patients at an early disease stage. Further multicenter prospective cohort studies are required.

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Events, n (%)</th>
<th>Hazard Ratio (95% Confidence Interval)</th>
<th>P Value</th>
<th>Rate of eGFR Decline (95% Confidence Interval), ml/min per 1.73 m²/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full model without race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>4 (2%)</td>
<td>1</td>
<td>—</td>
<td>−1.15 (−1.52 to −0.78)</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td>10 (2%)</td>
<td>1.20 (0.38 to 3.81)</td>
<td>0.76</td>
<td>−1.01 (−1.33 to −0.69)</td>
</tr>
<tr>
<td>Higher risk</td>
<td>48 (10%)</td>
<td>5.95 (2.15 to 16.53)</td>
<td>0.001</td>
<td>−1.21 (−1.68 to −0.73)</td>
</tr>
<tr>
<td>Highest risk</td>
<td>124 (57%)</td>
<td>45.13 (16.64 to 122.43)</td>
<td>&lt;0.001</td>
<td>−4.19 (−4.88 to −3.50)</td>
</tr>
<tr>
<td>P value for trend</td>
<td>—</td>
<td>—</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Full model with race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>4 (2%)</td>
<td>1</td>
<td>—</td>
<td>−1.07 (−1.44 to −0.71)</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td>10 (2%)</td>
<td>1.16 (0.37 to 3.71)</td>
<td>0.80</td>
<td>−1.12 (−1.45 to −0.79)</td>
</tr>
<tr>
<td>Higher risk</td>
<td>47 (10%)</td>
<td>5.65 (2.03 to 15.70)</td>
<td>0.001</td>
<td>−1.06 (−1.53 to −0.59)</td>
</tr>
<tr>
<td>Highest risk</td>
<td>125 (57%)</td>
<td>46.56 (17.16 to 126.34)</td>
<td>&lt;0.001</td>
<td>−4.34 (−5.01 to −3.67)</td>
</tr>
<tr>
<td>P value for trend</td>
<td>—</td>
<td>—</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3. Calibration plots demonstrated the full model without race generally underestimated the risk of primary outcome, whereas the full model with race was well calibrated for predicting 5-year risk. (A) Comparison of observed and predicted risks of the primary outcome during the follow-up period. Predicted risks are mean predicted risk curves (blue and yellow solid lines), observed risks were estimated by Kaplan–Meier method (red solid line), and the red dashed lines represent 95% confidence intervals of the observed risks. (B–E) Comparison of observed and predicted 5-year risks of the primary outcome in risk groups and tenths of predicted risk. Plots by risk groups are in (B) and (D), and plots by tenths of predicted risk are in (C) and (E). The dashed lines indicate perfect calibration, in which predicted risks are exactly the same as the observed risks. Vertical lines in observed groups represent 95% confidence intervals. Risk groups were on the basis of percentiles of the linear predictor: group 1 (low risk): <16th; group 2 (intermediate risk): 16th to 50th; group 3 (higher risk): 50th to 84th; and group 4 (highest risk: >84th).
to assess the predictive performance of the International IgA Nephropathy Prediction Tool and whether the tool could change clinical decisions or improve the prognosis for patients with IgA nephropathy (37).

Disclosures
All authors have nothing to disclose.

Funding
This work was supported by National Natural Science Foundation of China grant 81570645, Innovation Scientists and Technicians Troop Construction Projects of Henan Province grant 2018JR0014, Program for Science & Technology Innovation Talents in Universities of Henan Province grant 18HASTIT043, Major Project of Henan Medical Science and Technology Research Program grant 201501018, Science and Technology Huimin Project of Henan Province grant 162207310001, and Pathogenesis & Precision diagnosis and treatment of Chronic Kidney Disease Projects of Henan Province grant 162207310002, and Pathogenesis Project of Henan Medical Science and Technology Research in Universities of Henan Province grant 18HASTIT043, Major Project of Henan Medical Science and Technology Research Program grant 201501018, Science and Technology Huimin Project of Henan Province grant 162207310001, and Pathogenesis & Precision diagnosis and treatment of Chronic Kidney Disease grant 18210510002.

Supplemental Material
This article contains the following supplemental material online at http://cjasn.asnjournals.org/lookup/suppl?doi=10.2215/CJN.16021219/-/DCSupplemental.

Supplemental Figure 1. Histogram of the distribution in this external cohort (n = 1373) of the 5-year predicted risk of the primary outcome (50% decline in eGFR or kidney failure).

Supplemental Figure 2. Time-dependent receiver operating characteristic curve.

Supplemental Figure 3. The 5-year predicted risks of the patients who developed the primary outcome within 1 year (n = 33).

Supplemental Figure 4. Comparison of observed and predicted 5-year risks of the primary outcome in risk groups and tenths of predicted risk in patients with over 1-year follow-up (n = 1340).

Supplemental Figure 5. Decision curve analysis using 5-year predicted risk in patients with over 1-year follow-up (n = 1340).

Supplemental Figure 6. Comparison of observed and predicted 3-year risks of the primary outcome in risk groups and tenths of predicted risk in the full external cohort (n = 1373).

Supplemental Figure 7. Decision curve analysis using 3-year predicted risk in the full external cohort (n = 1373).

Supplemental Table 1. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis checklist for prediction model development and validation.

Supplemental Table 2. The characteristics of participants in the risk subgroups on the basis of the linear predictor of the full model without race.

Supplemental Table 3. The characteristics of participants in the risk subgroups on the basis of the linear predictor of the full model with race.

Supplemental Table 4. Concordance statistics, calibration slopes, net reclassification improvements, and integrated discrimination improvements of the full models in patients with over 1-year follow-up (n = 1340).

Supplemental Table 5. Associations of the risk groups with the composite outcome of 50% decline in eGFR or kidney failure in patients with over 1-year follow-up (n = 1340).

Supplemental Table 6. Risk reclassification of the primary outcome using 3-year predicted risk in the full external cohort (n = 1373).

References
5. Barbour SJ: Personalised risk stratification in IgAN – is it possible? Kidney Dis 4: 145–146, 2018


Received: December 31, 2019 Accepted: May 28, 2020

J.Z. and B.H. contributed equally to this work and are first authors.

Published online ahead of print. Publication date available at www.cjasn.org.