Nephrol Dial Transplant (2022) 0: 1–9
https://doi.org/10.1093/ndt/gfac225
Advance Access publication date 29 July 2022

Validation of two IgA nephropathy risk-prediction tools using a cohort with a long follow-up

Yngvar Lunde Haaskjold[1,2], Njål Gjærde Lura[3], Rune Bjørneklett[2,4], Leif Bostad[2,5], Lars Sigurd Bostad[2,4] and Thomas Knoop[1,2]

[1]Department of Medicine, Haukeland University Hospital, Bergen, Norway, [2]Renal Research Group, Department of Clinical Medicine, University of Bergen, Norway, [3]Department of Radiology, Haukeland University Hospital, Bergen, Norway, [4]Emergency Care Clinic, Haukeland University Hospital, Bergen, Norway and [5]Department of Pathology, Haukeland University Hospital, Bergen, Norway

Correspondence to: Yngvar Lunde Haaskjold; E-mail: yngvarhaaskjold@gmail.com

ABSTRACT

Background. Recently, two immunoglobulin A (IgA) nephropathy-prediction tools were developed that combine clinical and histopathologic parameters. The International IgAN Prediction Tool predicts the risk for 50% declines in the estimated glomerular filtration rate or end-stage kidney disease up to 80 months after diagnosis. The IgA Nephropathy Clinical Decision Support System uses artificial neural networks to estimate the risk for end-stage kidney disease. We aimed to externally validate both prediction tools using a Norwegian cohort with a long-term follow-up.

Methods. We included 306 patients with biopsy-proven primary IgA nephropathy in this study. Histopathologic samples were retrieved from the Norwegian Kidney Biopsy Registry and reclassified according to the Oxford Classification. We used discrimination and calibration as principles for externally validating the prognostic models.

Results. The median patient follow-up was 17.1 years. A cumulative, dynamic, time-dependent receiver operating characteristic analysis showed area under the curve values ranging from 0.90 at 5 years to 0.83 at 20 years for the International IgAN Prediction Tool, while time-naive analysis showed an area under the curve value at 0.83 for the IgA Nephropathy Clinical Decision Support System. The International IgAN Prediction Tool was well calibrated, while the IgA Nephropathy Clinical Decision Support System tends to underestimate risk for patients at higher risk and overestimates risk in the lower risk categories.

Conclusions. We have externally validated two prediction tools for IgA nephropathy. The International IgAN Prediction Tool performed well, while the IgA Nephropathy Clinical Decision Support System has some limitations.

Keywords: external validation, IgA nephropathy, machine learning, prediction models, prognostic tool

INTRODUCTION

Immunoglobulin A nephropathy (IgAN) is a common cause of end-stage kidney disease (ESKD) worldwide [1]. Most patients with IgAN are diagnosed as young adults based on clinical findings of microscopic hematuria and proteinuria, followed by a kidney biopsy, which is mandatory to confirm the diagnosis [2]. The clinical course is highly varied, with some patients rapidly progressing to ESKD, while the kidney function in other patients remains preserved [2, 3]. Clinical features, such as a reduced estimated glomerular filtration rate (eGFR), proteinuria, and hypertension, are known risk factors for progression to ESKD [3, 4]. The Oxford Classification, which is an IgAN histopathologic model that was established in 2009, combines four histological lesions associated with adverse outcomes: mesangial hypercellularity (M), endocapillary hypercellularity (E), segmental glomerulosclerosis (S), and tubular atrophy/interstitial fibrosis (T) [5, 6]. Crescents (C) was added to the model in 2016 [7].

Currently, there are no effective treatments for IgAN; however, several ongoing clinical trials are bringing hope for new therapeutic agents [8–10]. This development underscores the importance of accurate prediction tools that can aid clinicians in selecting patients for inclusion in studies and identifying patients that could benefit from future treatments.

Two IgAN prediction tools have recently been developed, both of which combine clinical and histopathologic parameters. Barbour *et al.* developed the International IgAN Prediction Tool (IIGAN-PT), which is a Cox proportional hazards model designed to predict the risk for 50% declines in the eGFR or ESKD up to 80 months after diagnosis [11]. This tool is derived from a multiethnic cohort, and it is available at https://qxmd.com/calculate-by-qxmd. The newly updated Kidney Disease: Improving Global Outcomes (KDIGO) guidelines encourage clinicians to use this tool with their patients [12]. External validation studies show that the

KEY LEARNING POINTS

What is already known about this subject?

- Two prognostic models for immunoglobulin A (IgA) nephropathy, the International IgAN Prediction Tool and the IgA Nephropathy Clinical Decision Support System have recently been developed.
- The International IgAN Prediction Tool has been externally validated in cohorts limited by short follow-up time. The IgA Nephropathy Clinical Decision Support System has yet to be externally validated.

What this study adds?

- External validation of both models using a cohort with long follow-up time.

What impact this may have on practice or policy?

- External validation of prediction models in different cohorts is necessary before the models are put into widespread clinical use.
- The International IgAN Prediction Tool might be used to predict kidney survival beyond 80 months.
- The IgA Nephropathy Clinical Decision Support System has some limitations, especially when predicting time to end-stage kidney disease.

tool performs well, but these studies are limited by short follow-up times [13–15].

Schena *et al.* developed the IgA Nephropathy Clinical Decision Support System artificial neural network model (ANN model), which uses ANNs to estimate the risk for ESKD up to 10 years after diagnosis [16]. The tool combines two models: a classifier model to predict the risk of ESKD and a regression model to predict the time to ESKD. The ANN model was derived from a European cohort, and it is available at https://igan.poliba.it. It has yet to be externally validated.

External validation of prognostic models is of great importance because it ensures that the models perform well in different cohorts before they are put to use in clinical practice [17, 18]. Therefore, the aim of this study was to validate the IIGAN-PT and ANN model using a Norwegian cohort of patients with IgAN that was retrieved from the Norwegian Kidney Biopsy Register. The follow-up period for this cohort was up to 28 years, which is longer than the respective derivation cohorts for each prediction tool. A longer time frame such as this is important for evaluating a slow-progressing disease such as IgAN.

MATERIALS AND METHODS

Study population

In this study, we included 306 patients from a cohort previously used to address the prognostic value of the Oxford Classification [19]. The patients all had IgAN, proven by biopsy, before 2010, an initial eGFR level above 30 mL/min/1.73 m^2, and histopathologic specimens that were available for reanalysis according to the Oxford Classification. The patient data were retrieved from the Norwegian Kidney Biopsy Registry (Bergen, Norway), which, since 1988, has compiled morphologic, clinical, biochemical, and immunologic data from patients in Norway subject to a kidney biopsy. Treatment data at the time of biopsy were obtained retrospectively from patient records. The clinical data collected during follow-up were obtained from patient records and the Norwegian Renal Registry, which is located at Rikshospitalet in Oslo, Norway.

Table 1. Parameters included in the prognostic models

IIGAN-PT (Barbour *et al.* [10])	ANN model (Schena *et al.* [15])
eGFR	Creatinine
Systolic blood pressure	Systolic blood pressure
Diastolic blood pressure	Diastolic blood pressure
Proteinuria	Proteinuria
Age	Age
Race	ACE inhibitor/ARB
ACE inhibitor/ARB	Sex
M	M
E	E
S	S
T	T
Immunosuppression	C
	Immunosuppression

ARB, angiotensin receptor blocker; ACE, angiotensin-converting enzyme.

The observation period was defined as the time between the diagnostic kidney biopsy and ESKD, death, or the end of April 2020.

This study was approved by the Western Norwegian Regional Committee for Medical and Health Research Ethics (Reference no. 2018/2130), and study participation was based on informed consent. The research was done according to the Declaration of Helsinki.

Histopathologic studies

Biopsy slides were retrieved from the Norwegian Kidney Biopsy Register and reclassified according to the Oxford Classification (MEST-C score) [5–7] by an experienced renal pathologist who was blinded to the clinical data.

Variables and outcome predictors

We collected the same clinical and histopathologic data that were used to develop the two prediction tools evaluated in this study [11, 16]. The clinical and histopathologic parameters for both tools are summarized in Table 1. The primary outcome when evaluating the IIGAN-PT was a composite outcome of 50% decline in eGFR or ESKD, while ESKD was the

primary outcome in the evaluation of the ANN model. In both derivation studies, ESKD was defined as eGFR less than 15 mL/min/1.73 m^2 {chronic kidney disease [CKD] stage 5}, start of dialysis, or receipt of a kidney transplant. Because reporting for CKD stage 5 only recently became routine in the Norwegian kidney registry, however, we defined ESKD as the start of dialysis or receipt of a kidney transplant. The prognostic index for IIGAN-PT was calculated using the reported formula from the derivation study. Because no formula are provided for the ANN model, we derived the predicted outcome probabilities by plotting the data from the validation cohort into the online tool and further used them as a prognostic index in the validation analysis [18].

Statistics

We evaluated discrimination and calibration as principles for the external validation of the prognostic models, as described by Royston and Altman [20]. The ANN model was evaluated in two steps: first evaluating the predicted binomial outcome of ESKD, and then evaluating the precision of the time-to-event estimate for the cases with an outcome.

Discrimination

We used a time-dependent receiver operating characteristic (tdROC) analysis to evaluate the prognostic performance of IIGAN-PT; we used a time-naive ROC analysis for the ANN-model. We used the 'timeROC' R package (R Foundation for Statistical Computing, Vienna, Austria) [21] to evaluate the cumulative prognostic performance at key time points of interest. The discrimination ability of each model was also assessed using the concordance index. We calculated the calibration slopes for the models by performing regression on the prognostic indexes in the current validation data set [22]. The prognostic indexes were categorized into risk groups by the cut points Royston and Altman recommended (16th, 50^{th}, and 84th centiles) [20]. Kaplan-Meier survival curves for risk groups were drawn for both models.

Calibration

We evaluated model calibration by drawing calibration plots for censored survival data for the IIGAN-PT model, comparing the predicted with the observed survival probabilities in calibration plots at 5, 10, 15, and 20 years. Survival plots comparing mean survival predictions from the Cox model and observed Kaplan-Meier survival estimates in risk groups were drawn for the IIGAN-PT model. We assessed model calibration for the ANN model by drawing a binomial outcome calibration plot, then comparing observed and predicted outcomes [20].

Other statistical considerations

We evaluated the proportional hazard assumptions by using the Schoenfeld test of residuals and by drawing a log-hazard plot for the IIGAN-PT. A Bland-Altman plot was used to evaluate the ANN-model's ability to predict time to ESKD, comparing predicted time to event with observed time to event.

All the data were analyzed using the R software package, version 4.0.3.

RESULTS

Clinical and histopathologic characteristics

In total, 306 patients from an all-Norwegian cohort were included in this study. A total of 234 of the patients were male (76.5%), and the mean and median patient follow-up periods were 16.5 years and 17.1 years, respectively. The mean age at the time of the biopsy was 37.4 years. The mean eGFR at the time of the biopsy was 78.4 mL/min/1.73 m^2, and the mean proteinuria level was 1.7 g/day. Renin-aldosterone-angiotensin system (RAAS) inhibitors were used frequently (70.9%), while only 6.5% of the patients received immunosuppressants. During the study period, 61 patients (20%) reached ESKD, and 17 patients died before reaching ESKD or end of follow-up. The baseline characteristics from our cohort as well as those of the derivation cohorts (as reported by Barbour [11] and Schena [16]), are shown in Table 2.

Discrimination

The calibration slope for IGAN-PT was 0.79 {95% confidence interval [CI], 0.68–0.97; $P = .03$; for H_0, slope = 1}. Conversely, the calibration slope for the ANN-model was 1.64 {95% CI, 1.22–2.07; $P < .001$; for H_0, slope = 1}, indicating potential local issues regarding calibration and discrimination for both models. The tdROC analysis showed that IIGAN-PT had excellent discrimination abilities at 5 years, with an area under the curve (AUC) value of 0.90 in predicting renal survival. The discriminatory ability of the tool declines over time, with an AUC value of 0.83 at 20 years. Time-naive ROC analysis of the ANN model showed good discriminatory abilities, with an AUC value of 0.83 for predicting ESKD. The concordance index was 0.80 for IIGAN-PT and 0.83 for the ANN model for predicting renal survival and ESKD, respectively. All values are summarized in Table 3. A direct comparison of the discriminatory abilities of IIGAN-PT and the ANN-model models is not appropriate because the two models are evaluating different end points and the concordance index and AUC values have been derived from different analyses.

Calibration

The calibration plots for censored survival data evaluating the IIGAN-PT model show overall acceptable alignment to observed renal survival probabilities, but it does tend to underestimate risk at 5 years and overestimate risk at 20 years for the patients with the highest observed risk (Fig. 1a). The survival plots comparing mean survival predictions from the Cox model and observed Kaplan-Meier survival estimates in risk groups show that accumulated over time, the Cox model slightly overestimates the risk for the patients in the most

Table 2. Baseline characteristics

Characteristics	Validation cohort	IIGAN-PT[a]	ANN model[a]
Patients, *n*	306	2781	948
Follow-up (years), median (IQR)	17.1 (12.9–21.3)	4.8 (3.0–7.6)	7.42 (4.2–11.2)
Follow-up (years), mean (SD)	16.6 (7)	NA	NA
Age at biopsy (years), median (IQR)	35.0 (25.0–46.0)	35.6 (28.2–45.4)	NA
Age at biopsy (years), mean (SD)	37.4 (14)	NA	40.6 (14)
Male sex, *n* (%)	234 (76.5)	1608 (57.8)	685 (72.3)
Race			
Caucasian, *n* (%)	306 (100)	1167 (42)	948 (100)[b]
Japanese, *n* (%)	0 (0)	569 (20.5)	0 (0)
Chinese, *n* (%)	0 (0)	1021 (36.7)	0 (0)
Other, *n* (%)	0 (0)	22 (0.8)	0 (0)
MEST score			
M = 1, *n* (%)	103 (33.7)	1054 (38.0)	307 (32.4)
E = 1, *n* (%)	81 (26.5)	478 (17.3)	108 (11.4)
S = 1, *n* (%)	168 (54.9)	2137 (77.0)	710 (74.9)
T1 = 1, *n* (%)	32 (10.5)	686 (24.7)	194 (20.5)
T2 = 2, *n* (%)	2 (0.65)	128 (4.6)	44 (4.6)
Crescents	70 (22.9)	953 (34.3)	NA
C = 1, *n* (%)	61 (19.9)	NA	86 (9.1)
C = 2, *n* (%)	9 (2.9)	NA	NA
Creatinine (μmol/L), median (IQR)	90.0 (74–114)	92.0 (70.7–123.8)	NA
Creatinine (mg/dL), median (IQR)	1.02 (0.84–1.29)	NA	1.20 (0.96–1.70)
eGFR (mL/min/1.73 m^2), median (IQR)	79.8 (57.2–101.0)	83.0 (57.6–108.0)	67.3 (44.9–89.9)
Proteinuria (g/day), median (IQR)	0.80 (0.3–2.2)	1.2 (0.7–2-2)	1.3 (0.60–2.5)
Systolic blood pressure (mm Hg), mean (SD)	135 (17)	NA	131.4 (18.6)
Diastolic blood pressure (mm Hg), mean (SD)	83 (11)	NA	83.2 (10.9)
MAP (mm Hg), median (IQR)	100.0 (93.0–107.0)	96.7 (88.7–106.3)	NA
MAP (mm Hg), mean (SD)	100.1 (12.4)	NA	100.2 (12.4)
RAAS, *n* (%)	217 (70.9)	862 (32.4)	577 (60.9)
Immunosuppressants, *n* (%)	20 (6.5)	252 (9.1)	258 (27.2)
Clinical outcome			
50% decline in eGFR, *n* (%)	90 (29.4)	420 (15.1)	NA
ESKD, *n* (%)	61 (19.9)	372 (13.4)	210 (22.2)

[a]Reported derivation cohorts. IQR, interquartile range; MAP, mean arterial blood pressure; NA, not applicable.
[b]European cohort.

Table 3. Discrimination: the concordance index and cumulative dynamic time-dependent ROC analysis

	IIGAN-PT, AUC (95% CI)	ANN model, AUC (95% CI)
Harrel's C index	0.80 (0.75–0.84)[a]	0.83 (0.77–0.89)[b]
AUC[c]	NA	0.83 (0.77–0.88)
5-year AUC	0.90 (0.85-0.0.94)	NA
10-year AUC	0.87 (0.82–0.92)	NA
15-year AUC	0.86 (0.80–0.91)	NA
20-year AUC	0.83 (0.75–0.89)	NA

[a]Time-dependent concordance index.
[b]Time-naive concordance index.
[c]Time-naive ROC analysis.

severe risk group and underestimates the risk for patients in risk group 1 (Fig. 2). Kaplan-Meier curves were drawn for both models, showing probability for primary end points in four different risk groups (Fig. 3).

The calibration plot for the ANN model, which compares predicted and observed risk regardless of time points, as intended by the model's design, shows no significant miscalibration in most groups. The ANN model, however, underestimates risk for patients in the highest risk group and overestimates risk in one of the groups with intermediate risk. There is also a tendency to overestimate risk in the lower risk categories, but there are fewer events in these categories to base the estimates on, and precise estimation of the calibration in these groups would require additional patients (Fig. 1b). The ANN model's ability to predict time to ESKD (Fig. 4), with a cutoff at 50% given by the online tool, achieves a specificity of 78%, but the sensitivity is only 41%.

Assumption of proportional hazard for IIGAN-PT using the Schoenfeld test of residuals yielded a significant *P*-value ($P < .001$), indicating a violation of the proportional hazard assumption. This finding was confirmed by a log-hazard plot showing a significant decline in the hazard ratio over time for the IIGAN-PT prognostic index.

In patients for whom the ANN model predicted ESKD, the model also supplied time-to-event estimates. In the 25 patients for whom the model correctly predicted ESKD, the model overestimated the time to event by 2.5 years on average, with 9.5 years and −14.5 years as the upper and lower limits of agreement, respectively, shown in a Bland-Altman plot (Fig. 4).

DISCUSSION

We have externally validated two new IgAN prognostic tools: IIGAN-PT, which was developed by Barbour *et al.*, and the ANN model, which was developed by Schena *et al.* IIGAN-PT was well calibrated and showed good discriminatory abilities

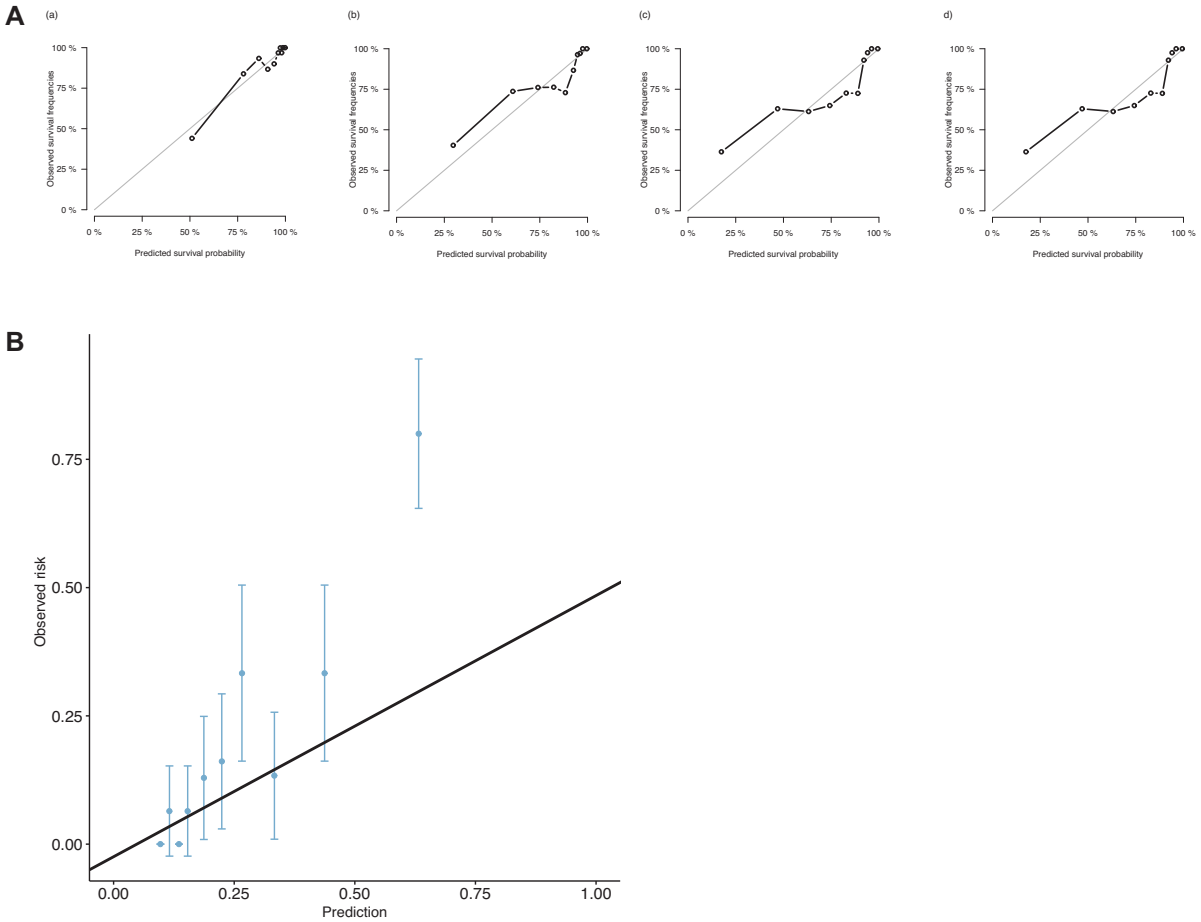

FIGURE 1: Calibration plots. **(A)** Plot comparing observed and predicted for renal survival in the IIGAN-PT at 5 (a), 10 (b), 15 (c), and 20 (d) years. **(B)** Plot comparing observed and predicted risk for the primary end point (ESKD) in the ANN model.

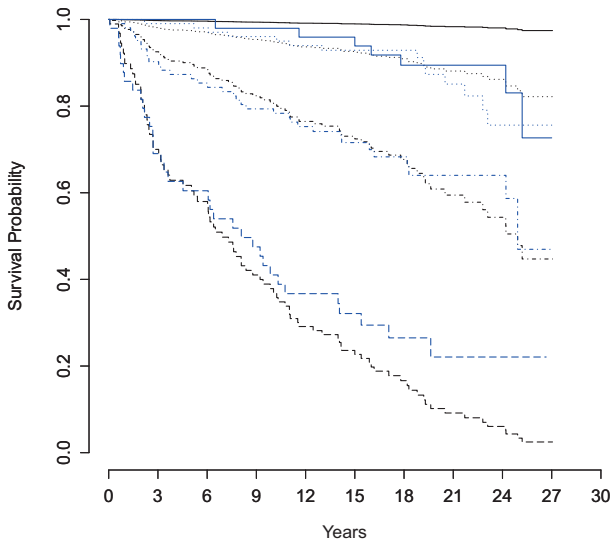

FIGURE 2: Comparison of the observed survival (blue lines) with mean predicted risk (black lines) from the IIGAB-PT.

for the first 10 years. The ANN model had good discrimination, but it has some local calibration issues, underestimating risk for patients with higher risk and overestimating risk for patients in the lower risk categories. Its ability to predict time to event in cases with predicted and actual ESKD was insufficient.

Before the current study, IIGAN-PT had been validated only at 80 months, and it is derived from a cohort with a short follow-up time; therefore, clinicians have been advised not to use the tool to predict prognoses at later stages [23].

IIGAN-PT is derived from a multiethnic cohort, while the ANN model included only European patients in their study cohort. The patients included in the present study were recruited from an all-Norwegian cohort; therefore, it will be important to validate both tools for different ethnic groups [23]. IIGAN-PT has previously been externally validated in a Chinese cohort and a combined Chinese–Argentinean cohort [13, 14], and the results from those studies are similar to the results reported herein.

There are some major differences in the development of the two models. The primary end point for IIGAN-PT is a composite end point of the first occurrence of a 50% decline in eGFR from baseline or ESKD, while the ANN model uses a hard end point, such as ESKD. Barbour *et al.* argue that a 50% decline in eGFR is a widely accepted surrogate end point [11], while Schena *et al.* suggest that one should use ESKD as an end point in a slow, progressive disease such as IgAN, where patients can have a permanent reduction in eGFR greater than

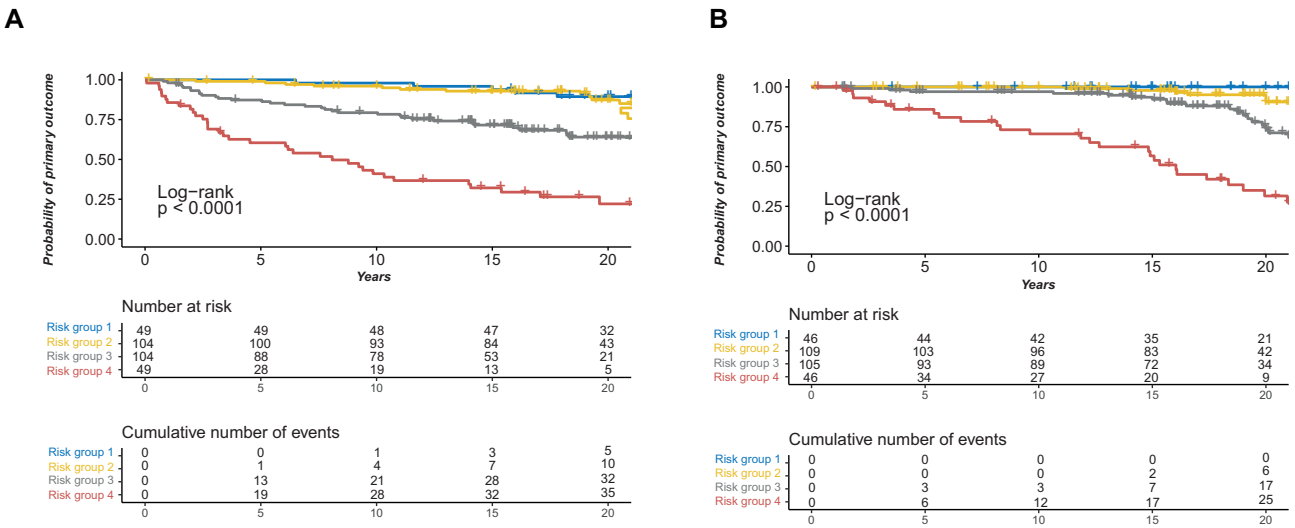

FIGURE 3: Kaplan-Meier curves for primary end points. **(A)** 50% decline in eGFR or ESKD for IIGAN-PT and **(B)** ESKD for the ANN model. Patients are divided into four risk groups: <16th, 16th–50th, 50th–84th, and >84th centiles.

50% for many years [16]. Notably, both prediction models included patients with eGFR less than 30 mL/min/1.73 m^2 in their derivation cohorts, but the ANN model also included 15 patients (1.5%) with an eGFR less than 15 mL/min/1.73 m^2 as well as active lesions, such as crescents, which could influence the performance of the ANN model in our validation cohort. The exclusion of patients with an eGFR under 15 mL/min/1.73 m^2 and C lesions by IIGAN-PT limits the tool's ability to identify patients with acute kidney failure and active lesions that could be responsive to treatment with glucocorticoids.

Both the IIGAN-PT and ANN model include well-known risk factors for disease progression to ESKD, such as eGFR, blood pressure, proteinuria level, and histopathologic parameters. They also include the use of immunosuppressants and RAAS inhibitors [3]. Barbour *et al.* chose to add race to their model and leave out crescents, arguing that crescents are strongly associated with race, and race is a more prominent predictor of ESRD [23]. We have previously described that even though crescents are independently a strong predictor for progressive decrease in renal function [24], adding C to the MEST score does not significantly improve its prognostic value [19]. Notably, the ANN model includes sex as a prognostic factor, which is supported by a recent study describing male sex as a risk factor for IgAN [25].

The IIGAN-PT and ANN models include 12 and 13 different variables, respectively. Even though the tools are available online, clinicians still must provide both clinical and histopathologic data to use them. As described previously, we found that the MEST score alone performed well as a prognostic model [19]. This finding could indicate that some clinical parameters, such as blood pressure, proteinuria, and eGFR, fluctuate, which makes them, to some extent, less accurate than histopathologic parameters, given a representative biopsy specimen. Histopathologic finding are isolated observations, however, and repeated biopsies have been shown to improve the prediction of ESKD in IgAN [26]. Further, recent studies have found proteinuria and eGFR slope to be the most important clinical variables in the evaluation of IgAN in clinical trials [27, 28].

In our cohort, only 6.5% of the patients received immunosuppressive treatment, which is lower than expected given the high rate of E lesions among the patients [7]. Additionally, S and T lesions were less frequent in our validation cohort, which could be the result of including only patients who had an eGFR over 30 mL/min/1.73 m^2 in the study, as this inclusion criterion may have excluded patients with a poor prognosis. Conversely, a larger proportion of the patients in our cohort received RAAS inhibitors, indicating that these patients were treated according to the KDIGO guidelines from 2012 [29]; therefore, they could be seen as representative, even though many of the biopsies were performed at an earlier stage [13, 14].

There are some methodological concerns when validating clinical decision tools. It is recommended that the prognostic index be derived from the regression formula from the derivation study [20], but it has been suggested that one should consider using the estimated risk score derived from the online tool if this score is the value intended for clinical use [18].

In the present study, we used the original formula when evaluating IIGAN-PT, while the ANN-model was evaluated by using risk scores derived from the online tool because no regression formula was available. This should be kept in mind when comparing the results from the study.

The use of machine learning (ML) could represent a paradigm shift for IgAN prognostic modeling because it combines genetic, proteomic, imaging, metabolic, and microbiome data with clinical and histopathologic information [30]. Two models have recently been developed based on different combinations of clinical and histopathologic data: A random forest model was developed by Liu *et al.* in 2018, and a model based on eXtreme Gradient Boosting (XGBoost) was developed by Chen *et al.* in 2019 [31, 32]. Proponents of ML models postulate that they may yield superior predictive performance over conventional regression models by capturing complex, nonlinear variable relationships

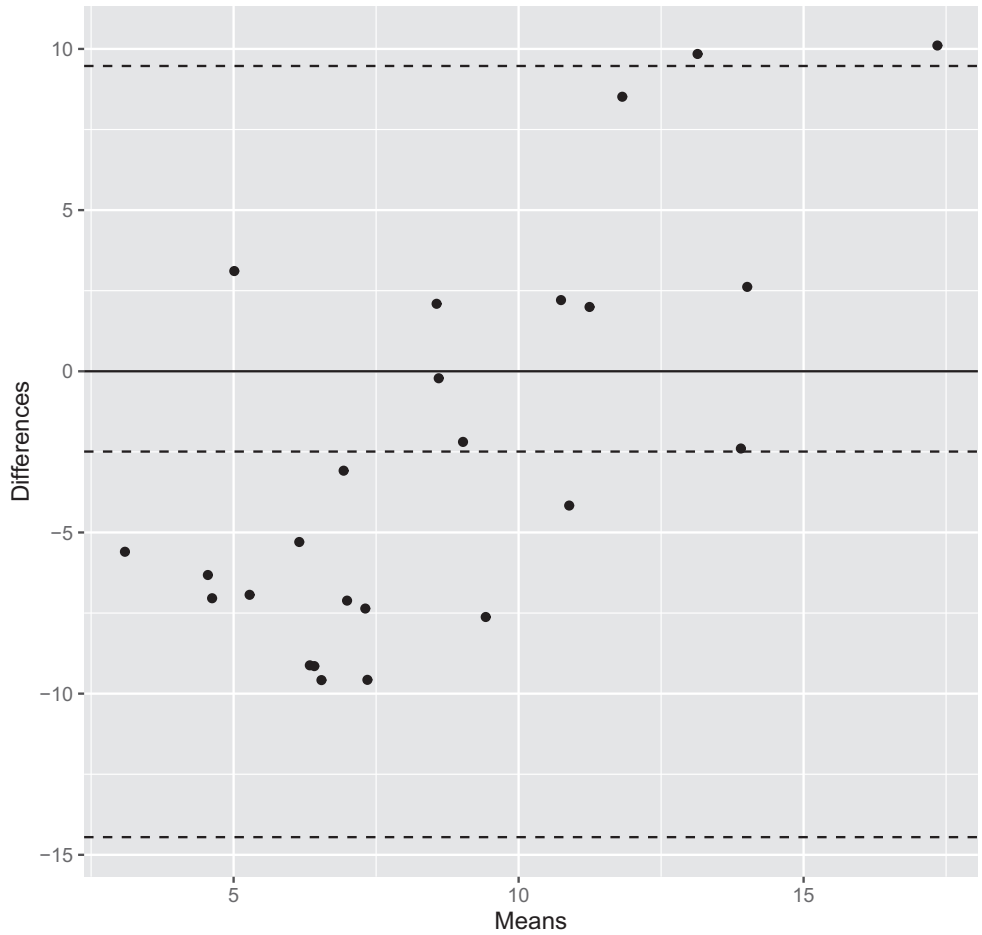

FIGURE 4: Bland-Altman plot describing time to event, as suggested by the ANN model, compared with observed cases.

in some data sets [33], and some models have provided more accurate prediction estimates than conventional statistical regression models [34, 35]. ML models are prone to overfitting, however; thus, validating the external performance of these models is crucial for determining whether promising results from a local data set are applicable to external data and, therefore, could have potential as a clinical prediction tool. Unfortunately, external validations of prediction models are scarce for both conventional regression and ML models [18, 36]. The application of the ANN model for predicting outcomes using only a few explanatory variables is also questionable because it is not clear that the ML model would achieve better performance than a conventional regression model and important information regarding the explanatory variables such as the variable coefficients is not available. It seems more appropriate to use ML modeling in high-dimensional data, where conventional regression models fail. In this study, the Cox model seems more appropriate for the prediction task because it handles time-to-event data.

Predicting future events is challenging; because of various innovations, treatment, and health care standard improvements, developing prediction models for future adverse events is like attempting to hit a moving target. The ambition of any prediction model should be to identify individuals who are at high risk of an adverse event because this would provide accurate prognostic information that would enable clinicians to offer individually tailored treatments to their patients and, hopefully, change the course of the poor prognostic predictions the model makes. Thus, successfully applied prediction models are not required to precisely predict what will happen; rather, they should yield probability estimates of what could happen.

Precise prognostic tools are of great importance for a heterogenous disease such as IgAN. Several IgAN prediction models have been developed over previous decades, but none of them has been put into widespread use [3, 31, 32, 37–42]. Currently, multiple ongoing clinical trials [8, 9] emphasize the significance of robust prognostic tools for selecting patients to participate in studies as well as those who will be eligible for specific treatments at a later stage. Schena *et al.* recently showed that their ANN model could be beneficial for matching patients to therapies [43], while Barbour *et al.* found through simulations that IIGAN-PT can be used to select patients specifically for immunosuppression treatments [44]. One must also not forget the value of identifying low-risk patients, which could prevent unnecessary treatments with potentially negative side effects and thus reduce health care costs.

The main limitations of this study are the retrospective study design, small sample size, and homogenous ethnicity. This study benefits from a long follow-up time compared with previous validation studies, however.

CONCLUSION

We have externally validated IIGAN-PT, developed by Barbour *et al.*, and the ANN model, developed by Schena *et al.* IIGAN-PT performed well, and clinicians can use it to predict the risk of a 50% decline in eGFR from baseline or ESKD for patients with IgAN. It should not be used, however, for risk prediction in patients with an eGFR under 15 mL/min/1.73 and active lesions. The ANN model has acceptable discrimination but tends to underestimate risk for patients with higher risk and overestimate risk in the lower risk categories when predicting risk of ESKD. Notably, the model's ability to predict time to ESKD is insufficient in this validation cohort.

FUNDING

No funding was obtained for this study.

AUTHORS' CONTRIBUTIONS

Y.L.H. and N.G.L. identified the study plot, collected clinical data, conducted statistical analysis, and drafted and approved the manuscript. R.B. identified the study plot, was responsible for the ethics committee application, conducted statistical analysis, and drafted and approved the manuscript. L.B. identified and reclassified biopsies and reviewed and approved the manuscript. L.S.N. drafted the ethics committee application, organized the database, and reviewed and approved the manuscript. T.K. collected clinical data, conducted statistical analysis, and drafted and approved the manuscript.

DATA AVAILABILITY STATEMENT

The data underlying this article cannot be shared publicly because of Norwegian regulations and the privacy of the individuals who participated in the study. The data may be shared upon reasonable request to the corresponding author if the request is accepted by the Regional Committee for Medical and Health Research Ethics and the local data protection official.

CONFLICT OF INTEREST STATEMENT

Thomas Knoop has been principal investigator and Yngvar Lunde Haaskjold has been sub-investigator for the Novartis LNP023 × 2203/APPLAUSE-IgAN trial at Haukeland University Hospital. The remaining authors have no conflict of interest to declare.

REFERENCES

1. Schena FP, Nistor I. Epidemiology of IgA nephropathy: a global perspective. *Semin Nephrol* 2018;**38**:435–42. https://doi.org/10.1016/j.semnephrol.2018.05.013
2. Wyatt RJ, Julian BA. IgA nephropathy. *N Engl J Med* 2013;**368**:2402–14. https://doi.org/10.1056/NEJMra1206793
3. Bartosik LP, Lajoie G, Sugar L *et al.* Predicting progression in IgA nephropathy. *Am J Kidney Dis* 2001;**38**:728–35. https://doi.org/10.1053/ajkd.2001.27689
4. D'Amico G. Natural history of idiopathic IgA nephropathy and factors predictive of disease outcome. *Semin Nephrol* 2004;**24**:179–96. https://doi.org/10.1016/j.semnephrol.2004.01.001
5. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society; Cattran DC, Coppo R, Cook HT *et al.* The Oxford Classification of IgA Nephropathy: rationale, clinicopathological correlations, and classification. *Kidney Int* 2009;**76**:534–45. https://doi.org/10.1038/ki.2009.243
6. Working Group of the International IgA Nephropathy Network and the Renal Pathology Society; Roberts ISD, Cook HT, Troyanov S *et al.* The Oxford Classification of IgA Nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int* 2009;**76**:546–56. https://doi.org/10.1038/ki.2009.168
7. Trimarchi H, 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**:1014–21. https://doi.org/10.1016/j.kint.2017.02.003
8. Selvaskandan H, Cheung CK, Muto M *et al.* New strategies and perspectives on managing IgA nephropathy. *Clin Exp Nephrol* 2019;**23**:577–88. https://doi.org/10.1007/s10157-019-01700-1
9. Cheung CK, Rajasekaran A, Barratt J *et al.* An update on the current state of management and clinical trials for IgA nephropathy. *J Clin Med* 2021;**10**:2493. https://doi.org/10.3390/jcm10112493
10. Floege J, Rauen T, Tang SC. Current treatment of IgA nephropathy. *Semin Immunopathol* 2021;**43**:717–28. https://doi.org/10.1007/s00281-021-00888-3
11. Barbour SJ, Coppo R, Zhang H *et al.* Evaluating a new international risk-prediction tool in IgA nephropathy. *JAMA Intern Med* 2019;**179**:942–52. https://doi.org/10.1001/jamainternmed.2019.0600
12. Rovin BH, Adler SG, Barratt J *et al.* Executive summary of the KDIGO 2021 Guideline for the Management of Glomerular Diseases. *Kidney Int* 2021;**100**:753–79. https://doi.org/10.1016/j.kint.2021.05.015
13. Zhang Y, Guo L, Wang Z *et al.* External validation of the international risk-prediction models of IgA nephropathy in an Asian-Caucasian cohort. *Kidney Int Rep* 2020;**5**:1753–63. https://doi.org/10.1016/j.ekir.2020.07.036
14. Zhang J, Huang B, Liu Z *et al.* External validation of the international IgA nephropathy prediction tool. *Clin J Am Soc Nephrol* 2020;**15**:1112–20. https://doi.org/10.2215/CJN.16021219
15. Ouyang Y, Zhao Z, Li G *et al.* A validation study comparing risk prediction models of IgA nephropathy. *Front Immunol* 2021;**12**:753901. https://doi.org/10.3389/jimmu.2021.753901
16. Schena FP, Anelli VW, Trotta J *et al.* Development and testing of an artificial intelligence tool for predicting end-stage kidney disease in patients with immunoglobulin a nephropathy. *Kidney Int* 2021;**99**:1179–88. https://doi.org/10.1016/j.kint.2020.07.046
17. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;**19**:453–73. https://doi.org/10.1002/(sici)1097-0258(20000229)19:4⟨453::aid-sim350⟩3.0.co;2-5
18. Ramspek CL, Jager KJ, Dekker FW *et al.* External validation of prognostic models: what, why, how, when and where? *Clin Kidney J* 2021;**14**:49–58. https://doi.org/10.1093/ckj/sfaa188
19. Haaskjold YL, Bjørneklett R, Bostad L *et al.* Utilizing the MEST score for prognostic staging in IgA nephropathy. *BMC Nephrol* 2022;**23**:1–11. https://doi.org/10.1186/s12882-021-02653-y
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Method* 2013;**13**:1–15. https://doi.org/10.1186/1471-2288-13-33
21. Pérez Fernández S, Martínez Camblor P, Filzmoser P *et al.* nsROC: an R package for non-standard ROC curve analysis. *R Journal* 2018;**10**:55–77.
22. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? *J Clin Epidemiol* 2020;**118**:93–9. https://doi.org/10.1016/j.jclinepi.2019.09.016
23. Barbour SJ, Feehally J. Predicting the future in immunoglobulin A nephropathy: a new international risk prediction tool. *Nephrol Dial Transplant* 2020;**35**:379–82. https://doi.org/10.1093/ndt/gfz181
24. Haas M, Verhave JC, Liu Z-H *et al.* A multicenter study of the predictive value of crescents in IgA nephropathy. *J Am Soc Nephrol* 2017;**28**:691–701. https://doi.org/10.1681/ASN.2016040433

25. Wen D, Tang Y, Tan L *et al.* Sex disparities in IgA nephropathy: a retrospective study in Chinese patients. *Int Urol Nephrol* 2020;**53**:315–323. https://doi.org/10.1007.s11255-020-02631-7
26. Jullien P, Laurent B, Berthoux F *et al.* Repeat renal biopsy improves the Oxford Classification-based prediction of immunoglobulin A nephropathy outcome. *Nephrol Dial Transplant* 2020;**35**:1179–86. https://doi.org/10.1093/ndt/gfy341
27. Inker LA, Heerspink HJL, Tighiouart H *et al.* GFR slope as a surrogate end point for kidney disease progression in clinical trials: a meta-analysis of treatment effects of randomized controlled trials. *J Am Soc Nephrol* 2019;**30**:1735–45. https://doi.org/10.1681/ASN.2019010007
28. Inker LA, Mondal H, Greene T *et al.* Early change in urine protein as a surrogate end point in studies of IgA nephropathy: an individual-patient meta-analysis. *Am J Kidney Dis* 2016;**68**:392–401. https://doi.org/10.1053/j.ajkd.2016.02.042
29. Kidney Disease: Improving Global Outcomes (KDIGO) Glomerulonephritis Work Group. KDIGO Clinical Practice Guideline for Glomerulonephritis. *Kidney Int Suppl* 2012;**2**:139–274. https://doi.org/10.1038/kisup.2012.12
30. Bülow RD, Dimitrov D, Boor P *et al.* How will artificial intelligence and bioinformatics change our understanding of IgA in the next decade? *Semin Immunopathol* 2021;**43**:739–52. https://doi.org/10.1007.s00281-021-00847-y
31. Chen T, Li X, Li Y *et al.* Prediction and risk stratification of kidney outcomes in IgA nephropathy. *Am J Kidney Dis* 2019;**74**:300–9. https://doi.org/10.1053/j.ajkd.2019.02.016
32. Liu Y, Zhang Y, Liu D *et al.* Prediction of ESRD in IgA nephropathy patients from an Asian cohort: a random forest model. *Kidney Blood Press Res* 2018;**43**:1852–64. https://doi.org/10.1159/000495818
33. Chen JH, Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. *N Engl J Med* 2017;**376**:2507. https://doi.org/10.1056/NEJMp1702071
34. Bertsimas D, Dunn J, Velmahos GC *et al.* Surgical risk is not linear: derivation and validation of a novel, user-friendly, and machine-learning-based predictive optimal trees in emergency surgery risk (Potter) calculator. *Ann Surg* 2018;**268**:574–83. https://doi.org/10.1097/SLA.0000000000002956
35. Orfanoudaki A, Chesley E, Cadisch C *et al.* Machine learning provides evidence that stroke risk is not linear: the non-linear Framingham stroke risk score. *PLoS One* 2020;**15**:e0232414. https://doi.org/10.1371/journal.pone.0232414
36. Siontis GC, Tzoulaki 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. https://doi.org/10.1016/j.jclinepi.2014.09.007
37. Berthoux F, Mohey H, Laurent B *et al.* Predicting the risk for dialysis or death in IgA nephropathy. *J Am Soc Nephrol* 2011;**22**:752–61. https://doi.org/10.1681/ASN.2010040355
38. Tanaka S, Ninomiya T, Katafuchi R *et al.* Development and validation of a prediction rule using the Oxford Classification in IgA nephropathy. *Clin J Am Soc Nephrol* 2013;**8**:2082–90. https://doi.org/10.2215/CJN.03480413
39. Xie J, Kiryluk K, Wang W *et al.* Predicting progression of IgA nephropathy: new clinical progression risk score. *PLoS One* 2012;**7**:e38904. https://doi.org/10.1371/journal.pone.0038904
40. Goto M, Wakai K, Kawamura T *et al.* A scoring system to predict renal outcome in IgA nephropathy: a nationwide 10-year prospective cohort study. *Nephrol Dial Transplant* 2009;**24**:3068–74. https://doi.org/10.1093/ndt/gfp273
41. 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**:2800–8. https://doi.org/10.1093/ndt/gfl342
42. Miyabe Y, Karasawa K, Akiyama K *et al.* Grading system utilising the total score of Oxford Classification for predicting renal prognosis in IgA nephropathy. *Sci Rep* 2021;**11**:3584. https://doi.org/10.1038/s41598-021-82967-x
43. Schena FP, Manno C, Anelli VW *et al.* MO260 Performance analysis of an artificial neural network tool to predict ESKD in patients with immunoglobulin A nephropathy (IGAN). *Nephrol Dial Transplant* 2021;**36**(suppl 1):gfab104.0018. https://doi.org/10.1093/ndt/gfab104.0018
44. 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**:1009–19. https://doi.org/10.1016/j.kint.2020.04.042

Received: 14.3.2022; Editorial decision: 16.7.2022