The Improved Kidney Risk Score in ANCA Vasculitis for Clinical Practice and Trials

Sebastian Bate[1,2], Dominic McGovern[3,4,5,6], Francesca Costigliolo[7,8], Pek Ghe Tan[9,10], Vojtech Kratky[11,12], Jennifer Scott[13], Gavin Chapman[14], Nina Brown[15,16], Lauren Floyd[15,17], Benoit Brilland[18], Eduardo Martín-Nares[19], Mehmet Fethullah Aydın[20], Duha Ilyas[15,21], Arslan Butt[16], Eithne Nic an Riogh[13], Marek Kollar[22], Jennifer S Lees[3,4], Abdülmecit Yıldız[20], Andrea Hinojosa-Azaola[19], Ajay Dhaygude[17], Stephen A Roberts[1,2], Avi Rosenberg[23], Thorsten Wiech[24], Charles D Pusey[9,25], Rachel B Jones[5,6], David RW Jayne[5,6], Ingeborg Bajema[26], J Charles Jennette[27], Kate I Stevens[3,4], Jean Francois Augusto[18], Juan Manuel Mejía-Vilet[19], Neeraj Dhaun[14], Stephen McAdoo[9,25], Vladimir Tesar[11,12], Mark A Little[13], Geetha Duruvu[28], and Silke R Brix[1,21,29]

[1]Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust and [2]Centre for Biostatistics, Division of Population Health, Health Services Research, and Primary Care, University of Manchester, Manchester, UK
[3]Glasgow Renal and Transplant Unit, Queen Elizabeth University Hospital, and [4]School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, UK
[5]Department of Medicine, University of Cambridge, and [6]Vasculitis Clinic, Department of Renal Medicine, Addenbrooke's Hospital, Cambridge, UK
[7]Division of Nephrology, Dialysis and Transplantation, University of Genova, and [8]Department of Internal Medicine and IRCCS Ospedale Policlinico San Martino, Genova, Italy
[9]Imperial College Renal and Transplant Centre, Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK
[10]Renal Unit, Northern Health, Victoria, Australia
[11]1st Faculty of Medicine, Charles University, and [12]Department of Nephrology, General University Hospital, Prague, Czechia
[13]Trinity Kidney Centre, Trinity College Dublin, Ireland
[14]University/BHF Centre for Cardiovascular Science, University of Edinburgh and Department of Renal Medicine, Royal Infirmary of Edinburgh, Edinburgh, UK
[15]Division of Cardiovascular Sciences, University of Manchester, Manchester, UK
[16]Renal Department, Salford Royal Hospital, Northern Care Alliance NHS Foundation Trust, UK
[17]Renal Department, Royal Preston Hospital, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK
[18]Service de Néphrologie-Dialyse-Transplantation, CHU d'Angers, Angers, France
[19]Departments of Immunology and Rheumatology, Nephrology and Mineral Metabolism, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
[20]Division of Nephrology, Bursa Uludağ University School of Medicine, Bursa, Turkey
[21]Renal, Transplantation and Urology Unit, Manchester University NHS Foundation Trust, Manchester, UK
[22]Department of Pathology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic
[23]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[24]University Medical Center Hamburg-Eppendorf, Institute of Pathology, Hamburg, Germany
[25]Centre for Inflammatory Disease, Department of Immunology and Inflammation, Imperial College London, London, UK
[26]Department of Pathology, Groningen University Medical Center, Groningen, The Netherlands
[27]Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[28]Division of Nephrology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[29]Division of Cell Matrix Biology and Regenerative Medicine, University of Manchester, UK

Running title: AKRiS, the Improved ANCA Risk Score

Correspondence: Dr Silke R Brix, Renal, Urology and Transplantation Unit, Manchester University NHS Foundation Trust, Manchester Royal Infirmary, Oxford Road, Manchester, M13 9WL, United Kingdom, Phone 0044 161 276 4540, Email silke.brix@mft.nhs.uk

SIGNIFICANCE STATEMENT

Reliable prediction tools are needed to personalize treatment in anti-neutrophil cytoplasmic antibody (ANCA) glomerulonephritis. Over 1,500 patients were collated in an international longitudinal study to revise the ANCA Kidney Risk Score. The score showed satisfactory performance, mimicking the original study (Harrell's C=0.779). In the development cohort of 959 patients, no additional parameters aiding the tool were detected, but replacing the glomerular filtration rate with creatinine identified an additional cut-off. The parameter interstitial-fibrosis-tubular-atrophy was modified to allow wider access, risk points were reweighted, and a fourth risk group was created, improving predictive ability (C=0.831). In the validation, the new model performed similarly well with excellent calibration and discrimination (n=480, C=0.821). The revised score optimizes prognostication for clinical practice and trials.

ABSTRACT

Background:

Reliable prediction tools are needed to personalize treatment in anti-neutrophil cytoplasmic antibody (ANCA) glomerulonephritis. A retrospective international longitudinal cohort was collated to revise the ANCA Renal Risk Score (ARRS).

Methods:

Primary endpoint was end-stage kidney disease with patients censored at last follow-up. Cox Proportional Hazards were used to reweight risk factors. Kaplan-Meier curves, Harrell's C statistic, Receiver Operating Characteristics and calibration plots were used to assess model performance.

Results:

Of 1591 patients, 1439 were included in the final analyses, 2:1 randomly allocated per centre to development and validation cohorts (52% male, median age 64 years). In the development cohort (n=959), the ARRS was validated, calibrated and parameters were reinvestigated modifying interstitial fibrosis and tubular atrophy (IFTA) allowing semiquantitative reporting. An additional cut-off for kidney function (K) was identified and serum creatinine replaced glomerular filtration rate (K0: < 250 µmol/l = 0, K1: 250-450 µmol/l = 4, K2: > 450 µmol/l = 11 points). The risk points for the percentage of normal glomeruli (N) and IFTA (T) were reweighted (N0: > 25% = 0, N1: 10-25% = 4, N2: < 10% = 7, T0: none/mild or < 25% = 0, T1: ≥ mild-moderate or ≥ 25% = 3 points), and four risk groups created: low (0 – 4 points), moderate (5 – 11), high (12 – 18) and very high (21). Discrimination was C=0.831, and the three-year kidney survival was 96%, 79%, 54%, and 19%, respectively. The revised score performed similarly well in the validation cohort with excellent calibration and discrimination (n=480, C=0.821).

Conclusion:

The updated score optimizes clinicopathologic prognostication for clinical practice and trials.

INTRODUCTION

Anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitides (AAV) are autoimmune disorders characterized by inflammation of small blood vessels resulting in irreversible organ damage. Kidney involvement is common and disease presentation is heterogeneous. Kidney failure is a major predictor for adverse patient outcomes, and the development of effective, reliable prognostication tools is crucial to enable personalized vasculitis care.

Bajema et al. highlighted in 1996 the importance of characterizing and quantifying histopathological lesions to predict kidney outcomes in vasculitis[1]. Cohort studies detected the importance of unaffected, normal glomeruli, alongside tubular atrophy and interstitial fibrosis, as prognostic markers[2-7]. Thereafter, prospective analyses confirmed the percentage of normal glomeruli, the degree of tubulointerstitial changes and arteriosclerosis as having predictive ability[8,9]. Formulae were proposed, combining kidney function at presentation with histopathological features to improve outcome prediction but were not widely adopted into clinical practice[8,9]. A chronicity index score, combining four histopathological changes of chronic inflammation (global glomerulosclerosis, sclerosed crescents, interstitial fibrosis and tubular atrophy) was described in 2014[10]. The chronicity index score inversely associated with dialysis independence at four months, but no futility threshold for therapeutic intervention was detected.

A histopathological classification for ANCA glomerulonephritis (ANCA GN) was proposed by Berden et al. in 2010, focusing exclusively on glomerular pathology; biopsies were described as focal, crescentic, sclerotic, or mixed depending on the majority of lesions[11-14]. The focal class demonstrated the best kidney survival, highlighting unaffected glomeruli as a most powerful predictor. Attractive due to its simplicity, the classes lacked discriminative power to predict kidney survival and the focus of the classification lies on distinguishing histological phenotypes[15-20].

The Mayo Clinical Chronicity Score (MCCS) proposed a standardized reporting of chronic changes in 2017 using the parameters global glomerulosclerosis, interstitial fibrosis, and

tubular atrophy and arteriosclerosis[21]. The score was not designed for a specific disease but as a consensus assessment for chronic damage. The MCCS was validated on a single center retrospective ANCA GN cohort demonstrating the association of chronic changes with decreased recovery of kidney function and increased risk of end-stage kidney disease (ESKD)[22].

The ANCA Renal Risk Score (ARRS) was developed on a prospective multicenter cohort in 2018[23]. It originated from a detailed investigation of glomerular and tubulointerstitial lesions, quantifying eight glomerular and four tubulointerstitial changes, to determine their predictive power. The proportion of normal glomeruli was superior to all combinations of lesions and combining pathological and biochemical data increased risk prediction of ESKD further. The ARRS is an aggregate of individually weighted parameters using the percentage of normal glomeruli, the degree of interstitial fibrosis and tubular atrophy (IFTA) in the kidney biopsy and the estimated glomerular filtration rate (eGFR, ml/min/1.73 m^2)[24].

The ARRS was found to have excellent discrimination and since its publication, multiple retrospective cohort studies have been performed on cumulatively over 2000 patients across three continents[25-39]. It was recently incorporated into a proposed tool to identify patients that may benefit from plasma exchange following the results of the PEXIVAS trial[40,41]. A recent meta-analysis of 11 ARRS validations with a median number of 130 patients per study (range 37 – 252) highlighted the need for larger-scale studies[38].

Cox Proportional Hazards (PHs) models allow easy comparison of discriminative ability through the use of Harrell's C[42]. Baseline survival, critical to calibration, is rarely reported alongside prognostic models and calibration of prognostic models therefore cannot be performed[43-47] despite improved guidance for the development and reporting of prognostic models by the TRIPOD statement[47,48]. Calibration tests the accuracy of a model by comparing a new test result with the standard. No calibration was performed in the eleven validation studies reported in the meta-analysis[20,23,25-27,32,33,35,37,49,50]. To widely adopt the score into practice, predictions need to be reliable and generalizable to different populations[51].

Therefore, we collated a cohort of over 1500 patients from 11 referral centers and national registries on three continents with up to 20 years of follow-up data to I) validate and calibrate the ARRS, II) investigate histological and biochemical parameters on a larger cohort for parameter accessibility and to determine whether additional predictors could be added to the model, III) refine risk parameters and risk groups, IV) propose, and V) validate and calibrate an improved score ready to assist clinicians in the management of ANCA GN and for use in a prospective trial of risk stratification.

METHODS

Patient cohort

This study was a retrospective longitudinal cohort study. Data were collected from eleven kidney referral centers and registries (Maine-Anjou Registry, Angers, France; Baltimore, USA; Bursa, Turkey; London, Manchester, Preston, Salford, UK; Mexico City, Mexico; Prague, Czechia; and the Irish and Scottish registries). Patients with biopsy-proven ANCA glomerulonephritis (GN) diagnosed from January 1987 to December 2021, with follow-up data recorded to February 2022, were included. Inclusion criteria were: age ≥18 years, pauci-immune glomerulonephritis with acute and/or chronic vasculitic lesions consistent with ANCA GN on biopsy, data for the three components of the ARRS and twelve weeks of follow-up. ANCA were detected by ELISA for myeloperoxidase (MPO) or proteinase 3 (PR3). Patients with double positive MPO and PR3 antibodies and additional other immune-mediated kidney diseases, e.g., anti-glomerular basement membrane (GBM) disease, IgA and lupus nephritis were excluded. The sample size was pragmatic based on all eligible patients from each center. The study was performed in accordance with the Declaration of Helsinki. Data were collected following the Caldicott principles and guidelines of the respective local ethics committees: Manchester Biobank REC 16/NW/0119, West of Scotland REC 5 20/WS/0181, St James'/Tallaght REC 019-09 List 33 (040), Angers University Hospital CE 2020/84, Johns Hopkins IRB00090103, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán IRE-2928-19-20-1, Bursa Uludağ University Faculty of Medicine 29.03.2016, 2016-5/6.

The endpoint was ESKD (measured by time to ESKD) and was defined as the need for kidney replacement therapy (KRT) for at least 12 weeks and its continuation to the last follow-up assessed by the patients' clinical teams. 'Time-zero' was the date of the kidney biopsy. Censoring for the endpoint was at last follow-up, end of study or death. Kidney function was measured in serum creatinine and eGFR, the latter using the chronic kidney disease epidemiology collaboration, CKD-EPI, equation without race.[24]

Kidney histology

Normal glomeruli were defined as glomeruli without any scarring, crescents or fibrinoid necrosis within the tuft. Cellular crescents were defined according to the consensus of the Renal Pathology Society (as crescents with a composition of more than 75% cells and fibrin and less than 25% fibrous matrix[52]). The original ARRS was calculated combining the individually weighted point scores for eGFR at presentation (G0: >15 ml/min/1.73m^2, G1: ≤15 ml/min/1.73m^2), percentage of normal glomeruli (N0: >25%, N1: 10-25%, N2: <10%) and the percentage of IFTA (T0: IFTA <25%, T1: IFTA ≥25%). Each parameter was assigned points as previously described (G0=0, G1=3, N0=0, N1=4, N2=6, T0=0, T1=2) and their summation produced the total score determining the allocation of patients into one of the three risk groups: low (0), moderate (2 - 7), and high (8 - 11 points).

Model building and statistics

Data were not imputed, and models were built on a complete-case basis as the amount of missing data (after excluded patients) was negligible (<0.01%).

Stratified by center, the cohort was randomly split in a ratio of 2:1 to create development and validation sets. First, the development cohort was used for the external validation and calibration of the original ARRS (objective I). Then, the development cohort was employed to investigate the accessibility of the score parameters (objective II) and to derive alternative models refining the score (objective III). Proposing the new score (objective IV), the validation cohort was used for the validation and calibration of the new models (objective V).

Model performance was assessed for discrimination using Harrell's *C* statistic, a generalization of the area under the receiver-operating characteristic for Cox models[42] as well as Receiver Operating Characteristic (ROC) curves. Calibration was performed as advised by McLernon et al[53] (objectives I and V). An expert panel of eight experienced nephrologists and nephropathologists was formed to judge the accessibility and adaptation of parameters (objective II). Multivariable models were built using Cox proportional hazard models (objective III). The panel investigated each significant parameter of the multivariable analysis for its suitability as a risk parameter, and backwards elimination was used with a threshold of P=0.1 to select among these (objective III). An *exploratory* regression tree built with R Package *rpart*[54] (using the default parameters) was used to determine whether differing cutoffs would improve the prediction model (objective III). The cutoffs along with the clinical judgement of the panel were used to form the risk groups to maximize utility.

We generated new risk points (objective IV) for the model with categorical covariates. For this, we used the β-coefficients from the Cox models and divided them by the smallest coefficient of the model to retain proportionality. We then tripled and rounded the values to the nearest integer. The β-coefficients were tripled rather than doubled as previously to highlight the difference between the weight of normal glomeruli and IFTA.

An additional continuous model using the same parameters as the final model was derived without cut-offs (objective IV). Baseline survival equations were calculated using the methods proposed by Royston and Altman.[47] Integrated in a web application, the continuous model was designed to provide survival chances informing clinicians and patients of a detailed prognosis.

The final models were compared with the original model using the validation cohort. The difference in performance was evaluated using a paired bootstrap sample of the C statistics (1000 replicates; objective V). Time-dependent Receiver Operating Characteristic (ROC) curves were calculated for 1, 3, and 5 years using the method described by Blanche et al[55] (objective V). The proportional hazards assumption was assessed to investigate the consistency of the risk prediction over time using the Grambsch and Therneau global test[56]

and visually reviewing the Kaplan Meier plots[53]. A competing risk sensitivity analysis investigating death prior to ESKD as a competing risk was performed using the Fine and Grey approach.[57] The validation cohort was split on either side of January 1, 2015, to assess the influence of different time periods on model performance (objective V).

Analyses were performed using R v4.2.1 (R Core Team, Vienna, Austria). The study was reported according to the TRIPOD guidelines.[58,59]

RESULTS

Basic Characteristics

A total of 1591 patient records were submitted by their centers of which 1439 patients met inclusion criteria for the validation (Supplemental Figure 1a). To create development and validation sets, the 1439 patients were randomly split in a 2:1 ratio by center into two cohorts with a development cohort consisting of 959 patients and a validation cohort of 480 patients. The basic characteristics of the cohorts are depicted in Table 1. Median follow-up was 3.6 years (interquartile range 1.1-5.9 years), 33% of patients had at least five years and 8.4% had at least ten years of follow-up (Table 2). Patient flow through the study is detailed in Supplemental Figure 1b.

Patient Outcomes

In the development cohort, 207 patients (21.6%) progressed to ESKD compared to 118 patients (24.6%) in the validation cohort (Table 2). A total of 185 patients of the development cohort required KRT at presentation and 83 (44.9%) of these recovered kidney function. In the validation set, 53/106 (50%) recovered kidney function. 206 (21.5%) and 109 (22.7%) patients died in the development and validation cohorts, respectively.

A total of 1114 patients (77.4%) were ESKD-free at their last follow-up (752 patients, 78.4% in the development and 362 patients, 75.4% in the validation cohort). A total of 199 (13.8%)

of these ESKD-free patients died during follow-up (135 patients, 14.1% in the development cohort and 64 patients, 13.3% in the validation cohort).

Objective I: Validation and Calibration of the Original ARRS

Validation of the Original ARRS

For the validation of the original ARRS, the development cohort was restricted to biopsies reporting the degree of IFTA by percentage (n=456). The discrimination of the ARRS was 0.779 (95% Confidence interval, CI, 0.739-0.819). The 3-year kidney survival was 75.8% (CI 71.8-80.1%). Regarding the risk groups, the 3-year kidney survival was 93.7% (CI 88.9-98.7%), 82.1% (76.7-87.8%), 49.6% (41.3-59.6%) in the low, moderate, and high-risk groups, respectively. The kidney survival of the three risk groups is depicted in Figure 1.

Calibration of the Original ARRS

Data from the original cohort[23] were re-analyzed to determine survival function and calibration. The predictions were based on the 115 patients in the original cohort. Supplemental Figure 2 demonstrates the calibration plots at 1-, 3- and 5-year follow-up. Calibration was reasonable for an external cohort (year 3 slope: 0.66, intercept: 0.27); the observed survival was slightly higher than predicted.

Objective II: Accessibility and Adaptation of Risk Parameters

Risk Parameter Interstitial Fibrosis and Tubular Atrophy (IFTA)

The reporting of IFTA is often performed in a semiquantitative manner. To allow worldwide availability of the risk score, we investigated the congruency of semiquantitative and quantitative reporting of IFTA. Supplemental Figure 3 demonstrates the correlation between quantitative and semiquantitative reporting of IFTA in the development cohort (n=269). IFTA reporting showed a satisfactory and close to perfect congruency in the low gradings of none and mild. Therefore, we retained the original cut-off of IFTA ≥ 25% and allowed the semiquantitative alternative of reporting with the cut-off of IFTA ≥ mild to moderate.

Revalidation of the Simplified Original ARRS

For the validation of the simplified original ARRS, the whole development cohort was used. The discrimination of the ARRS in this cohort was C=0.800 (CI 0.768-0.832, P<0.001), comparable to the original development cohort in 2018 (C=0.832). The 3-year kidney survival of the development cohort was 80.9% (CI 78.3-83.6%). Regarding the risk groups, the 3-year kidney survival was 94.8% (CI 91.9-97.9%), 86.1% (82.8-89.5%), 52.4% (45.7-60.0%) in the low, moderate, and high-risk groups, respectively. The kidney survival of the three risk groups is depicted in Supplemental Figure 4a.

Calibration of the Simplified Original ARRS

Supplemental Figure 4b demonstrates the calibration plots at 1-, 3- and 5-year follow-up. Calibration was again reasonable for an external cohort (year 3 slope: 0.7, intercept: 0.29); the observed survival was slightly higher than predicted. The baseline survival for a patient with ARRS=0 is detailed in Supplemental Equation 1.

There was strong agreement in the model between the original score with percentage-IFTA only and the simplified original score with all patients, demonstrating that semi-quantitative grading can be used in lieu of percentage grading.

Objective III: Review of Predictors of Kidney Survival

On univariable analyses, the percentage of normal glomeruli, cellular crescents, IFTA and the kidney function at presentation were associated with kidney outcome (Supplemental Table 1a). On further multivariable analyses, using Cox proportional hazard models, three parameters remained predictors of ESKD (Supplemental Table 1b). The logarithm of serum creatinine at presentation demonstrated a strong association with ESKD (hazard ratio (HR) 3.541, 95% CI 2.675-4.687, P<0.001); and the presence of a higher percentage of normal glomeruli associated with a reduced risk of ESKD (HR 0.977, CI 0.967-0.987, P<0.001). There was also a risk associated with more severe IFTA (HR 1.673, CI 1.159-2.415, P=0.006). Other parameters, such as the ANCA antibody subtype PR3 versus (vs) MPO failed to show an

association with outcome (HR 0.892, CI 0.637-1.250, P=0.506). While there was an association between the percentage of cellular crescents and kidney survival on the univariable analysis, there was no detectable association after adjustment for the percentage of normal glomeruli (HR 0.996, CI 0.991-1.002, P=0.175).

Objective IV: Revision and Modification of the ARRS

In the development cohort (n=959), backwards elimination with a threshold of P=0.1 was performed and no other clinico-pathological parameters were associated with ESKD on multivariable models. Consequently, no additional parameters were incorporated into the revised ARRS.

New Revised Model: the Score with Creatinine and Four Risk Groups

In acute kidney disease, eGFR compresses a broad range of creatinine values into a smaller range of eGFR values. Taking age, sex and ethnicity components into the equation does not improve the reliability of the measured function when there are acute changes in kidney function but reduces the discriminative power of the biochemical measures of kidney function. Therefore, a model with creatinine instead of eGFR was developed to maximize reliability[60]. On a regression tree analysis, the primary node was detected at a creatinine of 454 µmol/l. Further cut-offs were identified at 239 and 967 µmol/l, the latter was not incorporated due to the small sub-cohort of patients (Supplemental Figure 5). Additional cut-offs for the percentage of normal glomeruli were similar to the original tree analysis from 2018 at 7.3% and 27% and therefore no changes were deemed necessary. Using the above-described method, risk points were assigned as follows for creatinine: K0: < 250 µmol/l = 0 points; K1: 250-450 µmol/l = 4 points; K2: > 450 µmol/l = 11 points; percentage of normal glomeruli: N0: >25% = 0 points; N1: 10-25% = 4 points; N2: < 10% = 7 points; IFTA: T0: none/mild or < 25% = 0 points; T1: IFTA ≥mild to moderate or ≥25% = 3 points (Table 3a). An interaction between terms was considered but none was found. Risk groups were defined separating low (0-4), moderate (5-11), high (12-18) and very high-risk groups (21) (as depicted in Table 3b; note that the risk

points 1, 2, 19, 20 are not calculable). Figure 2 demonstrates the discriminative power and kidney survival of the risk groups of the new ANCA Kidney Risk Score (AKRiS) over time. The 3-year kidney survival was 96.0% (CI 94.1-98.0%), 79.4% (74.3-84.8%), 53.8% (45.0-64.3%), and 18.5% (10.8-31.8%) in the low-, moderate-, high- and very high-risk groups respectively (C=0.831, CI 0.801 – 0.861, P<0.001).

Continuous Score Model with Percentage Risk

A complementary continuous model was developed using the logarithm of creatinine, the percentage of normal glomeruli and the IFTA categories as covariates. The model demonstrated excellent discrimination (C=0.833; CI 0.805 – 0.861, P<0.001, Table 3c). The model allowed detailed estimates of kidney survival but required significant computation. We created a webpage enabling the calculation of the risk percentage for ESKD at 1, 3 and 5 years for the individual patient at the time of presentation using the three parameters serum creatinine, the percentage of normal glomeruli and the IFTA category. Supplemental Figure 6 has the link and QR code to access the software.

Objective V: Validation and Calibration of the AKRiS

Performance of the Score in the Validation Cohort

The discrimination of the AKRiS in the validation cohort was C=0.821 compared with C=0.783 for the original ARRS (difference = 0.038, CI 0.012 – 0.064, P=0.003). The time-dependent area under the curves (AUCs) for the AKRiS and the original ARRS were calculated. These showed sustained and consistently improved performance (Supplemental Figure 7). The AUCs were 0.844 vs 0.803 (P=0.026), 0.839 vs 0.812 (P=0.172) and 0.820 vs 0.777 (P=0.051) for 1, 3 and 5 years, respectively.

The 3-year kidney survival of the validation cohort was 76.6% (CI 72.6-80.7%). In regard to the risk groups of the AKRiS, the 3-year kidney survival was 96.5% (CI 94.0-99.1%), 75.1% (67.6-83.4%), 49.3% (39.4-61.8%), and 12.0% (4.2-34.7%) in the low-, moderate-, high- and

very high-risk groups, respectively (C=0.821, CI 0.744 – 0.846, P<0.001). In the calibration of the new models, there was strong agreement between the expected and the observed events in the validation cohort (year 3 slope: 0.985, intercept: 0.008). Supplemental Figure 8b demonstrates the calibration plots for the New and the Continuous AKRiS models at 1, 3 and 5 years of follow-up.

The proportional hazards assumption was assessed, and the Grambsch and Therneau global test[53,56] detected disproportionality due to the nature of the rapidly progressive disease (P<0.001). The non-proportionality demonstrated minor violations with robust predictions over time when inspected on the calibration plots.

Effects of Competing Risks and the Time of Presentation on the Performance of the AKRiS

In a sensitivity analysis, we used the Fine-Grey[57] method with death prior to ESKD as a competing risk. The AKRiS retained performance and calibration in the validation cohort when applied to a *competing-risks* setting (C=0.817 (0.787-0.847), Supplemental Figure 9). We investigated the influence of the score on patient survival and the model did not predict death (C=0.569).

The model was then tested on a temporal split by the presentation date of the validation cohort on January 1, 2015. The discrimination of the score was higher in patients presenting in the more recent time period with Harrell's $C_{\geq 2015}$=0.868 (CI 0.828-0.908) when compared to the time period prior to 2015 with Harrell's $C_{<2015}$=0.782 (CI 0.722-0.848, P=0.008, Supplemental Table 2).

DISCUSSION

Kidney and patient outcomes remain unsatisfactory in ANCA GN due to heterogeneity in disease presentation and treatment-related adverse events, with infection being the most common cause of death in the first year. Immunosuppressive therapy needs to be tailored to the individual patient to avoid treatment failure and minimize toxicity. Reliable prediction tools are urgently required to guide clinicians in treatment decisions. Consensus on relative risk will

also facilitate better outcome assessment in interventional trials through stratification of patients. Oncology has led the way to personalized therapy and established rigorous disease staging and patient fitness assessments to optimize treatment outcomes. There is a clear and urgent unmet need to personalize medicine in ANCA GN.

The ANCA Renal Risk Score (ARRS) was published in 2018 categorizing patients in risk groups according to histological and clinical parameters, discriminating kidney survival[23]. Validations of over 2000 patients on three continents demonstrated the reliability of the prediction tool. Here, we collated the largest cohort of ANCA GN patients with biopsy data to date, validated and estimated the calibration error in the ARRS and present a revised and improved score optimizing outcome prediction. The derivation and validation of the original score was missing long-term outcome data and previous validation studies have not provided meaningful assessments of calibration error. Collating a cohort of patients from different centres and national registries around the world allowed us to observe kidney survival in patients with up to 10-15 years of follow-up and to calibrate the original score. Gathering a large number of events enabled us to investigate our original risk parameters and assess the quality of the original model, updating and further improving the tool. The next step will be to assess the utility of the prediction tool to guide treatment.

We found no additional parameters aiding the tool but by switching from eGFR to creatinine, we identified an additional cut-off that improved discrimination. Numerous factors can interfere with laboratory measures of creatinine, but a wide range of creatinine values are comprised in fewer eGFR values diminishing the usability of eGFR in acute kidney injury (AKI). In rapidly progressive AKI, creatinine will take days before it reflects the true GFR but eGFR estimations were developed in patients with stable chronic kidney disease, are population-based and even less appropriate for use in AKI. We were also able to modify the parameter IFTA to allow wider access, to reweight the risk points (by factor 3) and to create a fourth risk group (very high-risk) providing more reliable predictions.

Different risk parameters for adverse outcome have been proposed in the past and anti-MPO antibody specificity has repeatedly been shown to associate with a higher risk for ESKD. In

our cohort, this effect was not significant after adjustment for other parameters. Despite the size of the cohort, anti-MPO specificity did not independently predict outcome and was not found to improve outcome prediction. Our findings would suggest the tool can be used in different populations with different basic characteristics[25,32].

Our study was limited by its retrospective design and the observation period including patients from 1987 to 2021 but no other design would have been feasible to achieve our aims of a large-scale validation with a long follow-up. We detected an improved model performance of the revised score over time highlighting the need for future updates of the prediction tool to adjust to changes in vasculitis treatment and patient survival. We were not able to assess the impact of ethnicity and extrarenal disease and aim to investigate these in the next update. Using real-world data has the benefit of creating data sets that include complex and severely ill patients that are usually not recruited to interventional trials. Using pathological findings for our scoring, interobserver variability needed to be factored in. In ANCA GN, reporting of glomerular and tubulointerstitial lesions has demonstrated significant variability and impacts the reliability of prediction tools. Here, we investigated the impact of cellular crescents on outcome prediction and did not detect any significant signal. A limitation is that we did not investigate other pathological lesions, e.g., intra- and extracapillary fibrinoid necrosis. However, the interobserver variability of this early lesion is significant and we believe that the impact on outcome prediction would have been diminished by its unreliable reporting. We also demonstrated the variability in IFTA reporting and the challenges of using that risk predictor. The higher degrees of IFTA exhibit substantial interobserver disagreement, making moderate to severe IFTA less suitable for prediction models. We showed congruency of quantitative and semiquantitative reporting of the lower degrees of IFTA; thus the use of the semiquantitative reporting will support a wider utility of the score.

A common measure of model performance is the C statistic, and in our case Harrell's C, the extension for time-to-event models, which in development ranges from 0.5 to 1. For models to be implemented in routine clinical practice, they require strong performance and clear clinical utility and usability. While one could add an ever-increasing number of predictors into a model,

it would be subject to overfitting as well as making the tool more difficult to use in practice. The designing of a prediction tool requires statistical and practical trade-offs. Van Royen et al. published several barriers preventing the implementation of clinical prediction models into practice[51] and we believe the AKRiS has overcome these: the score demonstrated that it is fit for purpose using routinely collected predictors on an easily measured outcome. With prospective trials incorporating the tool, there is a clear path to its widespread implementation and adoption.

Prognostic tools have failed to detect a treatment futility threshold in ANCA GN. Standard kidney biopsies encounter about 15-20 glomeruli per biopsy. This number seems unable to detect the cut-off of glomerular injury predictable of treatment failure in ANCA GN. Kidneys seem to function sufficiently to provide dialysis independence with a very small proportion of unaffected glomeruli. Additional factors such as hyperfiltration damage over time and relapsing disease impact long-term survival and are challenging to incorporate into current prediction tools. The highest-risk group demonstrated a three-year kidney survival of 18.5% and 12% in the development and validation cohorts. Of 54 patients who required kidney replacement therapy at diagnosis in the very high-risk group, seven recovered kidney function (13%). The AKRiS provides a valuable risk prediction for ESKD improving the education of healthcare professionals and patients. Given the predictive power of the model, the ultimate aim is the prospective use of the tool to inform therapy. At present, the tool can assist the clinical assessment when discussing risks and merits of different treatment strategies.

In conclusion, the ARRS provides a strong discriminative ability, and the update improves the existing paradigm. We recommend those using the original ARRS change to the updated AKRiS using creatinine. The revised and improved kidney score will provide stratification of patients for personalized treatments and in clinical trials and will allow clinicians to give reliable estimates of prognosis.

Funding

This work was supported by the Wellcome Trust Institutional Translational Partnership Award 222061/Z/20/Z to SRB.

Disclosures

The authors declare no financial or commercial conflicts of interest.

Acknowledgements

We would like to acknowledge the patients and participating physicians for enabling the collection of this data. We would also like to acknowledge the statistical input and discussions with Pierre Tennstedt and Eik Vettorazzi.

References

1. Bajema IM, Hagen EC, Hansen BE, et al. The renal histopathology in systemic vasculitis : an international survey study of inter- and intra-observer agreement. *Nephrol Dial Transplant*. 1996;11(10):1989-1995. doi:10.1093/ndt/11.10.1989
2. Bajema IM, Hagen EC, Hermans J, et al. Kidney biopsy as a predictor for renal outcome in ANCA-associated necrotizing glomerulonephritis. *Kidney Int*. 1999;56(5):1751-1758. doi:10.1046/j.1523-1755.1999.00758.x
3. Aasarød K, Bostad L, Hammerstrøm J, Jørstad S, Iversen BM. Renal histopathology and clinical course in 94 patients with Wegener's granulomatosis. *Nephrol Dial Transplantation*. 2001;16(5):953-960. doi:10.1093/ndt/16.5.953
4. Vergunst CE, van Gurp E, Hagen EC, et al. An index for renal outcome in ANCA-associated glomerulonephritis. *Am J Kidney Dis*. 2003;41(3):532-538. doi:10.1053/ajkd.2003.50115
5. Haroun MK, Stone JH, Nair R, Racusen L, Hellmann DB, Eustace JA. Correlation of Percentage of Normal Glomeruli with Renal Outcome in Wegener's Granulomatosis. *Am J Nephrol*. 2002;22(5-6):497-503. doi:10.1159/000065283
6. Kapitsinou PP, Ioannidis JPA, Boletis JN, et al. Clinicopathologic predictors of death and ESRD in patients with pauci-immune necrotizing glomerulonephritis. *Am J Kidney Dis*. 2003;41(1):29-37. doi:10.1053/ajkd.2003.50013
7. Neumann I, Kain R, Regele H, Soleiman A, Kandutsch S, Meisl FT. Histological and clinical predictors of early and late renal outcome in ANCA-associated vasculitis. *Nephrol Dial Transplantation*. 2005;20(1):96-104. doi:10.1093/ndt/gfh563
8. Hauer HA, Bajema IM, Van Houwelingen HC, et al. Determinants of outcome in ANCA-associated glomerulonephritis: A prospective clinico-histopathological analysis of 96 patients. *Kidney Int*. 2002;62(5):1732-1742. doi:10.1046/j.1523-1755.2002.00605.x
9. De Lind Van Wijngaarden RAF, Hauer HA, Bruijn JA, et al. Clinical and histologic determinants of renal outcome in ANCA-associated vasculitis : A prospective analysis of 100 patients with severe renal involvement. *J Am Soc Nephrol*. 2006;17(8):2264-2274. doi:10.1681/ASN.2005080870
10. Lee T, Gasim A, Derebail VK, et al. Predictors of treatment outcomes in ANCA-associated vasculitis with severe kidney failure. *Clin J Am Soc Nephrol*. 2014;9(5):905-913. doi:10.2215/CJN.08290813
11. Berden AE, Ferrario F, Bruijn JA, et al. Histopathologic Classification of ANCA-Associated Glomerulonephritis. *J Am Soc Nephrol*. 2010;21(10):1628-1636. doi:10.1681/ASN.2010050477
12. Chang D-y, Wu L-h, Liu G, Chen M, Kallenberg CGM, Zhao M-H. Re-evaluation of the histopathologic classification of ANCA-associated glomerulonephritis: a study of 121 patients in a single center. *Nephrol Dial Transplantation*. 2012;27(6):2343-2349. doi:10.1093/ndt/gfr643
13. Hilhorst M, Wilde B, Vriesman PvB, van Paassen P, Tervaert JWC. Estimating Renal Survival Using the ANCA-Associated GN Classification. *J Am Soc Nephrol*. 2013;24(9):1371-1375. doi:10.1681/ASN.2012090912
14. Bjørneklett R, Sriskandarajah S, Bostad L. Prognostic Value of Histologic Classification of ANCA-Associated Glomerulonephritis. *Clin J Am Soc Nephrol*. 2016;11(12):2159-2167. doi:10.2215/CJN.04800516
15. Moroni G, Binda V, Leoni A, et al. Predictors of renal survival in ANCA-associated vasculitis. Validation of a histopatological classification schema and review of the literature. *Clin Exp Rheumatol*. 2015;33(2 Suppl 89):56-63.
16. Andreiana I, Stancu S, Avram A, Taran L, Mircescu G. ANCA positive crescentic glomerulonephritis outcome in a Central East European cohort: a retrospective study. *BMC Nephrol*. 2015;16(1):90-90. doi:10.1186/s12882-015-0091-8
17. Tanna A, Guarino L, Tam FWK, et al. Long-term outcome of anti-neutrophil cytoplasm antibody-associated glomerulonephritis: evaluation of the international histological classification and other prognostic factors. *Nephrol Dial Transplantation*. 2015;30(7):1185-1192. doi:10.1093/ndt/gfu237

18. Córdova-Sánchez BM, Mejía-Vilet JM, Morales-Buenrostro LE, Loyola-Rodríguez G, Uribe-Uribe NO, Correa-Rotter R. Clinical presentation and outcome prediction of clinical, serological, and histopathological classification schemes in ANCA-associated vasculitis with renal involvement. *Clin Rheumatol*. 2016;35(7):1805-1816. doi:10.1007/s10067-016-3195-z
19. Chen Y-X, Xu J, Pan X-X, et al. Histopathological Classification and Renal Outcome in Patients with Antineutrophil Cytoplasmic Antibodies-associated Renal Vasculitis: A Study of 186 Patients and Metaanalysis. *J Rheumatol*. 2017;44(3):304-313. doi:10.3899/jrheum.160866
20. van Daalen EE, Trejo MAW, Göçeroğlu A, et al. Developments in the histopathological classification of ANCA-associated glomerulonephritis. *Clin J Am Soc Nephrol*. 2020;15(8):1103-1111. doi:0.2215/CJN.14561119
21. Sethi S, D'Agati VD, Nast CC, et al. A proposal for standardized grading of chronic changes in native kidney biopsy specimens. *Kidney Int*. 2017;91(4):787-789. doi:10.1016/j.kint.2017.01.002
22. Casal Moura M, Fervenza FC, Specks U, Sethi S. Kidney biopsy chronicity grading in antineutrophil cytoplasmic antibody-associated vasculitis. *Nephrol Dial Transplantation*. 2021;doi:10.1093/ndt/gfab250
23. Brix SR, Noriega M, Tennstedt P, et al. Development and validation of a renal risk score in ANCA-associated glomerulonephritis. *Kidney Int*. 2018;94(6):1177-1188. doi:10.1016/j.kint.2018.07.020
24. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612. doi:10.7326/0003-4819-150-9-200905050-00006
25. An X-N, Wei Z-N, Yao X-Y, et al. Evaluating renal outcome of ANCA-associated renal vasculitis: comparative study of two histopathological scoring systems. *Clin Exp Rheumatol*. 2021;39(129):S39-S45.
26. Gercik O, Bilgin E, Solmaz D, et al. Histopathological subgrouping versus renal risk score for the prediction of end-stage renal disease in ANCA-associated vasculitis. *Ann Rheum Dis*. 2020;79(5):675-676. doi:10.1136/annrheumdis-2019-216742
27. Villacorta J, Diaz-Crespo F, Guerrero C, Acevedo M, Cavero T, Fernandez-Juarez G. Long-term validation of the renal risk score for vasculitis in a Southern European population. *Clin Kidney J*. 2021;14(1):220-225. doi:10.1093/ckj/sfaa073
28. Brilland B, Boud'hors C, Copin M-C, et al. Assessment of Renal Risk Score and Histopathological Classification for Prediction of End-Stage Kidney Disease and Factors Associated With Change in eGFR After ANCA-Glomerulonephritis Diagnosis. *Front Immunol*. 2022;13:834878-834878. doi:10.3389/fimmu.2022.834878
29. Kant S, Costigliolo F, Brix SR, Fenaroli P, Rosenberg A, Geetha D. Application of the ANCA Renal Risk Score in the United States: A Single-Center Experience. *Kidney Med*. 2021;3(4):686-688. doi:10.1016/j.xkme.2021.04.005
30. Lim J-H, Han M-H, Kim Y-J, et al. Histopathologic and clinicopathologic classifications of antineutrophil cytoplasmic antibody-associated glomerulonephritis: A validation study in a korean cohort. *Kidney Res Clin Practice*. 2021;41(1):77-88. doi:10.23876/j.krcp.20.184
31. Saito M, Saito A, Abe F, et al. Evaluation of a newly proposed renal risk score for Japanese patients with ANCA-associated glomerulonephritis. *Clin Exp Nephrol*. 2022;26(8):760-769. doi:10.1007/s10157-022-02217-w
32. You X, Zhang J, Ding X, Zhang J, Zhou Q, Lu G. Predictors of renal outcomes in crescentic and mixed class of ANCA-associated glomerulonephritis. *Clin Nephrol*. 2021;95(2):81.
33. Mejía-Vilet JM, Martín-Nares E, Cano-Verduzco ML, Pérez-Arias AA, Sedano-Montoya MA, Hinojosa-Azaola A. Validation of a renal risk score in a cohort of ANCA-associated vasculitis patients with severe kidney damage. *Clin Rheumatol*. 2020;39(6):1935-1943. doi:10.1007/s10067-020-04936-5
34. Bai X, Guo Q, Lou Y, et al. Validation of the renal risk score for antineutrophil cytoplasmic antibody-associated glomerulonephritis in a Chinese population. *Clin Rheumatol*. 2021;40(12):5009-5017. doi:10.1007/s10067-021-05862-w

35. Tan PG, O'Brien J, Pusey CD, McAdoo SP. Validation of the ANCA renal risk score in a London cohort: potential impact of treatment on prediction outcome. *Kidney Int*. 2021;99(2):488-489. doi:10.1016/j.kint.2020.04.061
36. Wester Trejo MAC, van Daalen EE, Berden AE, et al. A renal risk score for ANCA-associated glomerulonephritis. *Kidney Int*. 2019;96(1):245-245. doi:10.1016/j.kint.2019.01.046
37. Li AS, Saleh C, Denley H, Patel M, Brix SR. ANCA renal risk score predicts outcome in the Manchester cohort. *Kidney Int*. 2019;96(1):246-247. doi:10.1016/j.kint.2019.03.022
38. Xia M, Yu R, Zheng Z, et al. Meta-Analytical Accuracy of ANCA Renal Risk Score for Prediction of Renal Outcome in Patients With ANCA-Associated Glomerulonephritis. *Front Med*. 2021;8doi:10.3389/fmed.2021.736754
39. McGovern DP, Lees JS, Traynor JP, et al. Outcomes in ANCA-Associated Vasculitis in Scotland: Validation of the Renal Risk Score in a Complete National Cohort. *Kidney Int Rep*. 2023;doi:10.1016/j.ekir.2023.05.029
40. Nezam D, Porcher R, Grolleau F, et al. Kidney histopathology can predict kidney function in ANCA-associated vasculitides with acute kidney injury treated with plasma exchanges. *J Am Soc Nephrol*. 2022;33(3):628-637. doi:10.1681/ASN.2021060771
41. Walsh M, Merkel PA, Peh C-A, et al. Plasma Exchange and Glucocorticoids in Severe ANCA-Associated Vasculitis. *New Engl J Med*. 2020;382(7):622-631. doi:10.1056/NEJMoa1803537
42. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei L-J. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105-1117. doi:10.1002/sim.4154
43. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7
44. Ramspek CL, de Jong Y, Dekker FW, van Diepen M. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrol Dial Transplantation*. 2020;35(9):1527-1538. doi:10.1093/ndt/gfz018
45. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *Br Med J*. 2020;369doi:10.1136/bmj.m1328
46. Dhiman P, Ma J, Andaur Navarro CL, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. *BMC Med Res Methodol*. 2022;22(1):1-16. doi:10.1186/s12874-022-01577-x
47. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13(1):1-15. doi:10.1186/1471-2288-13-33
48. Van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401-3415. doi:10.1002/1097-0258(20001230)19:24<3401::aid-sim554>3.0.co;2-2
49. Jebali H, Khadhar M, Mami I, et al. Predictors of renal outcomes in anti-neutrophil cytoplasmic antibody glomerulonephritis. *Saudi J Kidney Dis Transplantation*. 2020;31(1):182. doi:10.4103/1319-2442.279939
50. Boudhabhay I, Delestre F, Coutance G, et al. Reappraisal of renal arteritis in ANCA-associated vasculitis: Clinical characteristics, pathology, and outcome. *J Am Soc Nephrol*. 2021;32(9):2362-2374. doi:10.1681/ASN.2020071074
51. van Royen FS, Moons KGM, Geersing G-J, van Smeden M. Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eu Respir J*. 2022:2200250. doi:10.1183/13993003.00250-2022
52. Haas M, Seshan SV, Barisoni L, et al. Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the Renal Pathology Society. *Kidney Int*. 2020;98(5):1120-1134. doi:10.1016/j.kint.2020.08.006
53. McLernon DJ, Giardiello D, Van Calster B, et al. Assessing performance and clinical usefulness in prediction models with survival outcomes: Practical guidance for Cox proportional hazards models. *Ann Intern Med*. 2022;doi:10.7326/M22-0844

54. *Package 'rpart'*. Version 4.1.16. 2022. https://cran.r-project.org/web/packages/rpart/rpart.pdf
55. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381-5397. doi:10.1002/sim.5958
56. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3):515-526. doi:10.1093/biomet/81.3.515
57. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
58. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Br Surg*. 2015;102(3):148-158. doi:10.1002/bjs.9736
59. 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 Internal Med*. 2015;162(1):W1-W73. doi:10.7326/M14-0698
60. Shafi T, Zhu X, Lirette ST, et al. Quantifying Individual-Level Inaccuracy in Glomerular Filtration Rate Estimation: A Cross-Sectional Study. *Ann Intern Med*. 2022;doi:10.7326/M22-0610

TABLES

Table 1 | Baseline demographic factors

	All Patients (N = 1439)	Development Cohort (N = 959)	Validation Cohort (N = 480)
Age at diagnosis	64.0 (53.0-73.0)	63.4 (52.3-72.5)	64.5 (53.9-73.2)
Male sex	750 (52.1%)	484 (50.5%)	266 (55.4%)
Diagnosis			
GPA	568 (39.5%)	388 (40.5%)	180 (37.5%)
MPA	820 (57.0%)	538 (56.1%)	282 (58.8%)
EGPA	19 (1.3%)	14 (1.5%)	5 (1.0%)
ANCA negative PING	31 (2.2%)	19 (2.0%)	12 (2.5%)
ANCA type			
Myeloperoxidase	762 (53.0%)	509 (53.1%)	253 (52.7%)
Proteinase 3	579 (40.2%)	389 (40.6%)	190 (39.6%)
ANCA negative	98 (6.8%)	61 (6.4%)	37 (7.7%)
eGFR (ml/min/1.73m^2)	21.0 (10.9-36.4)	21.8 (11.3-38.0)	18.0 (9.8-35.0)
Creatinine (µmol/l)	234.0 (150.3-402.0)	225.0 (147.0-383.5)	257.0 (157.5-449.5)
KRT at presentation	291 (20.2%)	185 (19.3%)	106 (22.1%)
Glomeruli on biopsy	17.0 (12.0-23.0)	17.0 (12.0-23.0)	17.0 (12.0-23.2)
Percentage Normal Glomeruli	35.5 (14.3-65.2)	36.0 (14.3-66.7)	34.7 (14.3-62.5)
Percentage Crescentic Glomeruli	20.0 (5.9-46.4)	19.0 (5.8-44.4)	21.4 (6.3-50.0)
IFTA			
None – Mild	689 (47.9%)	469 (48.9%)	220 (45.8%)
Mild to Moderate - Severe	750 (52.1%)	490 (51.1%)	260 (54.2%)
ANCA Renal Risk Score	3.0 (0.0-7.0)	3.0 (0.0-7.0)	4.0 (2.0-7.0)
Risk Group			
Low	370 (25.7%)	258 (26.9%)	112 (23.3%)
Moderate	742 (51.6%)	484 (50.5%)	258 (53.8%)
High	327 (22.7%)	217 (22.6%)	110 (22.9%)

ANCA, Anti-neutrophil cytoplasmic antibody; eGFR, estimated glomerular filtration rate; EGPA, eosinophilic granulomatosis with polyangiitis; GPA, granulomatosis with polyangiitis; IFTA, interstitial fibrosis and tubular atrophy; KRT, kidney replacement therapy; MPA, microscopic polyangiitis; PING, Pauci-immune necrotizing glomerulonephritis.

Table 2 | Clinical outcomes

	All Patients (N = 1439)	Development Cohort (N=959)	Validation Cohort (N=480)
Follow-up (years)	3.6 (1.1-5.9)	3.6 (1.1-5.9)	3.6 (1.0-6.0)
Kidney recovery	136/291 (46.7%)	83/185 (44.9%)	53/106 (50%)
ESKD	325 (22.6%)	207 (21.6%)	118 (24.6%)
Death	315 (21.9%)	206 (21.5%)	109 (22.7%)
Death without ESKD	199 (13.8%)	135 (14.1%)	64 (13.3%)

ESKD, end stage kidney disease

Table 3a | Model parameters and performance

	N=456	**Original Model**	N=959	**Simplified Original Model**	**New AKRiS Model**		
Variable	n (%)	Points	n (%)	Points	β	HR	Points
eGFR (ml/min/1.73m^2)							
G0: ≥ 15	279 (61.2)	0	608 (63.4)	0			
G1: < 15	177 (38.8)	3	351 (36.6)	3			
Creatinine (µmol/l)							
K0: < 250			534 (55.7)		Ref		0
K1: 250 – 450			241 (25.1)		0.661	1.94 (1.30-2.89)	4
K2: > 450			184 (19.2)		1.886	6.59 (4.59-9.45)	11
Normal Glomeruli (%)							
N0: > 25	245 (53.7)	0	590 (61.5)	0	Ref		0
N1: 10 – 25	90 (19.7)	4	177 (18.5)	4	0.650	1.92 (1.30-2.82)	4
N2: < 10	121 (26.5)	6	192 (20.0)	6	1.199	3.32 (2.35-4.69)	7
IFTA (%)							
T0: < 25%	216 (47.4)	0					
T1: ≥ 25%	240 (52.6)	2					
IFTA (simplified)							
T0: none, mild, < 25%			469 (48.9)	0	Ref		0
T1: ≥ mild-moderate, ≥ 25%			490 (51.1)	2	0.527	1.69 (1.28-2.26)	3
C Statistic (95%CI)	**0.779 (0.739 – 0.819)**		**0.800 (0.768-0.832)**		**0.831 (0.801 – 0.861)**		

AKRiS, ANCA Kidney Risk Score; C, Harrell's concordance; eGFR, estimated glomerular filtration rate; IFTA, interstitial fibrosis and tubular atrophy.

Table 3b | Risk groups for the AKRiS

Group	Points
Low	0 – 4
Moderate	5 – 11
High	12 – 18
Very high	21
Risk points 1, 2, 19 and 20 not available	

Table 3c | Parameters for the continuous AKRiS

		β
Creatinine	Log (µmol/l)	1.250
Normal Glomeruli	Per 1%	-0.0167
IFTA	≥ mild-moderate or ≥ 25%	0.616
	C Statistic (95%CI)	**0.833 (0.805 – 0.861)**

Survival function in Supplemental Equation 2
IFTA, interstitial fibrosis and tubular atrophy; log, natural logarithm.

FIGURE LEGEND

Figure 1 | Kidney survival according to the stratification of the Original ANCA Renal Risk Score (ARRS). Kaplan-Meier curve depicting the development of end-stage kidney disease (ESKD) of patients with anti-neutrophil cytoplasmic antibody glomerulonephritis (ANCA GN). Patients of the development cohort are assigned points as per the ARRS according to the estimated glomerular filtration rate at presentation (eGFR, G0: >15 ml/min/1.73 m^2, G1: ≤15 ml/min/1.73 m^2), the percentage of normal glomeruli in the kidney biopsy (N0: >25%, N1: 10-25%, N2: <10%) and the degree of interstitial fibrosis and tubular atrophy (T0: <25%, T1: ≥25%). Points are calculated (G0=0, G1=3, N0=0, N1=4, N2=6, T0=0, T1=2) and risk groups created – low (0), moderate (2 - 7), and high-risk group (8 – 11). The number of patients at risk in each group at each time point is stated below the graph. Patient outcomes differ per risk group, *C*=0.779, P<0.001.

Figure 2 | Kidney survival according to the stratification of the New ANCA Kidney Risk Score (AKRiS). Kaplan-Meier curve depicting the development of end-stage kidney disease (ESKD) of patients with anti-neutrophil cytoplasmic antibody glomerulonephritis (ANCA GN) of the development (solid line) and validation cohort (dashed line). Patients are assigned points as per the AKRiS according to the serum creatinine (K0: <250 µmol/l, K1: 250-450 µmol/l, K2: >450 µmol/l), the percentage of normal glomeruli in the kidney biopsy (N0: >25%, N1: 10-25%, N2: <10%) and the degree for interstitial fibrosis and tubular atrophy (T0: none/mild or <25%, T1: ≥ mild-moderate or ≥25%). Points are calculated (K0=0, K1=4, K2=11, N0=0, N1=4, N2=7, T0=0, T1=3) and risk groups created – low (0-4), moderate (5-11), high (12-18) and very high (21). The number of patients at risk in each group at each time point is stated below the graph. Patient outcomes differ per risk group, *C*=0.831, P<0.001.