Singh K, et al., Fortune J Health Sci 2024
DOI:10.26502/fjhs.173

Kidney Disease Improving Global Outcomes (KDIGO) recommends that patients be referred to nephrology when their risk of ESKD exceeds 10-20% and points to existing tools, such as the Kidney Failure Risk Equation (KFRE),[2] to support this recommendation. The KFRE is a Cox regression model that predicts the time-to-ESKD in patients with known CKD. It has been validated and recalibrated in a number of international settings, and it has been shown to outperform nephrologists in estimating patients' risk.[1] Because it was trained with dialysis and kidney transplant as target outcomes, it is most useful in identifying patients requiring preparation for renal replacement therapy. However, as new therapies such as Sodium-Glucose Cotransporter-2 (SGLT2) inhibitors, become available and be effective earlier in the course of CKD, new tools are required that are able to differentiate risk sooner. The slope of estimated glomerular filtration rate (eGFR) has previously been demonstrated to be a useful clinical marker of progression and has the advantage of being calculable in earlier CKD stages. Predicting the eGFR slope may be more clinically relevant in early CKD,[3,4] but the rate of early CKD progression may also be driven by social determinants of health,[5–7] data for which are often not available in the electronic health record. In this study, we sought to evaluate whether CKD progression can be accurately predicted when social determinants of health are incorporated into prediction models. While social determinants were not directly measured, we used geospatial data in a national cohort of US Veterans to link patient-level data to census-tract-level social determinants of health information derived from publicly available data sources. We evaluated the ability of models to predict fast progression (as determined by eGFR slope) and time to ESKD.

Methods

Data source and study population

Our study used data from a national VA cohort drawing on data from 118 VA hospitals and their associated outpatient practices. We have complied with all relevant ethical regulations. The study was approved by the Institutional Review Boards of the VA Ann Arbor Healthcare System and the University of Michigan Medical School, and the need for informed consent was waived. We examined CKD progression among US Veterans who had known CKD in the VA health system. Data were obtained from 2006-2016, and the presence of CKD was defined based on any of the following: any measure of eGFR value < 60 mL/min/1.73m^2 using the CKD-EPI formula,[8] a urine albumin/creatinine ratio (ACR) >30mg/g of creatinine, or a CKD diagnosis by ICD-9 or ICD-10 codes. Patients were excluded if they had fewer than five eGFR values available over the period.

Outcomes

CKD progression was defined based on two separate outcomes: 1) rapid CKD progression based on the eGFR slope < -3.7 mL/min/1.73m^2 as a binary outcome, and 2) time-to-ESKD as a survival outcome. Veterans whose eGFR values were declining more steeply than -3.7 per year were considered "fast progressors" because this represented a slope that was 1 standard deviation more extreme than the mean eGFR slope. Of the overall cohort, 9.8% were considered fast progressors by this definition. The outcome of ESKD was identified by analysis of the linked VA data with the national ESKD Registry data, i.e., the US Renal Data System (USRDS).

Predictors

For each patient, patient-level variables were collected from the VA database along with social determinants of health obtained from public data sources linked at the census-tract and county level. Patient-level variables included demographic variables (age, sex, race, ethnicity), comorbidities, social history (smoking, alcohol use, and drug use), systolic and diastolic blood pressure values, laboratory data (serum creatinine based - baseline eGFR, baseline urine ACR, blood urea nitrogen, hematocrit, hemoglobin A1c, glucose), and exposure to potential and known nephrotoxins (including antibiotics, ACE inhibitors, antiretrovirals, non-steroidal anti-inflammatory medications, anti-angiogenesis medications, proton pump inhibitors, and lithium). Race and ethnicity were primarily included to understand their role in predicting CKD progression, and their relationship to social determinants of health. Combined with the social determinants of health predictors, the models considered 192 predictor variables.

Social determinants of health

Social determinants may play a role in the rate of CKD progression by affecting access to care, quality of care, and overall health. To examine their role, we included multiple variables related to social determinants of health. These included variables collected at the patient level relating to access to care (e.g., drive time and drive distance to the nearest primary, secondary, and tertiary care VA facilities), healthcare utilization (e.g., number of laboratory tests performed), and variables collected at either the census-tract or county levels. These were linked to public data sources by geocoding patient addresses to census tract and county levels (depending on the data source). Linked variables included neighborhood affluence, density of grocery stores and fast foods, whether a patient's neighborhood is considered a food desert, civic service density, environmental pollution (median $PM_{2.5}$ levels), and the percentage of nearby buildings constructed in each decade.

Model derivation and validation

We used separate tree ensemble models to predict the two outcomes: fast eGFR progression and time-to-ESKD.

Citation: Karandeep Singh, Jennifer Bragg-Gresham, Yun Han, Brenda Gillespie, William Weitzel, Susan Crowley, Rajiv Saran. Incorporating Social Determinants of Health in the Prediction of Chronic Kidney Disease Progression in a National Cohort of US Veterans. Fortune Journal of Health Sciences. 7 (2024): 151-157.

Singh K, et al., Fortune J Health Sci 2024
DOI:10.26502/fjhs.173

Gradient-boosted decision trees (GBDTs) were used for fast progression, and random survival forests were used for time-to-ESKD (details in Supplementary Methods).

For each of the models, patient data were randomly divided into training, tuning, and test cohorts. The tuning cohort was used to decide on early stopping for the GBDT models. For both models, tree ensemble models were trained on the training data and evaluated in the testing data.

Variable importance

For the GBDT model, variable importance was calculated for each variable using equation 45 from Friedman 2001 as implemented in the *h2o* R package.[9] For the random survival forest model, variable importance was calculated using the sum of test statistics impurity measure as implemented in the *ranger* R package.[10] Relationships between individual variables and outcomes were explored using partial dependence plots and Shapley summary plots.

Handling of missing data

Missing values were mean-imputed, and dummy variables were created to indicate missingness in the original data. This is because while the *h2o* R package tree ensemble implementation supports missing values, the *ranger* R package random survival forest implementation does not.

Results

Among patients with known CKD, 3,237,113 patients were identified who had available laboratory data to calculate eGFR slope. After excluding extreme values of slope (likely due to lab errors) and restricting to patients with more than 5 labs available to calculate eGFR slope over a maximum follow-up period of 10 years, 1,550,526 patients remained, of which 930,615 patients were randomly assigned to a training cohort, 309,831 to a tuning cohort, and 310,044 to a testing cohort. Baseline characteristics of the patient population, stratified by rapid progression, are provided in Table 1 and Supplementary Table 1.

Rapid eGFR Progression

The tree ensemble model had a C-statistic of 0.79 in the testing cohort in predicting rapid eGFR progression. The most important 20 variables are shown in Table 2.

Baseline eGFR was the most important contributor to the model identifying fast progressors, followed by age and the number of laboratory tests performed. Among the social determinants of health, the most important predictors were median $PM_{2.5}$ (a measure of environmental pollution), whether the patient's neighborhood was a food desert, the percentage of nearby buildings built in the 1970s, neighborhood affluence, the density of grocery stores, and civic service density.

Table 1: Comparisons of Means/Percentages of Characteristics in Slow vs. Fast Progressors.

Variable	Slow Progressors	Fast Progressors	p-value
Demographics			
Age (years)	66.6	62.8	<0.0001
Male	95.8%	95.9%	0.0049
White Race	60.7%	58.4%	<0.0001
Black race	11.3%	19.8%	<0.0001
Asian Race	0.32%	0.54%	<0.0001
American Indian/AK Native	0.37%	0.60%	<0.0001
Pacific Islander/HI Native	0.55%	0.64%	<0.0002
Race unknown	22.8%	14.0%	<0.0001
Hispanic ethnicity	4.0%	6.0%	<0.0001
Comorbidities			
Anemia	55.7%	60.5%	<0.0001
Cardiovascular Disease	17.6%	20.9%	<0.0001
Diabetes	35.8%	63.8%	<0.0001
Hypertension	49.9%	59.2%	<0.0001

Citation: Karandeep Singh, Jennifer Bragg-Gresham, Yun Han, Brenda Gillespie, William Weitzel, Susan Crowley, Rajiv Saran. Incorporating Social Determinants of Health in the Prediction of Chronic Kidney Disease Progression in a National Cohort of US Veterans. Fortune Journal of Health Sciences. 7 (2024): 151-157.

Singh K, et al., Fortune J Health Sci 2024
DOI:10.26502/fjhs.173

Table 2: Variable importance for predicting "fast progressors" among patients with known CKD.

Variable	Contribution to model (%)
Baseline eGFR	6.60%
Age	4.50%
Number of laboratory tests performed	4.00%
Systolic blood pressure	3.70%
BMI	3.40%
BUN	3.30%
Diastolic blood pressure	3.20%
Serum Albumin	3.20%
Glucose	3.20%
Baseline UACR	2.80%
Hemoglobin	2.50%
Median PM2.5	2.40%
Hematocrit	2.30%
Food deserts	2.30%
HbA1c	2.20%
% Building built 70-79	2.10%
Neighborhood affluence	2.10%
Grocery density	2.00%
Civic service density	2.00%

Patients' risk of rapid progression increases as their baseline eGFR increases based on a partial dependence plot (Figure 1). This finding is similar in the Shapley summary plot (Figure 2), which shows that the highest risk of rapid progression (towards the right side of the plot) generally occurs when the baseline eGFR values are high (depicted as red dots).

Predicting time-to-ESKD

The tree ensemble model had a C-statistic of 0.90 in the testing cohort in predicting time to ESRD. The most important 20 variables are shown in Table 3.

The most important variable in predicting time-to-ESKD was baseline eGFR, followed by the most recent outpatient blood urea nitrogen level and whether the patient had previously seen a nephrologist. The only social determinant of health among the top 20 variables was the presence of a nephrology visit, suggesting that social determinants might play a lesser role in predicting the time to ESKD, although this cannot be stated with certainty.

Discussion

In this study of US Veterans using clinical factors and social determinants of health in the prediction of CKD progression, we found that tree ensemble models can accurately predict the progression of CKD, with C-statistics of 0.79 for fast progression (earlier in the CKD continuum) and 0.90 for time to ESKD (later in the course of CKD). While baseline eGFR is the most important variable in predicting both outcomes, its relationship to the two outcomes is different. While a lower baseline eGFR indicates a higher risk of progression of ESKD, we found that it appears

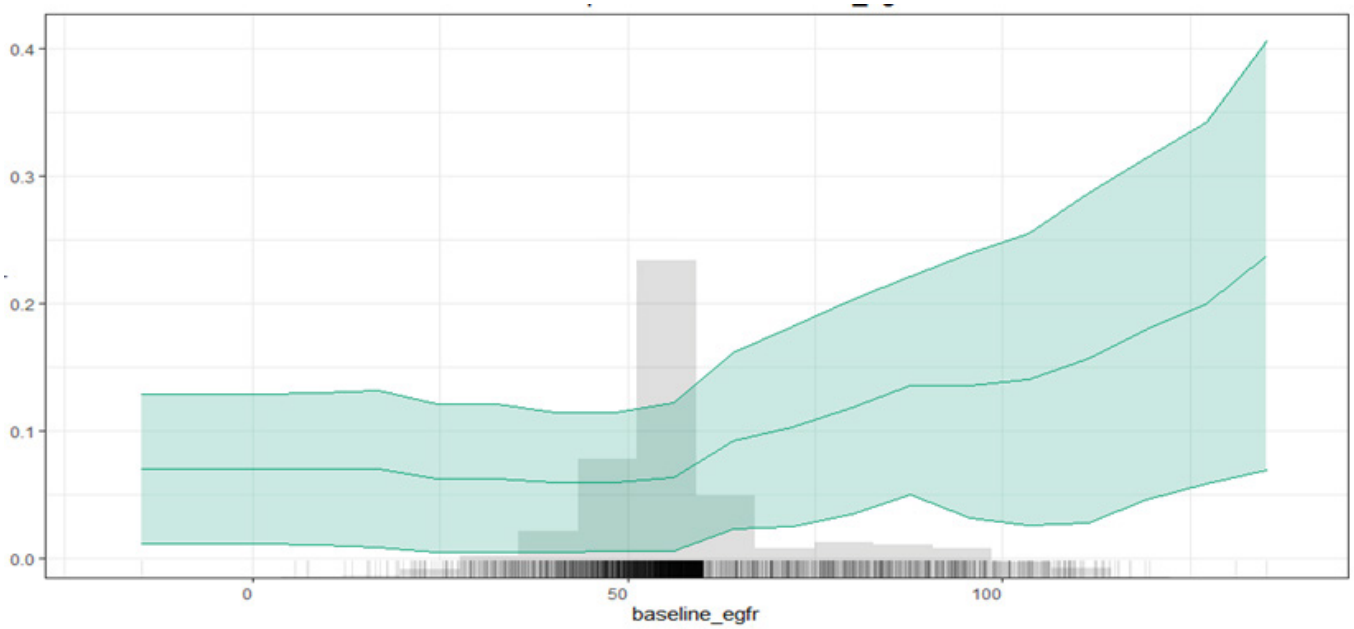

Figure 1: Partial dependence plot showing relationship between baseline eGFR and rapid CKD progression

Citation: Karandeep Singh, Jennifer Bragg-Gresham, Yun Han, Brenda Gillespie, William Weitzel, Susan Crowley, Rajiv Saran. Incorporating Social Determinants of Health in the Prediction of Chronic Kidney Disease Progression in a National Cohort of US Veterans. Fortune Journal of Health Sciences. 7 (2024): 151-157.

Singh K, et al., Fortune J Health Sci 2024
DOI:10.26502/fjhs.173

Figure 2: Shapley summary plot dependence plot showing relationship between baseline eGFR and rapid CKD progression

Table 3: Variable importance for predicting time-to-ESKD among patients with known CKD.

Variable	Relative importance (out of 100%)
Baseline eGFR	11%
Outpatient blood urea nitrogen lab	7.90%
Nephrology visit (binary)	4.80%
Diagnosis of chronic kidney disease	3.30%
Baseline urinary albumin/creatinine ratio	2.50%
Outpatient hematology lab	2.20%
Outpatient blood urea nitrogen (different source)	2.10%
Outpatient hematocrit lab	1.90%
Systolic blood pressure (mean)	1.80%
Age	1.70%
Systolic blood pressure in the prior year	1.70%
Systolic blood pressure 2 years ago	1.70%
Race	1.50%
Outpatient hemoglobin A1c	1.40%
Outpatient glucose	1.40%
Diastolic blood pressure 2 years ago	1.10%
Diastolic blood pressure in the past year	1.10%
Diastolic blood pressure (mean)	1.10%
Body mass index	1.00%
Any evidence of CKD (narrow definition)	0.98%

'protective' against a rapid decline in eGFR based on slope. Although counterintuitive, we speculate that this might be related to high values of eGFR (>100) represent glomerular hyperfiltration which is a far less pronounced in advanced stages of CKD. When early values of eGFR are high, future values of eGFR will commonly be lower both due to the fact that glomerular hyperfiltration often predicts subsequent kidney damage with progressive decline in eGFR. Beyond traditional risk factors for CKD progression, we also found social determinants of health to be important contributors to predicting risk based on tree ensemble variable importance. Important social determinants in predicting fast progressors included the number of laboratory tests performed, median $PM_{2.5}$, food deserts, the percentage of nearby buildings built in the 1970s, neighborhood affluence, grocery density, and civic service density. Social determinants were less relevant to the prediction of EKSD onset, where only the presence of a nephrology visit was considered in the top 20 important variables. This finding is interesting because it suggests that social determinants are more relevant earlier in the course of CKD, which is where eGFR slope is more relevant. This may be because earlier eGFR decline is more modifiable, and changing the diet (i.e., healthy food) may have more of an impact in early CKD. The number of laboratory tests may be indicative of the clinical need to monitor kidney function more frequently as well as greater eGFR variability and may not necessarily be reflective only of social factors. However, the ability to undergo multiple laboratory tests is suggestive of having both transportation to a laboratory testing facility and sufficient healthcare coverage to afford multiple tests.

Our observation that social factors were less relevant to

Citation: Karandeep Singh, Jennifer Bragg-Gresham, Yun Han, Brenda Gillespie, William Weitzel, Susan Crowley, Rajiv Saran. Incorporating Social Determinants of Health in the Prediction of Chronic Kidney Disease Progression in a National Cohort of US Veterans. Fortune Journal of Health Sciences. 7 (2024): 151-157.

predicting ESKD onset is not surprising. Even in the kidney failure risk equation (KFRE), a 4-variable model achieved a C-statistic of 0.91,[2] suggesting that demographic and biological risk factors capture a substantial degree of the risk of ESKD onset. Our model, which considered a much larger number of predictors, achieved a similar C-statistic of 0.90, suggesting that social determinants may play less of a role, although their role in ensuring progression in earlier stages of the disease is not excluded, and may be confounded by the advancing nature of the disease in later stages. The patients identified as high risk by the KFRE are generally those with lower eGFRs and thus are closer to the endpoint of renal replacement therapy. Thus, it may be too late for high-risk patients, i.e., those closer to the endpoint of renal replacement therapy, for social determinants to substantially appear to affect the risk, as they have already caused the most damage in earlier stages of the disease. Of the social factors we identified, having seen a nephrologist was the most important one. This is because the decision to start renal replacement therapy is typically made by a nephrologist and thus may represent an indicator that CKD has been clinically recognized, and that the patient has sufficient healthcare coverage to visit a nephrologist. This may also represent a case of 'confounding by indication' for ESKD treatment.

Our study has limitations. While our study was conducted in a large national cohort of US Veterans, factors that affect access to care in Veterans may be different from those affecting other populations, especially those relating to healthcare coverage. For example, the barriers to see a nephrologist may be lower or different in VA care settings as compared to other care settings, which could have affected our findings. While social determinants of health are generally well-captured by census tracts, there may be heterogeneity among social factors even within census tracts, which we did not capture as we did not measure social determinants at the patient level. Despite these limitations, our finding of differing risk factors of fast progression and ESKD onset suggest that interventions to consider in these groups may be different. Because social factors are more important in fast progression—an outcome that occurs earlier in the course of CKD—interventions on social factors may potentially modify risk in early CKD progression.

Disclosures

This material is based upon work supported (or supported in part) by the Department of Veterans Affairs, Office of Connected Health; VHA Department of Innovation Contract # 36C10B18C2768. The views expressed in this work are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. RS was funded as PI for this work through VHA Department of Innovation. He is Project Director for the CDC's Kidney Disease Surveillance System (KDSS), funded as a Cooperative Agreement with the CDC. He also receives funding from the Patient Centered Outcomes Research Institute (PCORI), the Michigan Department of Health and Human Services (MDHHS) and from the Michigan-wide Collaborative for Quality Improvement funded by Blue Cross Blue Shield of Michigan. He is on the advisory board of the National Kidney Foundation of Michigan. JBG and BWG are Co-PIs on the KDSS. KS receives grant funding from the National Institute of Diabetes and Digestive and Kidney Diseases, Blue Cross Blue Shield of Michigan, and Teva Pharmaceuticals for unrelated work. KS serves on a scientific advisory board for Flatiron Health.

Acknowledgements

(i) We are grateful to the following members of the VHA Advisory Committee: Susan Crowley, MD, (VA Connecticut Healthcare System), William (Rick) Weitzel, MD (VA Ann Abor Health System), Susan Wong, MD (Seattle VA), Ziyad Al-Aly, MD (St Louis VA), Rudolph Rodriguez, MD (Seattle VA), John Hotchkiss, MD (Pittsburgh VA), Matthew Vincenti, PhD (White River Junction, VT), Csaba Kovesdy, MD (Memphis VA), Charuhas Thakar, MD (Cincinnati VA), and (ii) Project Managers: Nirmala Rajaram, PhD (VA Ann Arbor Project Manager) and April Wyncott, MPH, MBA.

References

1. Potok OA, Nguyen HA, Abdelmalek JA, et al. Patients,' Nephrologists,' and Predicted Estimations of ESKD Risk Compared with 2-Year Incidence of ESKD. Clin J Am Soc Nephrol 14 (2019): 206-212.

2. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. JAMA 305 (2011): 1553-1559.

3. 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 30 (2019): 1735.

4. 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 30 (2019): 1746.

5. Norton JM, Moxey-Mims MM, Eggers PW, et al. Social Determinants of Racial Disparities in CKD. J Am Soc Nephrol 27 (2016): 2576.

6. Social Determinants of CKD Hotspots. Semin Nephrol 39 (2019): 256-262.

7. Ozieh MN, Garacci E, Walker RJ, et al. The cumulative

Singh K, et al., Fortune J Health Sci 2024
DOI:10.26502/fjhs.173

impact of social determinants of health factors on mortality in adults with diabetes and chronic kidney disease. BMC Nephrol 22 (2021): 1-10.

8. Inker LA, Schmid CH, Tighiouart H, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. N Engl J Med 367 (2012).
9. Friedman JH. Greedy function approximation: A gradient boosting machine. aos 29 (2001): 1189-1232.
10. ranger: A Fast Implementation of Random Forests. Comprehensive R Archive Network (CRAN).

Citation: Karandeep Singh, Jennifer Bragg-Gresham, Yun Han, Brenda Gillespie, William Weitzel, Susan Crowley, Rajiv Saran. Incorporating Social Determinants of Health in the Prediction of Chronic Kidney Disease Progression in a National Cohort of US Veterans. Fortune Journal of Health Sciences. 7 (2024): 151-157.