A Predictive Model for Progression of Chronic Kidney Disease to Kidney Failure

Navdeep Tangri, MD, FRCPC

Lesley A. Stevens, MD, MS, FRCPC

John Griffith, PhD

Hocine Tighiouart, MS

Ognjenka Djurdjev, MSc

David Naimark, MD, FRCPC

Adeera Levin, MD, FRCPC

Andrew S. Levey, MD

AN ESTIMATED 23 MILLION people in the United States (11.5% of the adult population) have chronic kidney disease (CKD) and are at increased risk for cardiovascular events and progression to kidney failure.[1-5] Similar estimates of burden of disease have been reported around the world.[6] Although there are proven therapies to improve outcomes in patients with progressive kidney disease, these therapies may also cause harm and add cost. Clinical decision making for CKD is challenging due to the heterogeneity of kidney diseases, variability in rates of disease progression, and the competing risk of cardiovascular mortality.[7,8] Accurate prediction of risk could facilitate individualized decision making, enabling early and appropriate patient care.[9,10]

Currently, there are no widely accepted predictive instruments for CKD progression; therefore, physicians must make ad hoc decisions about which patients to treat, risking delays in treatment in those who ultimately progress to kidney failure, or unnecessary treatment in those who do not progress. The severity of CKD has been recommended to guide treatment-related de-

See related articles.

Context Chronic kidney disease (CKD) is common. Kidney disease severity can be classified by estimated glomerular filtration rate (GFR) and albuminuria, but more accurate information regarding risk for progression to kidney failure is required for clinical decisions about testing, treatment, and referral.

Objective To develop and validate predictive models for progression of CKD.

Design, Setting, and Participants Development and validation of prediction models using demographic, clinical, and laboratory data from 2 independent Canadian cohorts of patients with CKD stages 3 to 5 (estimated GFR, 10-59 mL/min/1.73 m^2) who were referred to nephrologists between April 1, 2001, and December 31, 2008. Models were developed using Cox proportional hazards regression methods and evaluated using C statistics and integrated discrimination improvement for discrimination, calibration plots and Akaike Information Criterion for goodness of fit, and net reclassification improvement (NRI) at 1, 3, and 5 years.

Main Outcome Measure Kidney failure, defined as need for dialysis or preemptive kidney transplantation.

Results The development and validation cohorts included 3449 patients (386 with kidney failure [11%]) and 4942 patients (1177 with kidney failure [24%]), respectively. The most accurate model included age, sex, estimated GFR, albuminuria, serum calcium, serum phosphate, serum bicarbonate, and serum albumin (C statistic, 0.917; 95% confidence interval [CI], 0.901-0.933 in the development cohort and 0.841; 95% CI, 0.825-0.857 in the validation cohort). In the validation cohort, this model was more accurate than a simpler model that included age, sex, estimated GFR, and albuminuria (integrated discrimination improvement, 3.2%; 95% CI, 2.4%-4.2%; calibration [Nam and D'Agostino χ^2 statistic, 19 vs 32]; and reclassification for CKD stage 3 [NRI, 8.0%; 95% CI, 2.1%-13.9%] and for CKD stage 4 [NRI, 4.1%; 95% CI, −0.5% to 8.8%]).

Conclusion A model using routinely obtained laboratory tests can accurately predict progression to kidney failure in patients with CKD stages 3 to 5.

cisions.[11] Severity is commonly staged according to the level of glomerular filtration rate (GFR) estimated from serum creatinine. Reporting estimated GFR when serum creatinine is measured has increased awareness of CKD and referrals to nephrologists, but estimated GFR is not sufficient for clinical decision making.[12,13]

Recent studies have shown that albuminuria provides additional prognostic information for progression to kidney failure.[14] Some studies have examined the use of estimated GFR and albuminuria in prediction models, with additional

clinical and laboratory data, but these models are either specific to a particular type of kidney disease or not externally validated.[7,15-18] The ideal model to

Author Affiliations: Department of Medicine (Drs Tangri, Stevens, and Levey) and Biostatistics Research Center, Tufts Clinical and Translational Science Institute (Dr Griffith and Mr Tighiouart), Tufts Medical Center, Boston, Massachusetts; Department of Medicine, University of British Columbia, and British Columbia Provincial Renal Agency, Vancouver, British Columbia, Canada (Dr Levin and Ms Djurdjev); and Department of Medicine, Sunnybrook Hospital, University of Toronto, Toronto, Ontario, Canada (Dr Naimark).
Corresponding Author: Navdeep Tangri, MD, FRCPC, Department of Medicine, Tufts Medical Center, 800 Washington St, PO Box 391, Boston, MA 02111 (ntangri@tuftsmedicalcenter.org).

predict progression would be accurate, easy to implement, and highly generalizable across a spectrum of patients with CKD in independent populations.

Using data from 2 separate CKD cohorts, the goal of our study was to develop and externally validate an accurate but simple prediction model for progression of CKD. The goal was also to use variables routinely measured in patients with CKD to create a model to predict progression to kidney failure that could be easily implemented in clinical practice. We were especially interested in models that rely solely on information available to a clinical laboratory, enabling reporting the risk of kidney failure with laboratory test results.

METHODS
Study Population
Development Cohort. The development cohort was derived from the nephrology clinic electronic health record (EHR) at Sunnybrook Hospital, a part of the University of Toronto Health Network, Toronto, Ontario, Canada. Patients with CKD stages 3 to 5 (estimated GFR, <60 mL/min/1.73 m^2) at the time of initial nephrology referral were included and were followed up between April 1, 2001, and December 31, 2008 (eFigure 1, available at http://www.jama.com). Outcomes were ascertained by reviewing clinic records as well as through a matching algorithm with the Toronto Regional Dialysis Registry.

Validation Cohort. The validation cohort was derived from the British Columbia CKD Registry (Patient Registration and Outcome Management Information System), which captures clinical and laboratory data on all patients referred to nephrologists in the province. Patients with CKD stages 3 to 5 at the time of initial nephrology referral between January 1, 2001, and December 31, 2009, were included. Outcomes such as dialysis, death, and transplantation are all captured in the database, which matches all kidney failure outcomes with provincial and national registry.[17]

The study was reviewed and approved by the institutional review boards at Sunnybrook Hospital, University of Toronto, Toronto, Ontario, Canada; University of British Columbia, Vancouver, British Columbia, Canada; and Tufts Medical Center, Boston, Massachusetts. Informed consent was waived at all participating institutions because of the presence of de-identified data and lack of feasibility of obtaining informed consent from all participants in the cohorts.

Variables
Candidate Dependent Variables. Candidate dependent variables were selected by face validity and included demographic variables, including age and sex; physical examination variables, including blood pressure and weight; comorbid conditions, including diabetes, hypertension, and etiology of kidney disease; and laboratory variables from serum and urine collected at the initial nephrology visit (eTable 1). All predictor variables were obtained at baseline from the nephrology clinic EHR in the development data set. The baseline laboratory value was defined as the first test result within 365 days of the initial estimated GFR. Baseline values were restricted to +/–90 days in a sensitivity analysis. Comorbid conditions were categorized as present or absent at the time of initial nephrology visit.

Using a formula derived from the Irbesartan in Diabetic Nephropathy Trial study,[19] 24-hour urinary protein excretion was transformed to an albumin-to-creatinine ratio. The albumin-to-creatinine ratio was log transformed due to its skewed distribution. The remainder of the laboratory variables and physical examination characteristics were evaluated as linear predictors. The presence of colinearity was examined using a correlation matrix, followed by evaluation of variance inflation factors and magnitude of standard errors. Variables with more than 30% missing values were not included in the analysis. All other missing data were imputed using the multiple imputation technique with 5 imputations based on PROC MI in SAS version 9.1 (SAS Institute Inc, Cary, North Carolina). The complete case was examined as a sensitivity analysis.

Independent Variable. The outcome of interest was kidney failure, which was defined by initiation of dialysis or kidney transplantation and censored for mortality before kidney failure. Although predicting mortality before kidney failure is also important for patients with CKD, predicting the risk of kidney failure alone is important for many decisions made by patients, physicians, and health care systems. The time horizons for risk prediction were 1, 3, and 5 years (eEquation). Risk categories for CKD progression have not been previously defined. Based on input from clinicians, we defined stage-specific CKD risk categories that would be useful for management decisions. In CKD stage 3, risk categories of 0% to 4.9%, 5.0% to 14.9%, and 15.0% or more risk of kidney failure over 5 years were considered as low, intermediate, and high risk, respectively. In CKD stage 4, risk categories of 0% to 9.9%, 10.0% to 19.9%, and 20% or more risk of kidney failure over 2 years were considered as low, intermediate, and high risk, respectively.

Statistical Analysis
Model Development. We developed a sequential series of models and compared those with more variables (ie, greater complexity) to simpler ones. We used a combination of clinical guidance and forward selection to determine variable selection. In univariate Cox proportional hazards regression models, variables not associated with kidney failure ($P > .10$) were excluded from further analyses. Improvement in model performance through addition of new candidate variables in multivariate Cox proportional hazards regression models was tested using metrics for discrimination and goodness of fit. Models 1 through 3 were developed using age and sex, estimated GFR, and albuminuria, successively, and compared with each other. Models 4 through 7 were developed by adding either clinical variables (diabetes and hypertension), physical examination variables (systolic and diastolic blood pressure and weight), laboratory variables of CKD severity (which were as-

sociated with the outcome in multivariate forward selection), or all of the above and compared with model 3.

Prediction Model Validation. Models 2 and 3, the most accurate model among models 4 through 6, and model 7 were evaluated in the external validation data set. The baseline hazard function and β coefficients from the developed model were fixed and applied to the validation data set.

Prediction Model Performance. We used a series of methods to evaluate the performance of the models in the development and the validation data sets. Metrics were assessed in the overall population and in relevant subgroups (age dichotomized at 65 years, sex, CKD stage 3 vs 4, urine albumin-to-creatinine ratio dichotomized at 300 mg/g, and diabetes status).

Discrimination. Discrimination refers to the ability of a model to correctly distinguish between 2 classes of outcomes (kidney failure vs no kidney failure). Concordance statistics (C statistics) and integrated discrimination improvement were computed as measures of discrimination.[20-23]

Calibration. Calibration describes how closely the predicted probabilities agree numerically with the observed outcomes. We compared the observed vs predicted risk of kidney failure for each quintile of predicted risk and determined the magnitude of the deviation using the Nam and D'Agostino χ^2 statistic.[24]

Goodness of Fit. Overall model fit for sequential models was compared using the Akaike Information Criterion (AIC), which takes into account both the statistical goodness of fit and the number of parameters required to achieve this particular degree of fit, by imposing a penalty for increasing the number of parameters.[22]

Reclassification. Reclassification refers to movement of patients from one class to another based on changes to assignment to risk categories. Reclassification improvement was quantified using the net reclassification improvement (NRI) statistic.[21]

Sensitivity Analysis. To evaluate the effect of definition of risk categories on reclassification, we calculated NRI for alternative risk thresholds in CKD stages 3 and 4 and using an alternative method that does not require categories.[25] To evaluate the effect of the competing risk of mortality before kidney failure on risk prediction, we compared the results of our models with competing risk models. The competing risk models goal was to estimate the cumulative incidence function, which is defined by the probability of reaching kidney failure expressed in terms of the cause-specific hazards in the following manner:

$$I_{Kidney\,Failure} = [\sum \lambda_{Kidney\,Failure} \times (t_i) \times S \times (t_{i-1})].$$

The quantities under the summation denote the instantaneous hazard of kidney failure at event time t_i and survival rate from kidney failure or death past event time t_{i-1}.[26]

All statistical analyses were performed using SAS version 9.2. Two-sided $P < .05$ was considered statistically significant.

RESULTS
Cohort Description

The development and validation cohorts included 3449 and 4942 patients, respectively (**TABLE 1** and

Table 1. Baseline Characteristics of Development and Validation Cohorts

Characteristics	No. (%) of Participants		P Value
	Development Cohort (n = 3449)	Validation Cohort (n = 4942)	
Demographics			
Age, mean (SD), y	70 (14)	69 (14)	<.001
≥65	2447 (71)	3292 (67)	NA
<65	1002 (29)	1650 (33)	NA
Male sex	1946 (56)	2833 (57)	.41
Physical examination, mean (SD)			
Systolic BP, mm Hg	130 (22)	138 (24)	<.001
Diastolic BP, mm Hg	71 (12)	74 (13)	<.001
Weight, kg	76 (18)	79 (20)	<.001
Comorbid conditions			
Diabetes	1278 (37)	1907 (38)	.16
Vascular disease[a]	1386 (40)	1305 (26)	<.001
History of current or previous smoking	776 (23)	1149 (23)	.71
Laboratory data			
GFR, mL/min/1.73 m²			
Baseline, mean (SD)	36 (13)	31 (11)	<.001
30-59	2303 (67)	2407 (49)	NA
15-29	926 (27)	2095 (42)	NA
<15	220 (6)	440 (9)	NA
Serum creatinine, mean (SD), mg/dL	2.23 (1.31)	2.30 (0.84)	.02
Hemoglobin, mean (SD), g/dL	12.4 (1.8)	11.8 (1.9)	<.001
Serum calcium, mean (SD), mg/dL	9.36 (0.64)	9.16 (0.70)	<.001
Serum phosphate, mean (SD), mg/dL	3.96 (0.93)	3.93 (0.84)	.30
Serum albumin, mean (SD), g/dL	4.0 (0.5)	3.9 (0.5)	<.001
Serum bicarbonate, mean (SD), mEq/L	26 (4)	25 (4)	<.001
Urine albumin-to-creatinine ratio, mg/g			
Median (IQR)	93 (378)	110 (756)	<.001
<30	814 (24)	973 (20)	NA
30-299	1124 (33)	1915 (39)	NA
≥300	1511 (43)	2054 (41)	NA
Outcome			
Observation time, d	757 (748)	1114 (635)	<.001
Kidney failure events	386 (11)	1177 (24)	<.001
Dialysis	358 (93)	1123 (95)	NA
Transplantation	28 (7)	54 (5)	NA

Abbreviations: BP, blood pressure; GFR, glomerular filtration rate; IQR, interquartile range; NA, not applicable.
SI conversions: To convert serum creatinine to μmol/L, multiply by 88.4; serum calcium to mmol/L, multiply by 0.25; serum phosphate to mmol/L, multiply by 0.323.
[a]Vascular disease is defined as presence of coronary artery disease or peripheral vascular disease.

eFigure 1. Patients in the development and validation cohorts were similar in age and sex and had similar prevalence of diabetes and smoking. Kidney disease severity was greater (lower mean estimated GFR and higher mean albuminuria) and median follow-up was longer in the validation cohort than in the development cohort. The proportion of events was greater in the validation cohort compared with the development cohort (1177 patients with kidney failure [24%] vs 386 patients with kidney failure [11%]).

Prediction Model Performance in the Development Cohort

The hazard ratios for the variables and statistics for discrimination and goodness of fit for successive models in the development data set are shown in TABLE 2. Model 1, including age and sex only, performed poorly (C statistic, 0.561; 95% confidence interval [CI], 0.529-0.593). The C statistic improved with the inclusion of estimated GFR in model 2 (0.892; 95% CI, 0.874-0.910; $P<.001$) and albuminuria in model 3 (0.910; 95% CI, 0.894-

0.926; $P<.001$), did not improve with the addition of diabetes and hypertension in model 4 (0.909; 95% CI, 0.893-0.925; $P=.40$), and did improve with the inclusion of blood pressure and body weight in model 5 (0.915; 95% CI, 0.899-0.931; $P<.001$) and laboratory values in model 6 (0.917; 95% CI, 0.901-0.933; $P<.001$). Despite a similar C statistic, the AIC was lower for model 6 than for model 5 (4432 vs 4463, respectively). The inclusion of all variables in model 7 improved both the C statistic and the AIC compared with model 3 (0.921 [95% CI, 0.905-0.937] vs 0.910 [95% CI, 0.894-0.926] and 4378 vs 4520, respectively). Given these results, models 1, 4, and 5 were not considered in further evaluation steps.

Prediction Model Performance in the Validation Cohort

eTable 2 compares discrimination of models 2, 3, 6, and 7 overall and in subgroups in the development and validation cohorts. In both cohorts, the C statistic was higher for model 6 compared with models 2 and 3 in the entire popu-

lation, and in subgroups ($P<.001$ for all comparisons). In the validation cohort, no further improvement was observed with the additional nonlaboratory variables (0.835; 95% CI, 0.819-0.851 vs 0.841; 95% CI, 0.825-0.857; $P=.90$ for model 7 vs model 6). eTable 3 shows discrimination at years 1, 3, and 5. At all times, both the C statistic and integrated discrimination improvement were greater for model 6 compared with models 2 and 3 ($P<.001$ for all comparisons). Model 6 was more accurate than model 3 (integrated discrimination improvement, 3.2%; 95% CI, 2.4%-4.2%).

The FIGURE shows observed vs predicted probability of kidney failure at 3 years for models 2, 3, and 6 in the validation cohort. The mean absolute difference between the observed and predictive probabilities over quintiles of risk was lower with model 6 compared with models 2 and 3 (1.9% vs 2.7% and 3.1%, respectively), and the Nam and D'Agostino χ^2 statistic also indicated improved fit with model 6 compared with models 2 and 3 (χ^2 statistic, 19 vs 37 and 32, respectively).

Table 2. Hazard Ratios and Goodness of Fit for Sequential Models in the Development Data Set[a]

| Variable	Models						
	1	2	3	4	5	6	7
Baseline GFR, per 5 mL/min/1.73 m²		0.54	0.57	0.58	0.60	0.61	0.64
Age, per 10 y	0.86	0.75	0.80	0.80	0.79	0.82	0.82
Male sex	1.03[b]	1.46	1.26	1.27	1.34	1.16[b]	1.26
Log spot urine ACR[c]			1.60	1.61	1.55	1.42	1.37
Diabetes				0.86[b]			0.88[b]
Hypertension				1.17[b]			0.89[b]
Systolic BP, per 10 mm Hg					1.15		1.14
Diastolic BP, per 10 mm Hg					1.10		1.15
Body weight, per 10 kg					0.91		0.91
Serum albumin, per 0.5 g/dL						0.84	0.83
Serum phosphate, per 1.0 mg/dL						1.27	1.34
Serum bicarbonate, per 1.0 mEq/L						0.92	0.93
Serum calcium, per mg/dL						0.81	0.82
C statistic[d]	0.56	0.89	0.91	0.91	0.92	0.92	0.92
Akaike Information Criterion[d]	5553	4834	4520	4521	4463	4432	4378
P value		<.001	<.001	.40	<.001	<.001	<.001

Abbreviations: ACR, albumin-to-creatinine ratio; BP, blood pressure; GFR, glomerular filtration rate.
[a] Data are presented as hazard ratios unless otherwise specified. Models 2, 3, and 6 columns indicate models based on laboratory data. P values are for comparison of C statistics between successive models, except for models 5, 6, and 7, which are compared with model 3.
[b] Hazard ratios with $P>.05$; all other hazard ratios are significant (ie, $P<.05$).
[c] Hazard ratio for ACR represents a 1.0 higher ACR on the natural log scale. For the average patient with 20 mg/g of albuminuria, this represents an increase to 55 mg/g.
[d] Null values for C statistic and Akaike Information Criterion are 0.50 and 5569, respectively. Higher values for C statistic and lower values for Akaike Information Criterion indicate better models.

Figure. Observed vs Predicted Probability of Kidney Failure at 3 Years Using Models 2, 3, and 6 in the Validation Cohort
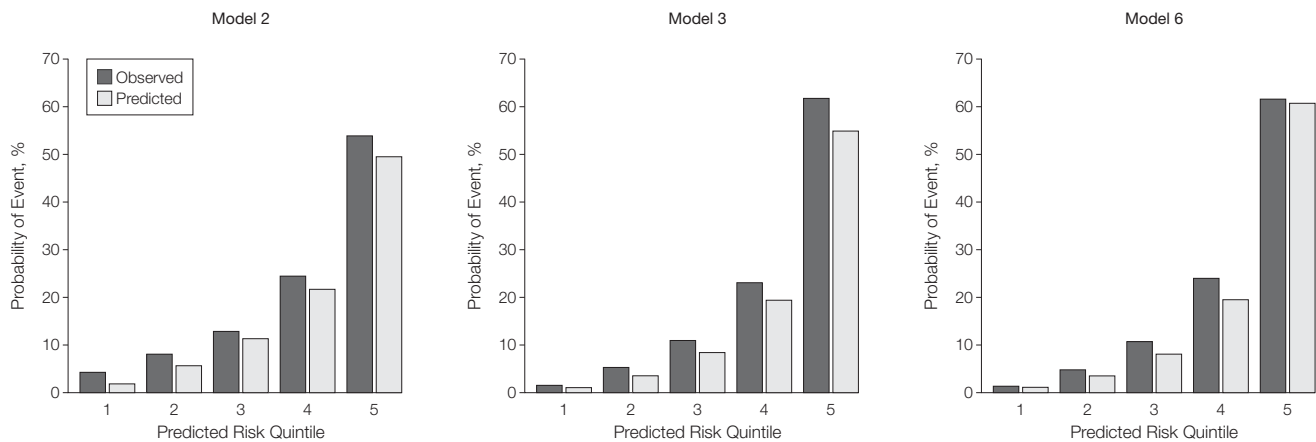
The predicted and observed event probability estimates represent the mean predicted probability from the Cox proportional hazards regression model and the mean observed probability from the population (Kaplan-Meier estimate) divided into quintiles of predicted probability. Predicted risk categories for quintiles 1 through 5 correspond with 0% to 4.3%, 4.4% to 8.1%, 8.2% to 12.9%, 13.0% to 24.5%, and 24.6% to 53.9%, respectively, for model 2; 0% to 1.6%, 1.7% to 5.3%, 5.4% to 11.0%, 11.1% to 23.1%, 23.2% to 61.7%, respectively, for model 3; and 0% to 1.4%, 1.4% to 4.8%, 4.9% to 10.7%, 10.8% to 24.0%, 24.1% to 61.6%, respectively, for model 6. Nam and D'Agostino χ^2 statistic is 37, 32, and 19 for models 2, 3, and 6, respectively.

The NRI risk for CKD stages 3 and 4 overall (**TABLE 3**) and by relevant subgroups (eTable 4) in the validation cohort were analyzed. Overall, model 6 outperformed models 2 and 3 with an NRI of 50.4% (95% CI, 42.7%-58.1%) and 8.0% (95% CI, 2.1%-13.9%), respectively, for CKD stage 3, and 26.7% (95% CI, 20.1%-33.3%) and 4.1% (95% CI, −0.5% to 8.8%), respectively, for CKD stage 4. The improvement in reclassification was similar across relevant subgroups.

Sensitivity Analyses

Reclassification analysis examining the effect of different thresholds for definitions of risk categories for CKD stages 3 and 4 showed consistent improvement with model 6 over models 2 and 3. Similarly, a category-less NRI showed a 30% (95% CI, 16%-44%) improvement in the overall reclassification for model 6 compared with model 3. Survival analysis using the competing risk approach showed no significant differences between the observed Kaplan-Meier method estimate and the cumulative incidence function, or between the predicted probabilities of kidney failure from the Cox proportional hazards regression model and the competing risk model (eFigure 2 and eFigure 3).

Table 3. Net Reclassification in the Validation Data Set for CKD Stages 3 and 4[a]

	No. (%) of Participants		
	NRI Events	Non-NRI Events	NRI and Non-NRI Events, No. (%) [95% CI]
Models	**CKD Stage 3**		
	(n = 248)	(n = 2159)	
3 vs 2	76 (30.6)	296 (13.7)	372 (44.4) [36.5 to 52.2]
6 vs 2	91 (36.7)	296 (13.7)	387 (50.4) [42.7 to 58.1]
6 vs 3	21 (8.5)	−11 (−0.5)	10 (8.0) [2.1 to 13.9]
Models	**CKD Stage 4**		
	(n = 400)	(n = 1695)	
3 vs 2	10 (2.5)	374 (22.1)	384 (24.6) [17.7 to 31.4]
6 vs 2	5 (1.3)	432 (25.5)	437 (26.7) [20.1 to 33.3]
6 vs 3	−2 (−0.5)	78 (4.6)	76 (4.1) [−0.5 to 8.8]

Abbreviations: CI, confidence interval; CKD, chronic kidney disease; NRI, net reclassification improvement.
[a] Risk categories for CKD stage 3 are 0% to 4.9%, 5.0% to 14.9%, and 15.0% or more over 5 years, and for CKD stage 4 are 0% to 9.9%, 10.0% to 19.9%, and 20.0% or more over 2 years.

COMMENT

We have developed and validated a set of risk prediction models for progression to kidney failure among patients with moderate to severe CKD. Accuracy was robust across relevant subgroups. Our models use laboratory data that are obtained routinely in patients with CKD and could be easily integrated into a laboratory information system or a clinic EHR. The incremental improvement in NRI and integrated discrimination improvement using our best model (model 6) compared with simpler models (models 2 and 3) is greater than the improvements observed with recent refinements to prediction tools used in other chronic diseases, such as the addition of C-reactive protein to traditional cardiovascular risk factors and the addition of serum sodium concentration to the Model for End-Stage Liver Disease score.[27,28]

Among patients with CKD, there can be considerable heterogeneity in risk for progression to kidney failure. For example, **TABLE 4** shows clinical and laboratory findings in 2 hypothetical patients with the same estimated GFR and kidney failure risk predictions from mod-

Table 4. Predicted Probability of Kidney Failure for 2 Hypothetical Patient Profiles Using Our Prediction Models[a]

	Probability of Kidney Failure, %	
Model	Patient A (70-year-old male, with estimated GFR of 30 mL/min/1.73 m² and urine ACR of 200 mg/g)[b]	Patient B (50-year-old male, with estimated GFR of 30 mL/min/1.73 m² and urine ACR of 50 mg/g)[c]
2	19.8	32.7
3	16.3	13.6
6	26.0	10.7

Abbreviations: ACR, albumin-to-creatinine ratio; GFR, glomerular filtration rate.
[a] For risk calculator, see http://www.jama.com. For smartphone app, see http://www.qxmd.com/Kidney-Failure-Risk-Equation.
[b] Patient A had the following serum laboratory values: calcium (9.0 mg/dL), phosphate (4.5 mg/dL), albumin (3.5 g/dL), and bicarbonate (21 mEq/L).
[c] Patient B had the following serum laboratory values: calcium (9.8 mg/dL), phosphate (3.8 mg/dL), albumin (4.0 g/dL), and bicarbonate (26 mEq/L).

els 2, 3, and 6. Predictions based on estimated GFR alone are not sufficient for risk prediction in these 2 patients. The addition of the laboratory variables in models 3 and 6 provides a substantially different risk prediction compared with model 2. Compared with model 3, model 6 increases the predicted risk by 9.7% for patient A and decreases the risk by 2.9% for patient B.

Risk prediction has gained increasing attention over the last 2 decades, with emerging literature suggesting improved patient outcomes with individualized risk prediction and with advances in information technology that allow for easy implementation of risk prediction models as components of EHRs.[29-33] The availability of these risk prediction tools and their inclusion in clinical practice guidelines have led to better adherence to treatment guidelines and have encouraged individual decision making.[31-33] Despite these benefits, the lack of easily applicable and externally validated models have delayed the widespread integration of risk prediction in all fields of medicine.[34,35]

Our models rely on demographic data and laboratory markers of CKD severity to predict the risk of future kidney failure. Similar to previous investigators from Kaiser Permanente and the RENAAL study group,[15,16] we find that a lower estimated GFR, higher albuminuria, younger age, and male sex predict faster progression to kidney failure. In addition, a lower serum albumin, calcium, and bicarbonate, and a higher serum phosphate also predict a higher risk of kidney failure and add to the predictive ability of estimated GFR and albuminuria. These markers may enable a better estimate of measured GFR or they may reflect disorders of tubular function or underlying processes of inflammation or malnutrition.[36,37]

Although these laboratory markers have also previously been associated with progression of CKD, our work integrates them all into a single risk equation (risk calculator and eTable 5, available at http://www.jama.com, and smartphone app, available at http://www.qxmd.com/Kidney-Failure-Risk-Equation). In addition, we demonstrate no improvement in model performance with the addition of variables obtained from the history (diabetes and hypertension status) and the physical examination (systolic blood pressure, diastolic blood pressure, and body weight). Although these other variables are clearly important for diagnosis and management of CKD, the lack of improvement in model performance may reflect the high prevalence of these conditions in CKD and imprecision with respect to disease severity after having already accounted for estimated GFR and albuminuria.

Our risk prediction models have important implications for clinical practice, research, and public health policy. For example, in CKD stage 3, the relative contribution of the nephrologist vs the primary care physician to CKD care is uncertain.[38] Using our models, lower-risk patients could be managed by the primary care physician without additional testing or treatment of CKD complications; whereas, higher-risk patients could receive more intensive testing, intervention, and early nephrology care. Similarly, in CKD stage 4, the timing of appropriate predialysis interventions remains uncertain.[39] Using our models, different risk thresholds could be used to triage patients for decisions

regarding dialysis modality education, vascular access creation, and preemptive transplantation. Furthermore, our models could be used to select higher-risk patients for enrollment in clinical trials and for evaluation of risk-treatment interactions.[40,41] In addition, our models may also be useful for identifying high-risk patients with CKD stage 3 for public health interventions, thereby improving the cost-effectiveness of CKD care.[42]

The strengths of our analysis are the development and validation of highly accurate prediction tools. The models are practical because all the variables included are collected routinely in clinical care for CKD. The models can be easily integrated into laboratory information systems and clinic EHRs, enabling integration of a risk prediction tool as a clinical decision support aid in outpatient care. In addition, the models were developed in a single-payer health care system, thereby disparities in referral patterns and CKD care were less likely to be present.[43] In addition, we have proposed several possible applications for our risk prediction tools in clinical decision making for CKD stages 3 and 4, which can be tested in pragmatic cluster randomized trials and formal cost-effectiveness analyses, that could help translate our models to the bedside.

Our analysis has a few limitations. First, our study population consisted of 2 distinct nephrology referred CKD populations; therefore, the generalizability of our findings to a nonreferred CKD population remains to be determined. Second, although we included patients with a wide spectrum of CKD and ranges of estimated GFR, certain ethnic minorities with increased risk for kidney failure, such as black and Native American individuals are underrepresented. Third, we did not explicitly model the risk of all-cause mortality in our CKD population. Although patients with CKD are at higher risk for mortality, knowing the probability of kidney failure among the survivors has clinical use. In addition, we did not observe a systematic overprediction of kidney failure risk in our models and using a competing-risk ap-

proach did not significantly change our predicted probabilities.

In conclusion, we have developed and validated highly accurate predictive models for progression of CKD to kidney failure. Our best model uses routinely available laboratory data and can predict the short-term risk of kidney failure with accuracy and could be easily implemented in a laboratory information system or an EHR. External validation in multiple diverse CKD cohorts and evaluation in clinical trials are needed.

REFERENCES

1. Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298(17):2038-2047.
2. Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305.
3. Sarnak MJ, Greene T, Wang X, et al. The effect of a lower target blood pressure on the progression of kidney disease. *Ann Intern Med*. 2005;142(5):342-351.
4. Sarnak MJ, Levey AS. Cardiovascular disease and chronic renal disease: a new paradigm. *Am J Kidney Dis*. 2000;35(4)(suppl 1):S117-S131.
5. Weiner DE, Tighiouart H, Amin MG, et al. Chronic kidney disease as a risk factor for cardiovascular disease and all-cause mortality. *J Am Soc Nephrol*. 2004;15(5):1307-1315.
6. Zhang QL, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. *BMC Public Health*. 2008;8:117.
7. Keane WF, Zhang Z, Lyle PA, et al. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol*. 2006;1(4):761-767.
8. Keith DS, Nichols GA, Gullion CM, et al. Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. *Arch Intern Med*. 2004;164(6):659-663.
9. Taal MW, Brenner BM. Predicting initiation and progression of chronic kidney disease: developing renal risk scores. *Kidney Int*. 2006;70(10):1694-1705.
10. Taal MW, Brenner BM. Renal risk scores: progress and prospects. *Kidney Int*. 2008;73(11):1216-1219.
11. Levey AS, Coresh J, Balk E, et al. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ann Intern Med*. 2003;139(2):137-147.
12. Hemmelgarn BR, Zhang J, Manns BJ, et al. Nephrology visits and health care resource use before and after reporting estimated glomerular filtration rate. *JAMA*. 2010;303(12):1151-1158.
13. Jain AK, McLeod I, Huo C, et al. When laboratories report estimated glomerular filtration rates in addition to serum creatinines, nephrology consults increase. *Kidney Int*. 2009;76(3):318-323.
14. Matsushita K, van der Velde M, Astor BC, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts. *Lancet*. 2010;375(9731):2073-2081.
15. Geddes CC, Fox JG, Allison ME, et al. An artificial neural network can select patients at high risk of developing progressive IgA nephropathy more accurately than experienced nephrologists. *Nephrol Dial Transplant*. 1998;13(1):67-71.
16. Johnson ES, Thorp ML, Platt RW, Smith DH. Predicting the risk of dialysis and transplant among patients with CKD. *Am J Kidney Dis*. 2008;52(4):653-660.
17. Levin A, Djurdjev O, Beaulieu M, Er L. Variability and risk factors for kidney disease progression and death following attainment of stage 4 CKD in a referred cohort. *Am J Kidney Dis*. 2008;52(4):661-671.
18. Mackinnon B, Fraser EP, Cattran DC, et al. Validation of the Toronto formula to predict progression in IgA nephropathy. *Nephron Clin Pract*. 2008;109(3):c148-c153.
19. Parving HH, Lehnert H, Bröchner-Mortensen J, et al. The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. *N Engl J Med*. 2001;345(12):870-878.
20. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004;23(13):2109-2123.
21. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27(2):157-172.
22. Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating.* New York, NY: Springer; 2009.
23. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.
24. D'Agostino RB, Nam BH. *Evaluation of the Performance of Survival Analysis Models: Discrimination and Calibration Measures.* San Francisco, CA: Elsevier; 2004:1-25.
25. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30(1):11-21.
26. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26(11):2389-2430.
27. Kim WR, Biggins SW, Kremers WK, et al. Hyponatremia and mortality among patients on the liver-transplant waiting list. *N Engl J Med*. 2008;359(10):1018-1026.
28. Wilson PW, Pencina M, Jacques P, et al. C-reactive protein and reclassification of cardiovascular risk in the Framingham Heart Study. *Circ Cardiovasc Qual Outcomes*. 2008;1(2):92-97.
29. Batsis JA, Lopez-Jimenez F. Cardiovascular risk assessment: from individual risk prediction to estimation of global risk and change in risk in the population. *BMC Med*. 2010;8:29.
30. D'Agostino RB Sr, Grundy S, Sullivan LM, et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA*. 2001;286(2):180-187.
31. Leslie WD, Morin S, Lix LM. A before-and-after study of fracture risk reporting and osteoporosis treatment initiation. *Ann Intern Med*. 2010;153(9):580-586.
32. Pozen MW, D'Agostino RB, Selker HP, et al. A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease. *N Engl J Med*. 1984;310(20):1273-1278.
33. Selker HP, Beshansky JR, Griffith JL, et al. Use of the acute cardiac ischemia time-insensitive predictive instrument (ACI-TIPI) to assist with triage of patients with chest pain or other symptoms suggestive of acute cardiac ischemia. *Ann Intern Med*. 1998;129(11):845-855.
34. Müller-Riemenschneider F, Holmberg C, Rieckmann N, et al. Barriers to routine risk-score use for healthy primary care patients. *Arch Intern Med*. 2010;170(8):719-724.
35. van Steenkiste B, van der Weijden T, Stoffers HE, Grol R. Barriers to implementing cardiovascular risk tables in routine general practice. *Scand J Prim Health Care*. 2004;22(1):32-37.
36. de Brito-Ashurst I, Varagunam M, Raftery MJ, Yaqoob MM. Bicarbonate supplementation slows progression of CKD and improves nutritional status. *J Am Soc Nephrol*. 2009;20(9):2075-2084.
37. Schwarz S, Trivedi BK, Kalantar-Zadeh K, Kovesdy CP. Association of disorders in mineral metabolism with progression of chronic kidney disease. *Clin J Am Soc Nephrol*. 2006;1(4):825-831.
38. Blantz RC. Handing out grades for care in chronic kidney disease: nephrologists versus non-nephrologists. *Clin J Am Soc Nephrol*. 2007;2(2):193-195.
39. O'Hare AM, Bertenthal D, Walter LC, et al. When to refer patients with chronic kidney disease for vascular access surgery: should age be a consideration? *Kidney Int*. 2007;71(6):555-561.
40. Kent D, Hayward R. Subgroup analyses in clinical trials. *N Engl J Med*. 2008;358(11):1199.
41. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients. *JAMA*. 2007;298(10):1209-1212.
42. Manns B, Hemmelgarn B, Tonelli M, et al. Population based screening for chronic kidney disease: cost effectiveness study. *BMJ*. 2010;341:c5869.
43. Lasser KE, Himmelstein DU, Woolhandler S. Access to care, health status, and health disparities in the United States and Canada. *Am J Public Health*. 2006;96(7):1300-1307.