CKD Progression Prediction in a Diverse US Population: A Machine-Learning Model

Joseph Aoki, Cihan Kaya, Omar Khalid, Tarush Kothari, Mark A. Silberman, Con Skordis, Jonathan Hughes, Jerry Hussong, and Mohamed E. Salama

Rationale & Objective: Chronic kidney disease (CKD) is a major cause of morbidity and mortality. To date, there are no widely used machine-learning models that can predict progressive CKD across the entire disease spectrum, including the earliest stages. The objective of this study was to use readily available demographic and laboratory data from Sonic Healthcare USA laboratories to train and test the performance of machine learning-based predictive risk models for CKD progression.

Study Design: Retrospective observational study

Setting & Participants: The study population was composed of deidentified laboratory information services data procured from a large US outpatient laboratory network. The retrospective data set included 110,264 adult patients over a 5-year period with initial estimated glomerular filtration rate (eGFR) values between 15-89 mL/min/1.73 m^2.

Predictors: Patient demographic and laboratory characteristics.

Outcomes: Accelerated (ie, >30%) eGFR decline associated with CKD progression within 5 years.

Analytical Approach: Machine-learning models were developed using random forest survival methods, with laboratory-based risk factors analyzed as potential predictors of significant eGFR decline.

Results: The 7-variable risk classifier model accurately predicted an eGFR decline of >30% within 5 years and achieved an area under the curve receiver-operator characteristic of 0.85. The most important predictor of progressive decline in kidney function was the eGFR slope. Other key contributors to the model included initial eGFR, urine albumin-creatinine ratio, serum albumin (initial and slope), age, and sex.

Limitations: The cohort study did not evaluate the role of clinical variables (eg, blood pressure) on the performance of the model.

Conclusions: Our progressive CKD classifier accurately predicts significant eGFR decline in patients with early, mid, and advanced disease using readily obtainable laboratory data. Although prospective studies are warranted, our results support the clinical utility of the model to improve timely recognition and optimal management for patients at risk for CKD progression.

Visual Abstract included

Complete author and article information provided before references.

Correspondence to J. Aoki (Joseph.aoki@hawaiilabs.com)

Kidney Med. 5(9):100692. Published online June 24, 2023.

doi: 10.1016/j.xkme.2023.100692

Chronic kidney disease (CKD) affects over 37 million people in the United States and is a tremendous source of morbidity and mortality worldwide.[1,2] CKD is strongly associated with hematologic,[3,4] metabolic bone,[5-7] cerebrovascular[8-10] and cardiovascular[11-13] disease, and it is the most common cause of kidney failure, requiring dialysis or transplant.[1] As a complex disorder that affects critical organ systems, CKD accounts for a disproportionate health care system expenditure, with recent studies estimating an overall Medicare cost of $87.2 billion in 2019.[1,14] However, up to 90% of patients with CKD (including 40% of patients with severe CKD) are unaware of their diagnosis, which precludes timely evaluation and optimal management of the disease.[15,16]

Kidney Disease Improving Global Outcomes (KDIGO) clinical practice guidelines classify CKD into categories based on glomerular filtration rate [≥90 (G1), 60-89 (G2), 45-59 (G3a), 30-44 (G3b), 15-29 (G4), <15 mL/min/1.72 m^2 (G5)] and albuminuria [<30 (A1), 30-299 (A2), ≥300 mg/g creatinine (A3)]. The clinical stage is established most often by laboratory evaluation for serum creatinine-based estimated glomerular filtration rate (eGFR) and urine albumin-creatinine ratio (UACR), respectively.[17]

Based primarily on studies from the Alberta Kidney Disease Network (AKDN),[18] CKD progression is defined by 1 or more of the following: (1) worsening in the glomerular filtration rate category and a ≥25% drop in the eGFR from the baseline; or (2) a decline in the eGFR by >5 mL/min/1.73 m^2 per year (rapid progression). To curb progression to kidney failure in this at-risk population, KDIGO recommends that patients with progressive CKD receive aggressive treatment and specific management measures.[17]

To assist with planning for kidney replacement therapy, KDIGO also endorses using validated risk prediction tools to identify patients at high risk for kidney failure. Currently, the Kidney Failure Risk Equation (KFRE) is the most widely accepted risk prediction model for this adverse outcome. Initially derived from a Canadian cohort of 8,391 persons with intermediate-advanced CKD (stages 3a-5) and using Cox proportional hazards regression methods, the 4-variable (age, sex, eGFR, and UACR) model uses demographic characteristic data and spot laboratory test results to accurately predict the 2-year and 5-

PLAIN-LANGUAGE SUMMARY

Defined by a significant decrease in estimated glomerular filtration rate (eGFR), chronic kidney disease (CKD) progression is strongly associated with kidney failure. However, to date, there are no broadly used resources that can predict this clinically significant event. Using machine-learning techniques on a diverse US population, this cohort study aimed to address this deficiency and found that a 5-year risk prediction model for CKD progression was accurate. The most important predictor of progressive decline in kidney function was the eGFR slope, followed by the urine albumin-creatinine ratio and serum albumin slope. Although further study is warranted, the results showed that a machine-learning model using readily obtainable laboratory information accurately predicts CKD progression, which may inform clinical diagnosis and management for this at-risk population.

year risk of progression to kidney failure with excellent discrimination (AUC = 0.84).[19]

Although several studies have further validated KFRE across a variety of clinical settings[20-24] a limitation of the classifier is that, by definition, it is designed exclusively to predict final loss of function. Recently, work by Wang et al[16] reported that in a large retrospective cohort, over 85% of patients with CKD had either no or a low (<3%) risk of kidney failure according to the 5-year KFRE. This represents a substantial population of patients with CKD for whom further risk stratification is warranted. However, to date, there are no broadly adopted machine-learning tools that predict a clinically significant and much earlier event—progressive CKD—across the wide spectrum of disease. Addressing this knowledge gap would identify high risk patients within this heterogeneous population who may benefit from targeted treatment and management at an earlier stage to preserve kidney function and prevent adverse outcomes.[25]

Using longitudinal clinical laboratory data across a geographically diverse cohort, the purpose of our study was to use machine-learning methods to develop a readily obtainable risk prediction model for >30% decline in eGFR or progressive CKD. Because earlier studies exclusively interrogated mid-late CKD stage (G3-5) cohorts, our investigation aimed to build a classifier that incorporates routine variables across early (G2), moderate (G3), and advanced (G4) stages of CKD.

METHODS

Study Population

The study was reviewed and approved by the institutional review board at Western Institutional Review Board—Copernicus Group (WCG) (approval number 20222952). Informed consent was waived because of the lack of feasibility of obtaining consent from all participants, and that the data reviewed were deidentified, record-based, and retrospective in nature. The study cohort included deidentified data from 330,238 participants from the Northeast, Southwest, Mid-South, and West or Pacific regions of the United States. The overall median age was 65 years old, and 51% of the participants were men. The cohort included participants with a broad spectrum of baseline eGFR results associated with G2 (60%), G3 (36%), and G4 (4%). The percentage of participants with initial albuminuria values of A1, A2, and A3 were 65%, 23%, and 12%, respectively.

Data Collection and Measurements

Creatinine testing was performed on the same high-throughput instrument platform across all testing sites and was standardized according to best practices, such as calibration traceable to an isotope dilution mass spectrometry (IDMS) reference measurement procedure. The eGFR was calculated using the 2009 CKD Epidemiology Collaboration (CKD-EPI) creatinine equation without the race-based coefficient.[26]

The training and test cohorts were derived from the laboratory information system (LIS) at outpatient facilities within Sonic Healthcare USA. To evaluate particularly early-to-advanced CKD, the cohort was selected for patients with eGFR values between 89 and 15 mL/min/1.73 m^2 (G2-G4) and dates of service between January 1, 2017, and December 31, 2021. The outcomes were determined by reviewing records contained in the LIS.

Candidates were selected based on the availability of sufficient demographic data (eg, age and sex) and laboratory results for eGFR, albumin, and UACR. The minimum criteria for inclusion in the study were: 3 eGFR values over a span of ≥12 months and ≥1 value for serum albumin and UACR. On the basis of previous work[27-32] these data and additional clinical laboratory values were evaluated as continuous or time-based variables. Where applicable, slopes for specific variables were calculated using the linear regression function within the Python Sklearn linear model. The slopes were derived from baseline values and adjusted using an annual time interval.

To broadly ensure all criteria for observed CKD progression[17] were met within the cohort, such as participants in the G2 and G3a subsets, the outcome of interest was defined as >30% eGFR decline with confirmation at any point over the course of 5 years. The rationale for this outcome is further supported by several studies that suggest that >30% eGFR decline is a clinically meaningful surrogate end point for progressive kidney disease.[33-35]

Statistical Analysis

Models were built by creating training and testing datasets, using 80% of the data for training and 20% for independent testing, respectively. The training and testing datasets were stratified to ensure similar patient cohorts. After filtering, a random forest (RF) classifier was built using Sklearn version 1.1.1. Ten thousand trees were generated within the models to develop the classifier. Additional 5-

Table 1. Baseline Characteristics

	Training Set	Test Set
Patient Demographics		
Number of unique individuals	88,211	22,053
Age (y), median (IQR)	65 (12)	65(12)
Male (%)	44,860 (51)	11,080 (50)
Laboratory Values		
eGFR, mL/min/1.73 m^2, initial, median (IQR)	64 (24)	64 (24)
15-29 (%)	3,107(4)	783 (4)
30-59 (%)	32,284 (36)	8,018 (36)
60-89 (%)	52,820 (60)	13,252 (60)
Serum calcium, mg/dL, initial, median (IQR)	9.6 (0.6)	9.6 (0.6)
Hemoglobin A1c, (%), initial, median (IQR)	6.6 (1.9)	6.6 (1.9)
Serum albumin, g/dL, initial, median (IQR)	4.4 (0.5)	4.4 (0.5)
Urine albumin-to-creatinine ratio, mg/g		
UACR, initial, median (IQR)	14 (60)	14 (58)
<30 (%)	57,067 (65)	14,258 (65)
30-299 (%)	19,883 (23)	5,012 (22)
≥300 (%)	11,261 (12)	2,783 (13)
End point, >30% decline in eGFR		
Events (%)	15,301 (17)	3,821 (17)
Progression time, mo (IQR)	20.9 (22)	19.9 (22)
Observation time, mo (IQR)	40.0 (28.7)	40.0 (28.7)

Note: Progression time is defined as the average number of months identified for eGFR decline of 30% from baseline. Observation time is defined as the average number of months of laboratory data procured for the cohort.
Abbreviations: IQR, interquartile range; eGFR, estimated glomerular filtration rate.

fold cross validation was performed on the training set, and a representative model was used in the testing set to assess the performance. R version 4.2.1 was used to perform additional data wrangling, and statistical analysis was reported using the tidyverse packages and the equivalence package. In time-to-event analysis, participants were censored if the event (ie, >30% eGFR decline) occurred during the designated slope evaluation time period. Participants lost-to-follow-up who did not experience an event before the end of the maximum 5-year follow-up time were censored at 36 months from the last encounter.

Variable importance was calculated as part of the RF analysis. Candidate variables that were supported by the dataset and showed a meaningful impact to the classifier were reported as a percentage. The predictive accuracy was assessed using the area under the receiver-operator characteristic curve (AUC) on the test set. Calibrated and uncalibrated RF models were developed using isotonic and sigmoid calibrations from Scikit-learn. Curves comparing observed positive classification versus predicted probability of positive classification were also generated (Fig S1). In addition, calibration was performed with additional time series data going from 6 months to 36 months, and the 95% confidence interval and Brier score loss were generated for all models.

RESULTS

Baseline Characteristics

The initial registry included 330,238 adult patients aged 18 to 75 years old with an outpatient clinical laboratory encounter within the Sonic Healthcare USA network between January 1, 2017, and December 31, 2021 (Table 1). To account for the well-established challenge of inter-laboratory variation, creatinine analysis was performed on a single high-throughput instrument platform and standardized according to best practices, such as calibration traceable to an IDMS reference measurement procedure.[36] The eGFR was calculated using the 2009 CKD Epidemiology Collaboration (CKD-EPI) creatinine equation.[26] Patients with eGFR <15 mL/min/1.73 m^2 or >89 mL/min/1.73 m^2, no serum albumin or UACR results, or less than 12 months of follow-up data were omitted from the study.

After applying exclusion criteria, the abnormal eGFR dataset contained 110,264 patients (Fig 1) with a median age of 66 years and a similar distribution between men and women. The cohort included participants with a broad spectrum of baseline eGFR results associated with G2 (60%), G3 (36%), and G4 (4%) and initial UACR values of <30 mg/g (65%), 30-299 mg/g (23%), and ≥ 300 mg/g (12%). On average, the observation time was 40 months, and follow-up was adequate with 9 eGFR results per person (1.8 eGFR tests per person per year) over the course of the study.

Machine-Learning Models for CKD Progression

Models using 2 variables (initial eGFR and eGFR slope) or 7 variables (age, sex, initial eGFR, eGFR slope, initial UACR, initial serum albumin, and serum albumin slope) with 5-fold cross validation were trained and tested against a >30% decline in eGFR within 5 years. Compared with

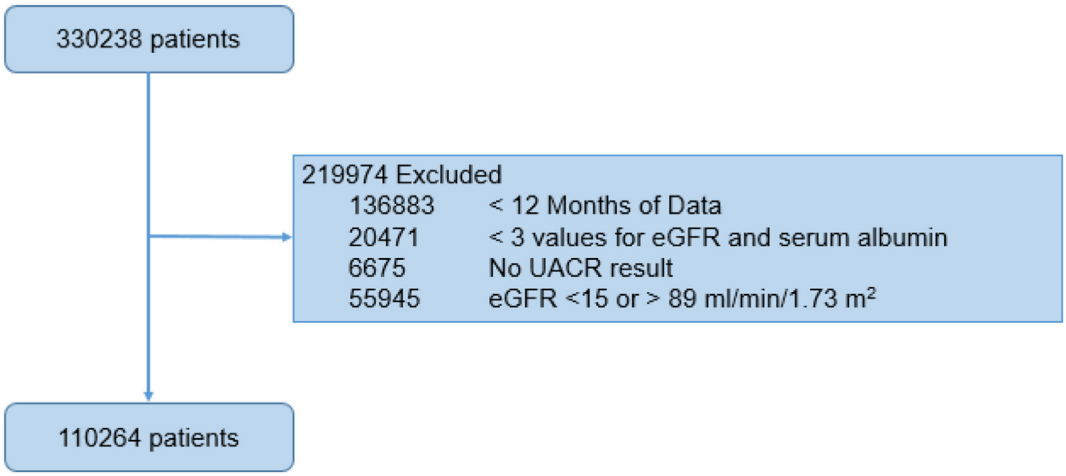


Figure 1. Participant flow diagram. The flow diagram depicts the number of adult participants between 18 and 75 years of age in the original dataset before removal because of exclusion criteria. Omitted participants included those with less than 12 months of data, less than 3 eGFR and serum albumin values, less than 1 UACR value, and initial eGFR <15 or >89 mL/min/1.73 m^2. The final data set included 110,264 participants.

the 2-variable model, the 7-variable model was more accurate and achieved an overall AUC of 0.85 (Fig 2).

Further analysis was performed on the 7-variable progressive CKD risk classifier. The uncalibrated model was similar to the calibrated model [S1], with a Brier score loss of 0.01 between the models. In sensitivity analysis, the cohort was selected exclusively for G2 or G3 disease, and the performance was similar (AUC = 0.85). Keeping with previous work by Inker et al,[34] the classifier improved when the eGFR slope input was extended from 6-month (AUC = 0.76, Brier score 0.017), 18-month (AUC = 0.81, Brier score 0.015), and 36-month (AUC = 0.83, Brier score 0.015) time intervals. Examined across regional cohorts, the model performed as expected across the Mid-South, West or Pacific, Southwest, and Northeast with AUC values of 0.83-0.88 (Brier score 0.013-0.02). For time-to-event and regional analysis, the 95% confidence interval was within 0.01.

Variable Importance

Overall, the change of the eGFR over time, so-called eGFR velocity or slope, was identified as the highest contributor

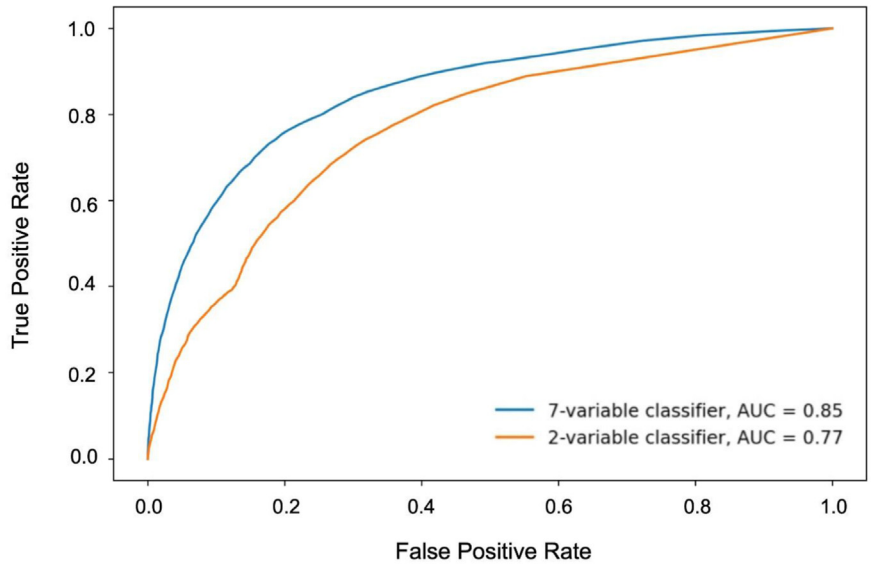


Figure 2. Receiver-operator characteristic curves of classifier models showing prediction performance for >30% decline in eGFR. The 2-variable classifier included initial eGFR and eGFR slope. The 7-variable classifier included initial eGFR, eGFR slope, UACR, initial serum albumin, serum albumin slope, age, and sex. eGFR, estimated glomerular filtration rate; UACR, urine albumin-creatinine ratio.

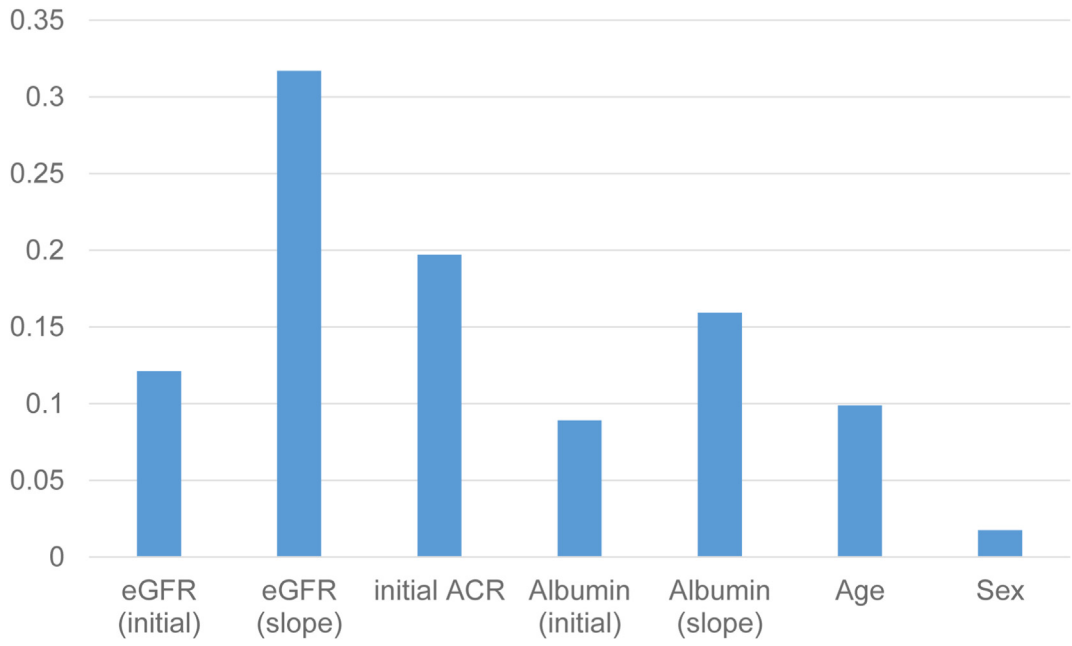


Figure 3. Variable importance. eGFR, estimated glomerular filtration rate; UACR, urine albumin-creatinine ratio.

to the risk prediction model. Other significant inputs to the model included initial eGFR, initial UACR, initial albumin, albumin slope, age, and sex (Fig 3).

DISCUSSION

Using machine-learning random forest survival (RFS) analysis, this retrospective cohort study trained and tested a risk prediction model for >30% decline in the eGFR among a large, diverse US adult population with abnormal kidney function. Consistent with previous studies, our findings further confirm the following static variables that are associated with substantial decline in kidney function: age, sex, eGFR, UACR, and serum albumin.[19,37] The accuracy of the progressive CKD risk classifier was robust, with an overall AUC value of 0.85.

A key strength of this investigation is that the risk prediction model is uniquely based on machine-learning RFS analysis that was able to interrogate for the effect of time-based variables. As opposed to using the Cox proportional hazards regression method and other regression models, recent reports suggest that RFS may be less prone to variance and better suited to solving complex problems with non-linear or continuous features.[38-40] As such, the machine-learning techniques deployed here identified that longitudinal results over time (eg, slope or velocity between multiple values) for select variables significantly enhance the overall performance.

Among continuous variables, the eGFR slope was the most significant contributor to the classifier. This finding supports numerous recommendations by authorities that highlight that interpreting the eGFR results is best performed over a series of studies rather than in isolation.[1,17,25] In addition, the importance of eGFR velocity as a contributor to predicting CKD progression is in line with other disease entities that identified change over time in relevant laboratory biomarker values as a clinically significant (and actionable) event. Paired examples include myocardial infarction, chronic lymphocytic leukemia, and prostate cancer with serial changes in troponin,[41] lymphocyte cell count,[42] and prostate-specific antigen,[43] respectively. Although further study is warranted, these results add to the growing body of evidence showing the important role of eGFR slope in predicting clinically significant disease outcomes, such as progressive CKD.[44-47]

Another strength of the progressive CKD risk classifier is that it leverages existing and readily obtainable information found within the highly used kidney profile[48] and could be broadly deployed in a value-based manner. However, the present data suggest that it may be effective using as little as 6-18 months of previous laboratory data, then optimized at 36 months with successive encounters. In one of the largest studies to date, Song et al[49] estimates that, beginning in the third decade of life, US patients undergo ~1.1 chemistry tests (including creatinine or eGFR) on average per year. This suggests that, by the fourth decade of life, procuring 18-36 months of longitudinal eGFR results at baseline is currently feasible for most laboratory service providers, well before the peak CKD prevalence at 65 years of age.[1] Furthermore, the model avoids novel and potentially costly biomarkers, and despite geographic, socio-economic, and ethnic differences, the classifier performed well across all major regions with an AUC range of 0.83-0.88. Together, these findings strengthen the overall usability, cost-efficiency, and generalizability of the model.

Finally, unlike the KFRE, which was designed to particularly predict kidney failure, this is among the first reported machine-learning models to accurately identify patients at risk for a significant decline in the eGFR associated with a more upstream event: progressive CKD. Moreover, the current model was developed for major stages of disease, such as those with largely preserved kidney function (eGFR of 60-89 mL/min/1.73 m^2). The importance and relevance of the findings in this report is highlighted by the recent consensus from the KDIGO controversies conference, which underscored the need to improve risk stratification, particularly for early CKD.[50] Although prospective studies are warranted, together, these findings suggest that the progressive CKD risk classifier could be used in conjunction with the KFRE to support accurate, value-based risk prediction at a significantly earlier stage to maximize the benefit of evidence-based interventions, such as more frequent monitoring, avoidance of inciting agents, patient education, treatment modification, and specialist referral.

Recently, work by Chan et al,[51] Ferguson et al,[52] and Grams et al.[53] successfully generated risk prediction models for CKD progression. Although there are similarities, the current study uniquely leverages a large and diverse US population, applies a standardized assay for creatinine, or eGFR, and exclusively uses routine laboratory results. By contrast, the Chan and Ferguson machine-learning models rely primarily on novel kidney disease biomarkers and a Canadian cohort, respectively.[51,52] Moreover, the effects of continuous variables, such as the eGFR and serum albumin slope, were not reported. Using logistic regression, the Grams et al[53] model included both laboratory and clinical variables (eg, systolic blood pressure, body mass index, and medication history). In addition, the meta-analysis showed that the eGFR slope had a negligent to modest effect on the classifier's performance.[53] Compared with the present findings, the dissimilar results for the eGFR slope may be explained, in part, by the different input variables, study design (observational cohort versus meta-analysis), and methodology (logistic regression vs machine learning or RFS).[54,55] Taken together, these data underscore the overall success of risk classifiers using a variety of independent variables and methods. To be sure, the findings from the emerging field are dynamic, and prospective, head-to-head studies are required to establish the optimal predictive model for progressive CKD.

The study has several limitations. First, the cohort defined CKD based exclusively on laboratory values rather than an established clinical diagnosis. However, multiple studies have shown that CKD in the general population is frequently undiagnosed[1,14] because of various reasons, such as a lack of awareness and a lack of clinical and administrative resources.[56] As a result, studies that exclusively evaluate clinically confirmed CKD are often limited to later stages of the disease and preclude the interrogation of a significant undiagnosed population. Because of the robust size and inclusion of undiagnosed disease, the findings seen here may be more applicable to the general population with CKD in the United States. If implemented broadly, it could also improve the overall recognition and diagnosis of disease. Second, the study did not particularly address how the classifier performs in the subset of patients with CKD with >5 mL/min/1.73 m^2 decline per year, so-called rapid CKD progression.[17] Because of the importance of eGFR slope in the model, it is uniquely positioned to capture rapid eGFR decline over time; however, further study is warranted to understand its role in this clinically significant population. Finally, the study design did not evaluate important clinical variables, such as body mass index, blood pressure, smoking, medications, and other CKD risk factors. Although incorporating these elements may have improved the accuracy, it would also have added more complexity and, arguably, decreased the overall application and usability of the model. However, these limitations raise an important question about how the risk prediction model functions in targeted populations associated with kidney disease. This question and others should be answered in prospective cohorts, enriched for cardiovascular disease, diabetes, and other disorders linked with progressive CKD.

In conclusion, the progressive CKD risk classifier was accurate and achieved an AUC of 0.85. The largest overall contributor to the model was the eGFR slope. Other significant risk factors included static or longitudinal features for UACR, serum albumin, age, and sex. The model was derived from routine demographic and laboratory values in a very large and geographically diverse US outpatient cohort, which strengthens its overall generalizability and usability. Used as a complement to and in conjunction with the well-established KFRE, the progressive CKD risk classifier has the potential to significantly improve timely recognition, risk stratification, and optimal management for a heterogeneous population with CKD at a much earlier stage for intervention.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Figure S1: Calibration curves comparing observed positive classification versus predicted probability of positive classification.

ARTICLE INFORMATION

Authors' Full Names and Academic Degrees: Joseph Aoki, MD, Cihan Kaya, PhD, Omar Khalid, PhD, Tarush Kothari, MD, Mark A. Silberman, MD, Con Skordis, BS, Jonathan Hughes, MD, PhD, Jerry Hussong, MD, Mohamed E. Salama, MD

Authors' Affiliations: Sonic Healthcare USA.

Address for Correspondence: Joseph Aoki, MD, 99-193 Aiea Heights Drive, Aiea, HI 96701. Email: Joseph.aoki@hawaiilabs.com

Authors' Contributions: Research idea and study design: JA, CK, TK, MAS, MES; data acquisition: OK, CS; data analysis/interpretation: JA, CK, OK, TK, MS, JoH, MES; statistical analysis: CK, OK; supervision or mentorship: JeH, JoH, MES. Each author contributed important intellectual content during manuscript

drafting or revision and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate.

Support: None.

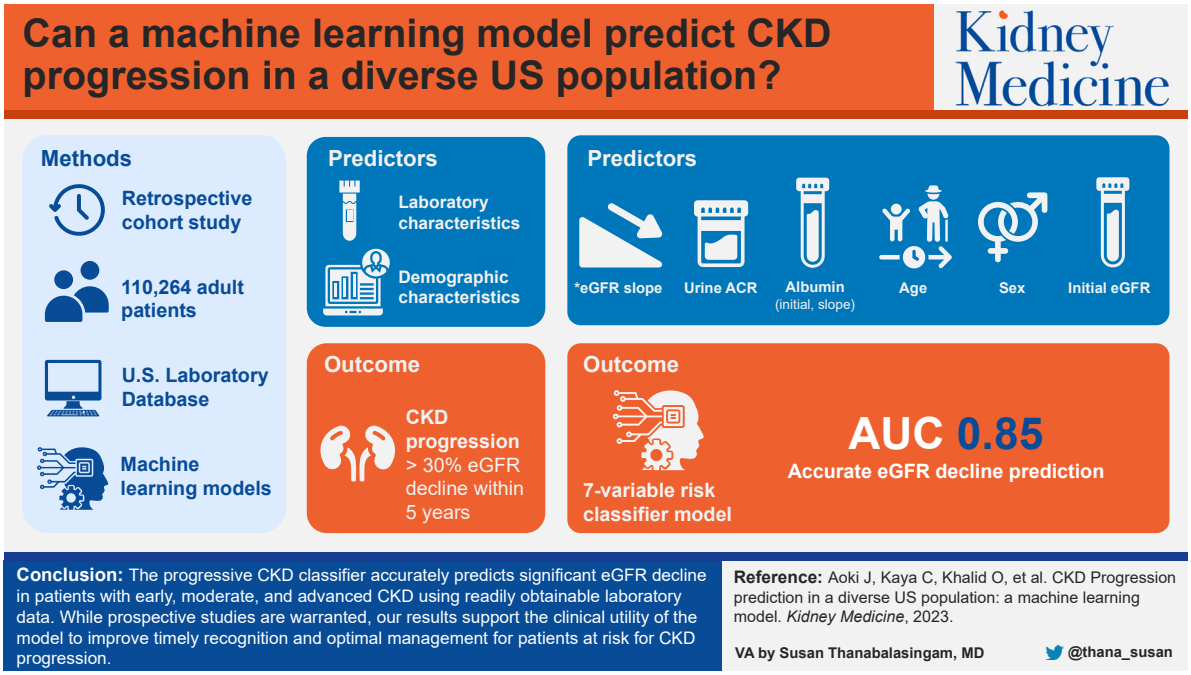
Financial Disclosure: The authors declare that they have no relevant financial interests.

Peer Review: Received December 20, 2022. Evaluated by 1 external peer reviewer, with direct editorial input from the Statistical Editor, an Associate Editor, and the Editor-in-Chief. Accepted in revised form May 9, 2023.

REFERENCES

1. Chronic Kidney Disease in the United States, 2021. US Department of Health and Human Services. Centers for Disease Control and Prevention. Accessed August 8, 2023. https://www.cdc.gov/kidneydisease/publications-resources/CKD-national-facts.html
2. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet*. 2017;389(10075):1238-1252. doi:10.1016/S0140-6736(16)32064-5
3. Babitt JL, Lin HY. Mechanisms of anemia in CKD. *J Am Soc Nephrol*. 2012;23(10):1631-1634. doi:10.1681/ASN.2011111078
4. Stauffer ME, Fan T. Prevalence of anemia in chronic kidney disease in the United States. *PLoS One*. 2014;9(1):e84943. doi:10.1371/journal.pone.0084943
5. Ketteler M, Block GA, Evenepoel P, et al. Diagnosis, evaluation, prevention, and treatment of chronic kidney disease-mineral and bone disorder: synopsis of the Kidney Disease: Improving Global Outcomes 2017 clinical practice guideline update. *Ann Intern Med*. 2018;168(6):422-430. doi:10.7326/M17-2640
6. Martin KJ, González EA. Metabolic bone disease in chronic kidney disease. *J Am Soc Nephrol*. 2007;18(3):875-885. doi:10.1681/ASN.2006070771
7. Ball AM, Gillen DL, Sherrard D, et al. Risk of hip fracture among dialysis and renal transplant recipients. *JAMA*. 2002;288(23):3014-3018. doi:10.1001/jama.288.23.3014
8. Vanent KN, Leasure AC, Acosta JN, et al. Association of chronic kidney disease with risk of intracerebral hemorrhage. *JAMA Neurol*. 2022;79(9):911-918. doi:10.1001/jamaneurol.2022.2299
9. Lee M, Saver JL, Chang KH, Liao HW, Chang SC, Ovbiagele B. Low glomerular filtration rate and risk of stroke: meta-analysis. *BMJ*. 2010;341:c4249. doi:10.1136/bmj.c4249
10. Masson P, Webster AC, Hong M, Turner R, Lindley RI, Craig JC. Chronic kidney disease and the risk of stroke: a systematic review and meta-analysis. *Nephrol Dial Transplant*. 2015;30(7):1162-1169. doi:10.1093/ndt/gfv009
11. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305. Published correction appears in *N Engl J Med*. 2008;18(4):4. doi:10.1056/NEJMoa041031
12. Meisinger C, Döring A, Löwel H; KORA Study Group. Chronic kidney disease and risk of incident myocardial infarction and all-cause and cardiovascular disease mortality in middle-aged men and women from the general population. *Eur Heart J*. 2006;27(10):1245-1250. doi:10.1093/eurheartj/ehi880
13. Choi Y, Jacobs DR Jr, Shroff GR, Kramer H, Chang AR, Duprez DA. Progression of chronic kidney disease risk categories and risk of cardiovascular disease and total mortality: coronary artery risk development in young adults cohort. *J Am Heart Assoc*. 2022;11(21):e026685. doi:10.1161/JAHA.122.026685
14. Saran R, Robinson B, Abbott KC, et al. US Renal Data System 2019 Annual Data Report: epidemiology of kidney disease in the United States. *Am J Kidney Dis*. 2020;75(1)(suppl 1):S1-S64.
15. Chu CD, McCulloch CE, Banerjee T, et al. CKD awareness among US adults by future risk of kidney failure. *Am J Kidney Dis*. 2020;76(2):174-183. doi:10.1053/j.ajkd.2020.01.007
16. Wang M, Peter SS, Chu CD, Tuot DS, Chen JH. Analysis of specialty nephrology care among patients with chronic kidney disease and high risk of disease progression. *JAMA Netw Open*. 2022;5(8):e2225797. doi:10.1001/jamanetworkopen.2022.25797
17. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney IntSuppl*. 2013;3:1-150.
18. Hemmelgarn BR, Clement F, Manns BJ, et al. Overview of the Alberta Kidney Disease Network. *BMC Nephrol*. 2009;10:30. doi:10.1186/1471-2369-10-30
19. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305(15):1553-1559. doi:10.1001/jama.2011.451
20. Tangri N, Grams ME, Levey AS, et al. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis. *JAMA*. 2016;315(2):164-174. Published correction appears in *JAMA*. 2016;315(8):822. doi:10.1001/jama.2015.18202
21. Ali I, Donne RL, Kalra PA. A validation study of the kidney failure risk equation in advanced chronic kidney disease according to disease aetiology with evaluation of discrimination, calibration and clinical utility. *BMC Nephrol*. 2021;22(1):194. doi:10.1186/s12882-021-02402-1
22. Hundemer GL, Tangri N, Sood MM, et al. Performance of the kidney failure risk equation by disease etiology in advanced CKD. *Clin J Am Soc Nephrol*. 2020;15(10):1424-1432. doi:10.2215/CJN.03940320
23. Grams ME, Sang Y, Ballew SH, et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int*. 2018;93(6):1442-1451. Published correction appears in *Kidney* Int. 2018;94(5):1025-1026.
24. Tangri N, Ferguson T, Komenda P. Pro: risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrol Dial Transplant*. 2017;32(5):748-751. doi:10.1093/ndt/gfx067
25. Vassalotti JA, Boucree SC. Integrating CKD into US primary care: bridging the knowledge and implementation gaps. *Kidney Int Rep*. 2022;7(3):389-396. doi:10.1016/j.ekir.2022.01.1066
26. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612. Published correction appears in *Ann Intern Med*. 2011;155(6):408. doi:10.7326/0003-4819-150-9-200905050-00006
27. Zoppini G, Targher G, Chonchol M, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. *Clin J Am Soc Nephrol*. 2012;7(3):401-408. doi:10.2215/CJN.07650711
28. Yokoyama H, Kanno S, Takahashi S, et al. Determinants of decline in glomerular filtration rate in nonproteinuric subjects

with or without diabetes and hypertension. *Clin J Am Soc Nephrol*. 2009;4(9):1432-1440. doi:10.2215/CJN.06511208
29. Chue CD, Edwards NC, Davis LJ, Steeds RP, Townend JN, Ferro CJ. Serum phosphate but not pulse wave velocity predicts decline in renal function in patients with early chronic kidney disease. *Nephrol Dial Transplant*. 2011;26(8):2576-2582. doi:10.1093/ndt/gfq787
30. Schumock GT, Andress D, E Marx S, Sterz R, Joyce AT, Kalantar-Zadeh K. Impact of secondary hyperparathyroidism on disease progression, healthcare resource utilization and costs in pre-dialysis CKD patients. *Curr Med Res Opin*. 2008;24(11):3037-3048. doi:10.1185/03007990802437943
31. Schumock GT, Andress DL, Marx SE, Sterz R, Joyce AT, Kalantar-Zadeh K. Association of secondary hyperparathyroidism with CKD progression, health care costs and survival in diabetic predialysis CKD patients. *Nephron Clin Pract*. 2009;113(1):c54-c61. doi:10.1159/000228076
32. Kronenberg F. HDL in CKD-the devil is in the detail. *J Am Soc Nephrol*. 2018;29(5):1356-1371. doi:10.1681/ASN.2017070798
33. Greene T, Teng CC, Inker LA, et al. Utility and validity of estimated GFR-based surrogate time-to-event end points in CKD: a simulation study. *Am J Kidney Dis*. 2014;64(6):867-879. doi:10.1053/j.ajkd.2014.08.019
34. Inker LA, Lambers Heerspink HJ, Mondal H, et al. GFR decline as an alternative end point to kidney failure in clinical trials: a meta-analysis of treatment effects from 37 randomized trials. *Am J Kidney Dis*. 2014;64(6):848-859. doi:10.1053/j.ajkd.2014.08.017
35. Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis*. 2014;64(6):821-835. doi:10.1053/j.ajkd.2014.07.030
36. Myers GL, Miller WG, Coresh J, et al. Recommendations for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. *Clin Chem*. 2006;52(1):5-18. doi:10.1373/clinchem.2005.0525144
37. Swartling O, Rydell H, Stendahl M, Segelmark M, Trolle Lagerros Y, Evans M. CKD Progression and Mortality Among Men and Women: A Nationwide Study in Sweden. *Am J Kidney Dis*. 2021;78(2):190-199.e1. doi:10.1053/j.ajkd.2020.11.026
38. Miao F, Cai YP, Zhang YX, Li Y, Zhang YT. Risk prediction of one-year mortality in patients with cardiac arrhythmias using random survival forest. *Comput Math Methods Med*. 2015;2015:303250. doi:10.1155/2015/303250
39. Zhou L, Wang H, Xu Q. Random rotation survival forest for high dimensional censored data. *Springerplus*. 2016;5(1):1425. doi:10.1186/s40064-016-3113-5
40. Wang H, Li G. A selective review on random survival forests for high dimensional data. *Quant Biosci*. 2017;36(2):85-96. doi:10.22283/qbs.2017.36.2.85
41. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). *Eur Heart J*. 2019;40(3):237-269. doi:10.1093/eurheartj/ehy462
42. Hallek M, Al-Sawaf O. Chronic lymphocytic leukemia: 2022 update on diagnostic and therapeutic procedures. *Am J Hematol*. 2021;96(12):1679-1705. doi:10.1002/ajh.26367
43. Nelson TJ, Javier-DesLoges J, Deka R, et al. Association of prostate-specific antigen velocity with clinical progression among African American and Non-Hispanic White men treated for low-risk prostate cancer with active surveillance. *JAMA Netw Open*. 2021;4(5):e219452. doi:10.1001/jamanetworkopen.2021.9452
44. Coresh J, Turin TC, Matsushita K, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA*. 2014;311(24):2518-2531. doi:10.1001/jama.2014.6634
45. Grams ME, Sang Y, Ballew SH, et al. Evaluating Glomerular Filtration Rate Slope as a Surrogate End Point for ESKD in Clinical Trials: An Individual Participant Meta-Analysis of Observational Data. *J Am Soc Nephrol*. 2019;30(9):1746-1755. doi:10.1681/ASN.2019010008
46. Inker LA, Heerspink HJL, Tighiouart H, et al. GFR slope as a surrogate end point for kidney disease progression in clinical trials: a meta-analysis of treatment effects of randomized controlled trials. *J Am Soc Nephrol*. 2019;30(9):1735-1745. doi:10.1681/ASN.2019010007
47. Greene T, Ying J, Vonesh EF, et al. Performance of GFR slope as a surrogate end point for kidney disease progression in clinical trials: a statistical simulation. *J Am Soc Nephrol*. 2019;30(9):1756-1769. doi:10.1681/ASN.2019010009
48. Chronic Kidney Disease Assessment and Diagnosis. American Society for Clinical Pathology. Accessed August 8, 2023. https://www.ascp.org/content/get-involved/institute-of-science-technology-policy/ckd-assessment-and-diagnosis#
49. Song Z, Lillehaugen T, Wallace J. Out-of-network laboratory test spending, utilization, and prices in the US. *JAMA*. 2021;325(16):1674-1676. doi:10.1001/jama.2021.0720
50. Shlipak MG, Tummalapalli SL, Boulware LE, et al. The case for early identification and intervention of chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) controversies conference. *Kidney Int*. 2021;99(1):34-47. doi:10.1016/j.kint.2020.10.012
51. Chan L, Nadkarni GN, Fleming F, et al. Derivation and validation of a machine learning risk score using biomarker and electronic patient data to predict progression of diabetic kidney disease. *Diabetologia*. 2021;64(7):1504-1515. doi:10.1007/s00125-021-05444-0
52. Ferguson T, Ravani P, Sood MM, et al. Development and external validation of a machine learning model for progression of CKD. *Kidney Int Rep*. 2022;7(8):1772-1781. doi:10.1016/j.ekir.2022.05.004
53. Grams ME, Brunskill NJ, Ballew SH, et al. Development and validation of prediction models of adverse kidney outcomes in the population with and without diabetes. *Diabetes Care*. 2022;45(9):2055-2063. doi:10.2337/dc22-0698
54. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. *JAMA*. 2000;283(15):2008-2012. doi:10.1001/jama.283.15.2008
55. Ooka T, Johno H, Nakamoto K, Yoda Y, Yokomichi H, Yamagata Z. Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: large-scale health check-up data in Japan. *BMJ Nutr Prev Health*. 2021;4(1):140-148. doi:10.1136/bmjnph-2020-000200
56. Boulware LE, Troll MU, Jaar BG, Myers DI, Powe NR. Identification and referral of patients with progressive CKD: a national study. *Am J Kidney Dis*. 2006;48(2):192-204. doi:10.1053/j.ajkd.2006.04.073


Can a machine learning model predict CKD progression in a diverse US population?
Kidney Medicine
Methods
Retrospective cohort study
110,264 adult patients
U.S. Laboratory Database
Machine learning models
Predictors
Laboratory characteristics
Demographic characteristics
Predictors
*eGFR slope
Urine ACR
Albumin (initial, slope)
Age
Sex
Initial eGFR
Outcome
CKD progression > 30% eGFR decline within 5 years
Outcome
7-variable risk classifier model
AUC 0.85
Accurate eGFR decline prediction
Conclusion: The progressive CKD classifier accurately predicts significant eGFR decline in patients with early, moderate, and advanced CKD using readily obtainable laboratory data. While prospective studies are warranted, our results support the clinical utility of the model to improve timely recognition and optimal management for patients at risk for CKD progression.
Reference: Aoki J, Kaya C, Khalid O, et al. CKD Progression prediction in a diverse US population: a machine learning model. Kidney Medicine, 2023.
VA by Susan Thanabalasingam, MD
@thana_susan