

Table of Contents

Supplemental Table 1

Discrimination and calibration of all pre-selected characteristics' (Full) model and parsimonious models

Supplemental Table 2

Comparisons of patients with diabetes and/or hypertension by physician recommendation for biopsy

Supplemental Figure 1

Distribution of diagnoses in patients without CKDu on biopsy

Supplemental Figure 2

Ranking of the relative strength of the nine pre-selected characteristics used in the Full model

Supplemental Figure 3

Calibration plots of the Full model and five parsimonious models

Supplemental Appendix 1

Definition of CKDu histological scoring

Supplemental Appendix 2

Methods for prediction modeling

Supplemental Table 1. Discrimination and calibration of all pre-selected characteristics' (Full) model and parsimonious models

	Full model	A	B	C	D	E
Predictors						
Urine dipstick trace or negative for protein	•	•	•	•	•	•
Age (cubic spline with 3 knots)	•	•	•	•	•	•
Hypertension	•		•			
Diabetes	•		•			
Serum albumin	•				•	•
Hematuria	•					
No pyuria	•			•		
Normal potassium	•					
Woman	•					
Urine dipstick negative x Serum albumin	•					•
On original data						
C-statistic	0.89	0.86	0.87	0.87	0.87	0.87
Sensitivity	83.72	76.74	76.74	74.91	76.74	79.07
Specificity	79.55	84.09	84.09	84.09	84.09	84.09
Positive Predictive Value	80.00	82.50	82.50	82.05	82.50	82.93
Negative Predictive Value	83.33	78.72	78.72	77.08	78.72	80.43
Bootstrap Validation						
C-statistic	0.82	0.85	0.84	0.85	0.85	0.84
Calibration slope	0.59	0.91	0.82	0.87	0.88	0.84

Each model was fit using logistic regression that included all the variables with a • in the model's column.

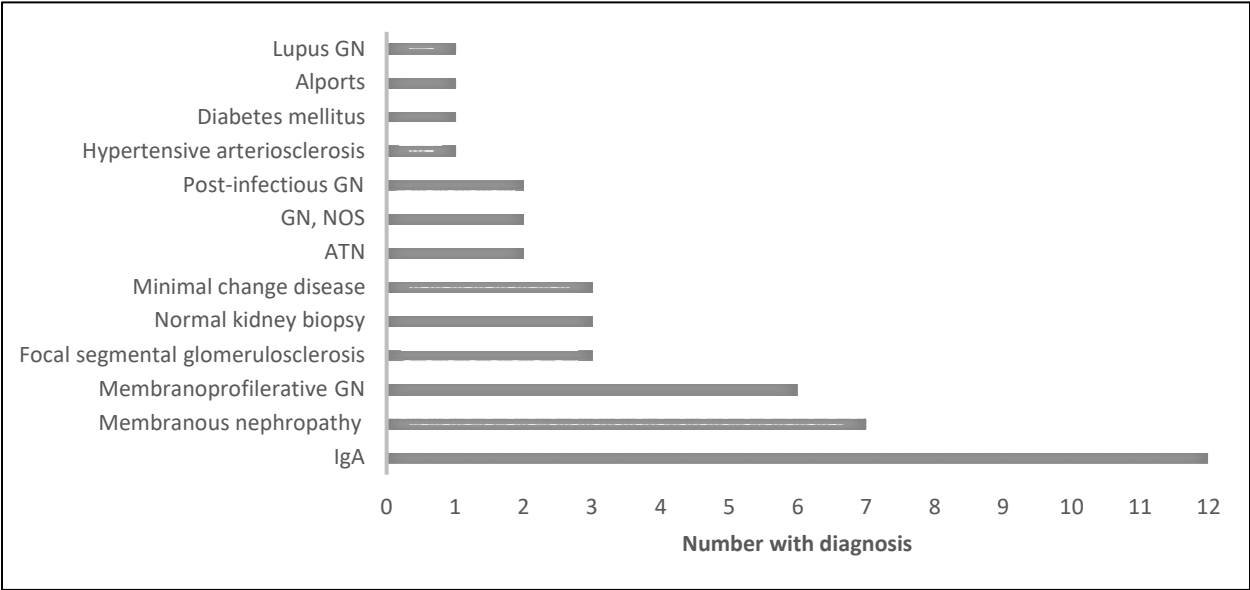
Supplemental Table 2a. Comparisons of patients with diabetes by physician recommendation for biopsy

	Biopsy recommended (n=20)	Biopsy not recommended (n=157)
Age	49.8 (11.2)	59.5 (10.6)
Women	5 (25.0)	50 (31.8)
Diabetes by self report	18 (90.0)	149 (94.9)
Diabetes by physician diagnosis	17 (85.0)	114 (72.6)
Hypertension	9 (45.0)	124 (79.0)
Albumin < 3.5 g/dL	7 (35.0)	9 (5.7)
Missing	4 (20.0)	53 (33.8)
Potassium < 3.5	1 (5.0)	6 (3.8)
Missing	2 (10.0)	22 (14.0)
Hematuria	2 (10.0)	9 (5.7)
Pyuria	3 (15.0)	35 (22.3)
Urine Dipstick negative for protein	10 (50.0)	66 (42.0)

Supplemental Table 2b. Comparisons of patients with hypertension by recommendation for biopsy

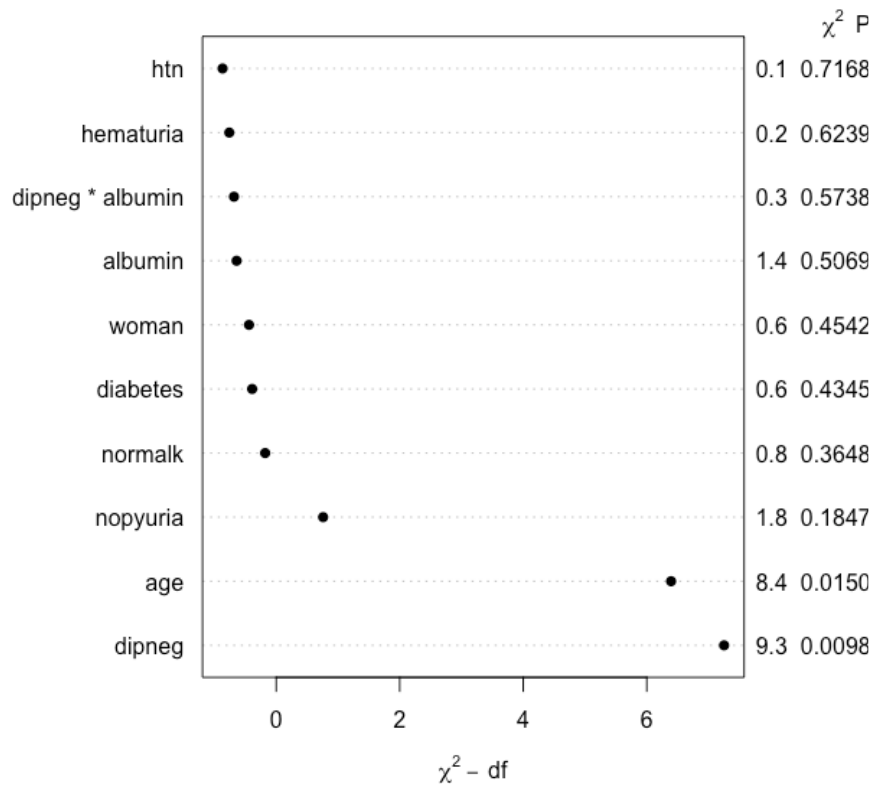
	Biopsy recommended (n=54)	Biopsy not recommended (n=300)
Age	47.4 (11.7)	59.8 (12.1)
Women	15 (27.8)	87 (29.0)
Hypertension by self report	40 (74.1)	217 (72.3)
Hypertension by physician diagnosis	27 (50.0)	149 (49.7)
Diabetes	9 (16.7)	124 (41.3)
Albumin < 3.5 g/dL	16 (29.6)	12 (4.0)
Missing	9 (16.7)	121 (40.3)
Potassium < 3.5	1 (5.0)	6 (3.8)
Missing	6 (11.1)	38 (12.7)
Hematuria	12 (22.2)	26 (8.7)
Pyuria	24 (44.4)	78 (26.0)
Urine dipstick negative for protein	26 (48.1)	156 (52.0)

Supplemental Figure 1. Distribution of diagnoses in patients without CKDu on biopsy



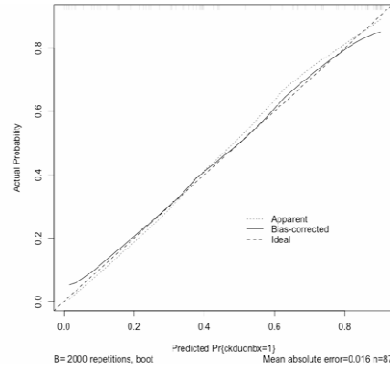
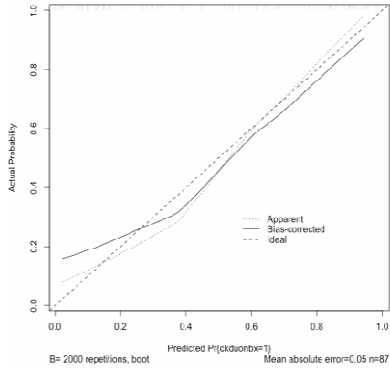
*Minimal change disease was diagnosed based minimal glomerular pathology and clinical history; Alport’s syndrome diagnoses had supportive family history.

Supplemental Figure 2. Ranking of the relative strength of the nine pre-selected characteristics used in the Full model

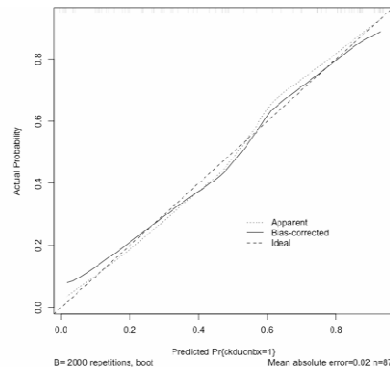
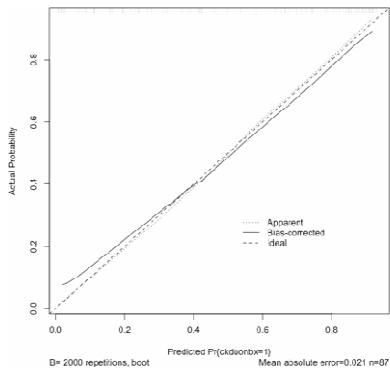


Supplemental Figure 3. Calibration plots of the Full model and five parsimonious models

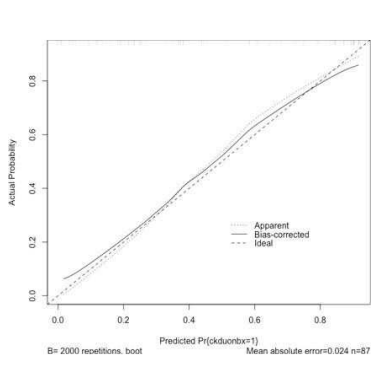
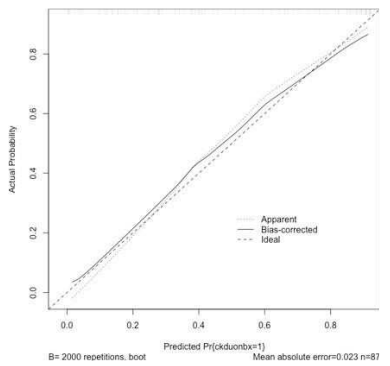
Full Model (validated calibration slope = 0.55) Model A (validated calibration slope = 0.91)



Model B (validated calibration slope = 0.82) Model C (validated calibration slope = 0.87)



Model D (validated calibration slope = 0.88) Model E (validated calibration slope = 0.84)



Supplemental Annex 1: Definition of CKDu histological scoring

	Definition	Grading(score)			
		0	1	2	3
C H R O N I C T Y I N D E X	1. Global glomerulosclerosis score (GS) = number of sclerosed glomeruli/total glomeruli (TG)x 100%	No GS	GS 0-25%	GS 26-50%	GS >50%
	2. Peri-Glo fibrosis (PGF) score =number of PGF/total glomeruli x 100	No PGF	PGF 0-25%	PGF 26-50%	PGF >50%
	3.Tubular Atrophy (TA)	No tubular atrophy	Tubular atrophy in up to 25% of the area of cortical tubules	Tubular atrophy involving 26 to 50% of the area of cortical tubules	Tubular atrophy of >50% of the area of cortical tubules
	4. Interstitial Fibrosis (IF)	No Interstitial fibrosis	Interstitial fibrosis tissue in 0 to 25% of cortical area	Interstitial fibrosis of 26 to 50% of cortical area	Interstitial fibrosis of >50% of cortical area
	5. Vascular Fibrous Intimal Thickening (CV)	No chronic vascular changes	Vascular narrowing of up to 25% luminal area by fibro-intimal thickening of arteries \pm breach of internal elastic lamina or presence of foam cells or occasional mononuclear cells*	Increased severity of changes described score 1 with 26 to 50% narrowing of vascular luminal area*	Severe vascular changes with >50% narrowing of vascular luminal area*
A C	6. Tubulitis (Tu)	No mononuclear cells in tubules	Foci with 1 to 4 cells/tubular cross section or 10 tubular cells	Foci with 5 to 10 cells/tubular cross section	

T I V I T Y I N D E X	7. Interstitial Inflammation (II)	No interstitial inflammation	0 to 25% of parenchyma inflamed cells	26 to 50% of parenchyma inflamed	>50% of parenchyma inflamed Indicate presence of remarkable numbers (>10% of total cells) of eosinophils, polys, or plasma cells (specify which)
O T H E R	8. Tubular BM damage				
	9. Tubular epithelia cell damage				
	10. Interstitial Eosinophils				
	11. Protein cast				
	12. Granular cast				
	13. RBC cast				

Supplemental Annex 2: Methods for prediction modeling

We used bootstrap validation, over 2000 bootstrap samples, to calculate the model validated c-statistic and calibration slope[1]. For missing serum potassium and albumin values, we imputed using fully conditioning specification using all variables in the full model in the imputation plus outcome and with calcium, sodium, weight, and systolic blood pressure as auxiliary variables. Imputations were performed stratified by dipstick proteinuria, which was not missing, to conform to the analysis model that included the interaction between urine protein dipstick and low serum albumin[2].

1. Harrell, F., *Regression modeling strategies: With application to linear models, logistic and ordinal regression, and survival analysis*. 2nd ed. Springer Science and Media. 2001, New York: Springer.
2. van Buuren, S., *Flexible imputation of missing data*. Chapman&Hall/CRC Interdisciplinary Statistics. 2012, Boca Raton, FL: CRC Press.

Supplemental Annex 2: Methods for prediction modeling

Each model takes the form

$$\text{Prob}\{\text{CKDu} = 1\} = \frac{1}{1 + \exp(-\mathbf{X}\hat{\boldsymbol{\beta}})}, \text{ where}$$

Full Model $\mathbf{X}\hat{\boldsymbol{\beta}} =$

$$\begin{aligned} & -8.278552 \\ & +0.1618171\text{age} - 5.297437 \times 10^{-5}(\text{age} - 23)_+^3 + 9.491241 \times 10^{-5}(\text{age} - 42)_+^3 + \\ & -4.193804 \times 10^{-5}(\text{age} - 66)_+^3 - 0.2322563 \text{ htn} + 0.7005494 \text{ diabetes} \\ & -0.4188918 \text{ hematuria} - 0.9505295 \text{ nopyuria} - 0.8154129 \text{ normalk} \\ & -0.6134991 \text{ woman} + 3.875809 \text{ dipneg} + 0.5699782 \text{ albumin} \\ & -0.4228312 \text{ dipneg} \times \text{albumin} \end{aligned}$$

Model A $\mathbf{X}\hat{\boldsymbol{\beta}} =$

$$\begin{aligned} & -8.243477 \\ & +0.1786645\text{age} - 6.920365 \times 10^{-5}(\text{age} - 23)_+^3 + 0.0001239899(\text{age} - 42)_+^3 + \\ & -5.478623 \times 10^{-5}(\text{age} - 66)_+^3 + 2.47526 \text{ dipneg} \end{aligned}$$

Model B $\mathbf{X}\hat{\boldsymbol{\beta}} =$

$$\begin{aligned} & -7.772454 \\ & +0.1666065\text{age} - 5.989618 \times 10^{-5}(\text{age} - 23)_+^3 + 0.000107314(\text{age} - 42)_+^3 + \\ & -4.741781 \times 10^{-5}(\text{age} - 66)_+^3 - 0.4147905 \text{ htn} + 0.4918624 \text{ diabetes} \\ & +2.478244 \text{ dipneg} \end{aligned}$$

Model C $\mathbf{X}\hat{\boldsymbol{\beta}} =$

$$\begin{aligned} & -8.55315 \\ & +0.1954188\text{age} - 8.022421 \times 10^{-5}(\text{age} - 23)_+^3 + 0.000143735(\text{age} - 42)_+^3 + \\ & -6.351083 \times 10^{-5}(\text{age} - 66)_+^3 - 0.6664385 \text{ nopyuria} + 2.644806 \text{ dipneg} \end{aligned}$$

Model D $\mathbf{X}\hat{\boldsymbol{\beta}} =$

$$\begin{aligned} & -8.833361 \\ & +0.1658592\text{age} - 6.076192 \times 10^{-5}(\text{age} - 23)_+^3 + 0.0001088651(\text{age} - 42)_+^3 + \\ & -4.810318 \times 10^{-5}(\text{age} - 66)_+^3 + 0.324278 \text{ albumin} + 2.141375 \text{ dipneg} \end{aligned}$$

Model E $\mathbf{X}\hat{\boldsymbol{\beta}}$ =

-10.19217

+0.1799686age - 6.65842 $\times 10^{-5}$ (age - 23) $^3_{+}$ + 0.0001192967(age - 42) $^3_{+}$

-5.271249 $\times 10^{-5}$ (age - 66) $^3_{+}$ + 3.813247 dipneg + 0.5662221 albumin

-0.4410596 dipneg \times albumin

$(\mathbf{x})_{+} = \mathbf{x}$ if $\mathbf{x} > 0$, 0 otherwise; hematuria=1 if urine red blood cells > 3 per high power field, 0 otherwise; nopyuria=1 if urine white blood cells > 3 per high power field, 0 otherwise; normalk =1 if serum potassium < 3.5 meq/L, 0 otherwise; dipneg=1 if urine dipstick protein trace or nil, 0 otherwise.