

Original Paper

# Electronic Phenotype for Advanced Chronic Kidney Disease in a Veteran Health Care System Clinical Database: Systems-Based Strategy for Model Development and Evaluation

Gajapathiraju Chamarthi<sup>1</sup>, MD; Tatiana Orozco<sup>2</sup>, PhD; Popy Shell<sup>2</sup>, MPH; Devin Fu<sup>2</sup>, MPH; Jennifer Hale-Gallardo<sup>2</sup>, PhD; Huangang Jia<sup>2</sup>, PhD; Ashutosh M Shukla<sup>1,2</sup>, MD

<sup>1</sup>Division of Nephrology, Hypertension and Transplantation, University of Florida, Gainesville, FL, United States

<sup>2</sup>Advanced Chronic Kidney Disease and Home Dialysis Program, North Florida/South Georgia Veteran Healthcare System, Gainesville, FL, United States

**Corresponding Author:**

Ashutosh M Shukla, MD

Advanced Chronic Kidney Disease and Home Dialysis Program

North Florida/South Georgia Veteran Healthcare System

1600 Archer Road

Gainesville, FL, 32610

United States

Phone: 1 352 273 8821

Fax: 1 352 392 5465

Email: [Ashutosh.Shukla@medicine.ufl.edu](mailto:Ashutosh.Shukla@medicine.ufl.edu)

## Abstract

**Background:** Identifying advanced (stages 4 and 5) chronic kidney disease (CKD) cohorts in clinical databases is complicated and often unreliable. Accurately identifying these patients can allow targeting this population for their specialized clinical and research needs.

**Objective:** This study was conducted as a system-based strategy to identify all prevalent Veterans with advanced CKD for subsequent enrollment in a clinical trial. We aimed to examine the prevalence and accuracy of conventionally used diagnosis codes and estimated glomerular filtration rate (eGFR)-based phenotypes for advanced CKD in an electronic health record (EHR) database. We sought to develop a pragmatic EHR phenotype capable of improving the real-time identification of advanced CKD cohorts in a regional Veterans health care system.

**Methods:** Using the Veterans Affairs Informatics and Computing Infrastructure services, we extracted the source cohort of Veterans with advanced CKD based on a combination of the latest eGFR value  $\leq 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  or existing International Classification of Diseases (ICD)-10 diagnosis codes for advanced CKD (N18.4 and N18.5) in the last 12 months. We estimated the prevalence of advanced CKD using various prior published EHR phenotypes (ie, advanced CKD diagnosis codes, using the latest single eGFR  $< 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , utilizing two eGFR values) and our operational EHR phenotypes of a high-, intermediate-, and low-risk advanced CKD cohort. We evaluated the accuracy of these phenotypes by examining the likelihood of a sustained reduction of eGFR  $< 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  over a 6-month follow-up period.

**Results:** Of the 133,756 active Veteran enrollees at North Florida/South Georgia Veterans Health System (NF/SG VHS), we identified a source cohort of 1759 Veterans with advanced nondialysis CKD. Among these, 1102 (62.9%) Veterans had diagnosis codes for advanced CKD; 1391 (79.1%) had the index eGFR  $< 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ ; and 928 (52.7%), 480 (27.2%), and 315 (17.9%) Veterans had high-, intermediate-, and low-risk advanced CKD, respectively. The prevalence of advanced CKD among Veterans at NF/SG VHS varied between 1% and 1.5% depending on the EHR phenotype. At the 6-month follow-up, the probability of Veterans remaining in the advanced CKD stage was 65.3% in the group defined by the ICD-10 codes and 90% in the groups defined by eGFR values. Based on our phenotype, 94.2% of high-risk, 71% of intermediate-risk, and 16.1% of low-risk groups remained in the advanced CKD category.

**Conclusions:** While the prevalence of advanced CKD has limited variation between different EHR phenotypes, the accuracy can be improved by utilizing two eGFR values in a stratified manner. We report the development of a pragmatic EHR-based

model to identify advanced CKD within a regional Veterans health care system in real time with a tiered approach that allows targeting the needs of the groups at risk of progression to end-stage kidney disease.

(*Interact J Med Res* 2023;12:e43384) doi: [10.2196/43384](https://doi.org/10.2196/43384)

## KEYWORDS

advanced chronic kidney disease; EHR phenotype; Veteran Health System; CKD cohort; kidney disease; chronic; clinical; database; data; diagnosis; risk; disease

## Introduction

Advanced chronic kidney disease (CKD) progressing to end-stage kidney disease (ESKD) is a huge burden for the US health care system [1]. Patients with advanced CKD are at increased risk for adverse outcomes, including progression to ESKD and death. Prior studies show that providing pre-ESKD nephrology care and comprehensive pre-ESKD education improves clinical outcomes; reduces health care costs; and increases home dialysis, transplantation utilization, and patient survival [2-6]. Despite these positive outcomes, approximately 40% of patients with incident ESKD in the United States have either limited (less than 6 months) or no access to nephrology care before initiating dialysis and even fewer (<1%) receive kidney disease education services [7,8]. Accurately identifying the advanced (stages 4 and 5) CKD population at risk for ESKD can facilitate targeted needs assessment studies to improve pre-ESKD nephrology care and provide comprehensive pre-ESKD education for this high-risk population [9].

Clinically, CKD is diagnosed by sustained alterations in the structure or function of the kidney for more than 3 months with implications for health. The Kidney Disease: Improving Global Outcomes (KDIGO) Work Group recommends staging CKD based on cause, estimated glomerular filtration rate (eGFR), and albuminuria [10]. Unfortunately, the asymptomatic nature of CKD creates a lack of awareness for patients and providers alike [1,11]. Investigators conventionally use the International Classification of Diseases (ICD)-based diagnosis codes or electronic health record (EHR)-based phenotypes according to the eGFR to identify patients with CKD in clinical databases [12]. These phenotypes recommend using two eGFR values below  $60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , obtained more than 90 days apart, to identify a population with CKD of stage 3 or higher in the databases [12]. However, similar guidance is not available to identify an advanced CKD population within clinical databases, and epidemiological investigations frequently use a single latest eGFR value while ascertaining the advanced CKD burden within the database [3,13,14]. Considering the variability in the frequency of measurement, pragmatic fluctuations in the serum creatinine value and concerns for intervening acute kidney injury (AKI) episodes can cause errors in classifying one's CKD stage [15]. Thus, there is a need to establish an optimal EHR-based method capable of identifying patients with advanced CKD within clinical databases in real time to improve kidney disease care and research.

Using the clinical database of the North Florida/ South Georgia (NF/SG) Veterans Health System (VHS), we sought to assess the burden of advanced CKD prevalence in real time using various EHR-recorded advanced CKD phenotypes within the

Veterans Health Administration (VHA) [14,16]. We further examined the accuracy of different EHR phenotypes for advanced CKD by prospectively following the cohorts for 6 months and assessed the number of Veterans remaining in the advanced CKD stage after the initial classification. Furthermore, considering the lack of consensus on EHR phenotyping for identifying an advanced CKD cohort within clinical databases, we also sought to explore a new tiered pragmatic method for estimating the Veteran cohort with advanced CKD in real time.

## Methods

### Data Source and Cohort Selection

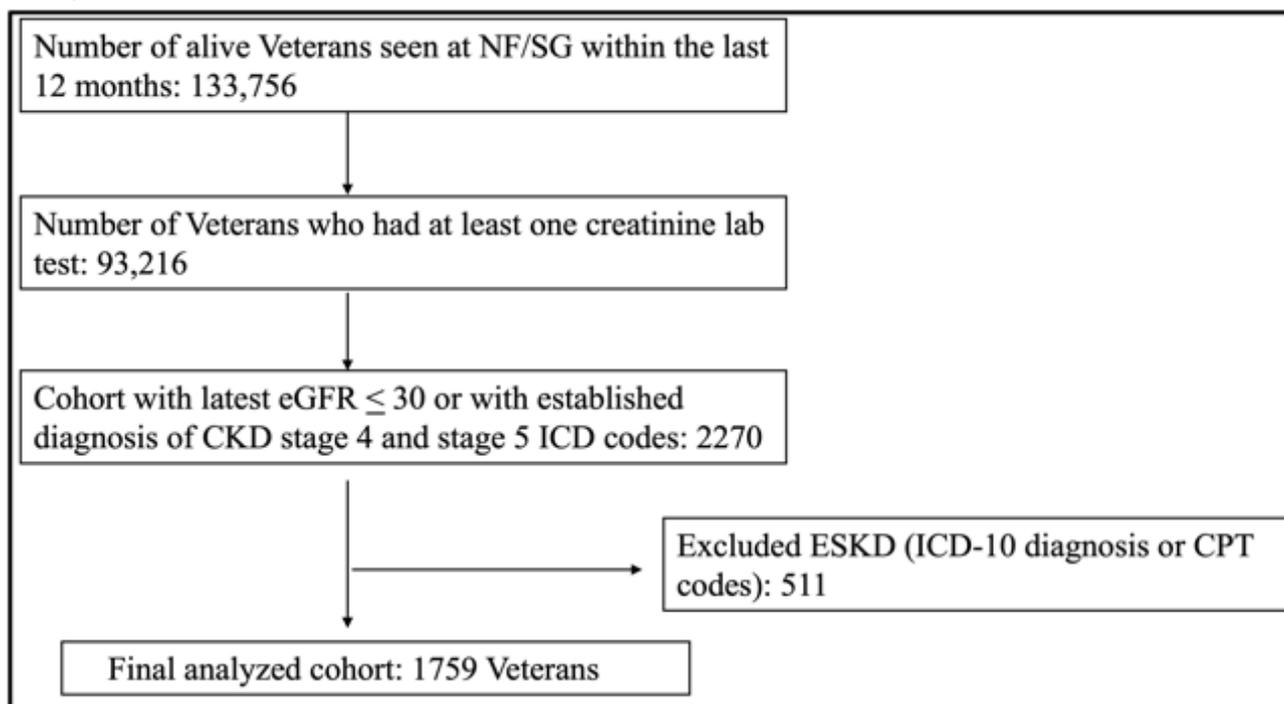
This study was conducted as a system-based strategy to identify all prevalent Veterans with advanced (stages 4 and 5) nondialysis CKD. The identified participants were then approached for enrollment in the Trial to Evaluate and Assess the effects of Comprehensive pre-ESKD education on Home dialysis among Veterans (TEACH-VET), which aims to assess the impact of a universal approach for comprehensive pre-ESKD education for all patients with advanced CKD on various clinical, patient-reported, and health services outcomes [17]. We used the Veterans Affairs (VA) Corporate Data Warehouse (CDW) and VA Informatics and Computing Infrastructure (VINCI) to identify the advanced CKD cohort. In brief, the VINCI services initially queried the VA CDW in April 2021 to identify all Veterans registered for service at NF/SG VHS during the 12 months prior to the data extraction (source cohort). The Veterans with an active laboratory value of creatinine were identified and their eGFR was calculated by applying the Modification of Diet in Renal Disease (MDRD) equation [18]. The use of the MDRD equation was determined by the then-prevalent method of eGFR estimation for the VINCI services. We then created a source cohort of Veterans with advanced CKD who either had the latest eGFR value  $\leq 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  (index eGFR) or an existing ICD-10 diagnosis code for advanced CKD (ICD-10 codes: N18.4 and N18.5) within the last 12 months (Figure 1). Patients on dialysis were excluded using the ICD-10 and Current Procedural Terminology (CPT) codes for dialysis (see Table S1 in Multimedia Appendix 1). The prevalence of advanced CKD was estimated in real time using various methods, including advanced CKD diagnosis codes or by eGFR phenotypes described in the literature (ie, by ICD-10 advanced CKD diagnosis codes, by using single [index] eGFR  $< 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , and by using the two eGFR values 90 days apart with the index eGFR  $< 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  and 90-day prior eGFR  $< 60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ ) [14,16,19]. The cumulative prevalence of CKD was calculated by combining the data extracted over 6 months. Patient-level data

included age, sex, race, ethnicity, religion, marital status, Veteran era, and residential zip codes used for defining the rurality by applying Rural-Urban Commuting Area codes. Statistical analyses were performed using R software version 4.0.4 (R Core Team, 2021) [20].

The source cohort (ie, April 2021 cohort) was divided into a high-, intermediate-, and low-risk advanced CKD cohort utilizing the latest (index) eGFR and 90-day prior eGFR and diagnostic codes (Table 1). Patients with both eGFR values below  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  were considered to have a high risk of advanced CKD, whereas those with one of the two eGFR values less than  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  but with the other value

$\geq 30$  but  $< 60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  were considered to have an intermediate risk of having advanced CKD. The intermediate-risk cohort with an index eGFR below  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  was further refined by excluding patients diagnosed with AKI within the 90 days prior to their latest eGFR values using ICD-10 codes. Veterans with both eGFR values  $\geq 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  but with diagnosis codes for advanced CKD were regarded as having a low risk of advanced CKD (Table 1). The source cohort was followed prospectively for 6 consecutive months until September 2021 using similar queries to examine the eGFR laboratory behavior of the patients with advanced CKD.

**Figure 1.** Selection of an advanced nondialysis chronic kidney disease (CKD) cohort at North Florida/South Georgia (NF/SG) Veterans Health System. CPT: Current Procedural Terminology; eGFR: estimated glomerular filtration rate; ESKD: end-stage kidney disease; ICD-10: International Classification of Diseases, Tenth Revision.



**Table 1.** Defining parameters for identifying cohorts at high, intermediate, and low risk of advanced chronic kidney disease (CKD).

Cohort	Index eGFR <sup>a</sup> ( $\text{ml}\cdot\text{min}^{-1}\cdot\text{m}^{-2}$ )	$\geq 90$ days prior eGFR ( $\text{ml}\cdot\text{min}^{-1}\cdot\text{m}^{-2}$ )	Additional criteria
High-risk advanced CKD	$< 30$	$< 30$	None
<b>Intermediate-risk advanced CKD</b>			
Subgroup 1	$< 30$	$\geq 30$ and $< 60$	Excluding AKI <sup>b</sup> using ICD-10 <sup>c</sup> codes (N17)
Subgroup 2	$\geq 30$ and $< 60$	$< 30$	Patients have ICD-10 codes for stage 4 and 5 CKD (N18.4 and N18.5)
Low-risk advanced CKD	$\geq 30$	$\geq 30$	Patients have ICD-10 codes for stage 4 and 5 CKD (N18.4 and N18.5)

<sup>a</sup>eGFR: estimated glomerular filtration rate; index eGFR refers to the latest eGFR at the time of extraction of the cohort.

<sup>b</sup>AKI: acute kidney injury.

<sup>c</sup>ICD-10: International Classification of Diseases, Tenth Revision.

## Outcomes

The primary goal of this study was to assess the prevalence and accuracy of various EHR phenotypes for extraction of an advanced CKD cohort in a clinical database utilizing diagnosis codes and eGFR models (ie, by ICD-10 advanced CKD diagnosis codes, by using single latest [index] eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , and by using the two eGFR values 90 days apart, with the index eGFR  $<30$  and 90 days prior eGFR  $<60$ ) and our tiered EHR phenotype (high, intermediate, and low risk). Considering that nearly one-third of Veterans do not regularly obtain laboratory testing from within the VA, the denominator population for estimating the prevalence of advanced CKD was judged by only including the Veterans with a valid creatinine value measured over the prior 12 months. Considering EHR phenotypes as a standard for identification of patients with advanced CKD, cross-sectional accuracy for identifying patients with advanced CKD using only ICD-10 codes was assessed by comparison with laboratory-based eGFR EHR phenotypes, analyzed by calculating the sensitivity and positive predictive value (PPV). A manual chart review was conducted in a small randomly selected sample to identify errors related to automated advanced nondialysis CKD identification. Prospective accuracy of all EHR phenotypes, including our pragmatic tiered approach of high-, intermediate-, and low-risk advanced CKD cohorts, was assessed by ascertaining the longitudinal follow-up of laboratory values and identifying the likelihood of remaining in the advanced CKD stage at the end of the 6-month follow-up.

## Ethical Approval

The regulatory approvals for the study were obtained from the institutional review board of the University of Florida (201900870). The study data are stored in secured systems at NF/SG VHS as per the institutional guidelines.

## Results

We identified 133,756 active enrollees with 93,216 enrollees having at least one value of measured creatinine during an outpatient or inpatient visit at NF/SG VHS in the prior 12 months. After excluding the Veterans with ESKD by additional ICD and CPT codes, a source cohort of 1759 Veterans was identified as either having the latest eGFR  $\leq 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  or an existing ICD-10 diagnosis code for advanced CKD (ICD-10 codes N18.4 and N18.5) within the last 12 months (Figure 1). The overall cohort had a mean age of 75 (SD 11.1) years and consisted of a predominantly male (95.8%) and white (67.8%) population. These Veterans lived approximately 126.3 (SD 229.5) miles from the nephrology service-providing VA center, with rural Veterans constituting a significant proportion (751/1759, 42.7%) of the cohort (Table 2). A manual chart review was performed on 116 records and 13 Veterans with ESKD were identified, yielding an 11.2% error rate for advanced nondialysis CKD identification.

Of the total cohort of 1759 Veterans, only 1102 (62.9%) had diagnosis codes for advanced CKD, whereas 1391 (79.1%) had

the latest (index) eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ . Incorporating two eGFR values where the latest eGFR was  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  and the 90-day prior eGFR was  $<60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , we found 1346 Veterans to have advanced CKD. We then categorized 928 (52.7%) as high risk, 480 (27.2%) as intermediate risk, and 315 (17.9%) as low risk of advanced CKD based on the definitions described above (Tables 1 and 2). The mean eGFR for the initial advanced CKD cohort was 26.2 (SD 12.1)  $\text{ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ . The mean eGFR was 27.7  $\text{ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  in the ICD codes group, while the mean eGFR in the latest (index) eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  group was 22  $\text{ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ . The mean eGFR was 20.3 (SD 6.6), 27.4 (SD 5.6), and 42.1 (SD 16.6)  $\text{ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  for the high-, intermediate-, and low-risk advanced CKD groups in the initial source cohort (Table 2). The prevalence of advanced CKD among Veterans at NF/SG VHS varied between 1% and 1.5% based on the phenotype for advanced CKD. Based on our definitions, the prevalence of advanced (high- and intermediate-risk) CKD at NF/SG VHS was approximately 1.5% (Table 3). The cumulative cohort over the 6 months yielded 1840 Veterans with high and intermediate risk (2% cumulative prevalence). The sensitivity of diagnosis codes was only 55%-65% compared to the eGFR phenotypes, and the PPV of ICD-10 diagnosis codes for advanced CKD varied between 55% and 74% (Table 4).

The source cohort was followed prospectively for 6 months to examine the variations and likelihood of a sustained reduced eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  across various EHR phenotypes. A total of 981 (55.8%) of the 1759 Veterans had at least one subsequent eGFR measurement in the initial April cohort (Table 5). The probability of any subsequent eGFR measurement above  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  after the index eGFR in the cohort defined by ICD codes was 38.3%, and was approximately 12.7% and 12.8% in cohorts defined by index eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  and two eGFR phenotypes with index eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  and 90-day prior eGFR  $<60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$ , respectively. Similarly, the probability of having any subsequent eGFR value above  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  after the index eGFR measurement was 7.1%, 35.7%, and 90% in the high-, intermediate-, and low-risk group, respectively. The probability of Veterans remaining in an advanced CKD stage (stages 4 and 5) noted by the recent eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  at the end of follow-up was 65.3% in the group identified by the ICD codes, whereas the probability improved to 90% in the group defined by single (index) eGFR  $<30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  and the group defined by the index eGFR and 90-day prior eGFR method. Similarly, the probability of Veterans remaining in an advanced CKD stage at the end of the follow-up period was 94.2%, 71.0%, and 16.1% for high-, intermediate-, and low-risk groups, respectively (Figure 2, Table 5, and Table S2 in Multimedia Appendix 1).

**Table 2.** Demographic data for the source cohort.

Characteristics	Total cohort (N=1759)	ICD-10 <sup>a</sup> code N18.4 or N18.5 (n=1102)	Index eGFR <sup>b</sup> <30 (n=1391)	Index eGFR <30 and 90 days prior eGFR <60	High-risk advanced CKD <sup>c</sup> (n=928)	Intermediate-risk advanced CKD (n=480)	Low-risk advanced CKD (n=315)
eGFR, mean (SD)	26.2 (12.1)	27.7 (13.7)	22.0 (6.5)	22.0 (6.5)	20.3 (6.6)	27.4 (5.6)	42.1 (16.6)
Age (years), mean (SD)	75.3 (11.1)	75.5 (10.9)	75.0 (11.0)	75.2 (10.8)	75.3 (11.0)	75.2 (10.3)	75.5 (12.2)
Sex (male), n (%)	1686 (95.8)	1057 (95.9)	1334 (95.9)	1290 (95.8)	889 (95.8)	461 (96.0)	301 (95.6)
<b>Race, n (%)</b>							
Black	386 (21.9)	232 (21.1)	311 (22.4)	299 (22.2)	216 (23.3)	94 (19.6)	69 (21.9)
White	1192 (67.8)	758 (68.8)	936 (67.3)	911 (67.7)	616 (66.4)	339 (70.6)	212 (67.3)
Other or unknown	181 (10.3)	112 (10.2)	144 (10.4)	136 (10.1)	96 (10.3)	47 (9.8)	34 (10.8)
Hispanic ethnicity, n (%)	35 (2.0)	23 (2.1)	31 (2.2)	30 (2.2)	23 (2.5)	9 (1.9)	2 (0.6)
Rural, n (%)	751 (42.7)	475 (43.1)	574 (41.3)	558 (41.5)	378 (40.7)	200 (41.7)	154 (48.9)
Married, n (%)	1087 (61.9)	661 (60.1)	875 (63.0)	846 (63.0)	574 (62.0)	311 (64.9)	179 (57.0)
<b>Service era, n (%)</b>							
Pre-Vietnam	372 (21.1)	244 (22.1)	287 (20.6)	282 (21.0)	207 (22.3)	88 (18.3)	72 (22.9)
Vietnam	1012 (57.5)	635 (57.6)	803 (57.7)	785 (58.3)	528 (56.9)	292 (60.8)	172 (54.6)
Post-Vietnam and other	375 (21.3)	223 (20.2)	301 (21.6)	279 (20.7)	193 (20.8)	100 (20.8)	71 (22.5)
Distance to VA <sup>d</sup> (station 573), mean (SD)	126.3 (229.5)	130.3 (249.0)	127.9 (231.0)	126.6 (225.0)	127.7 (221.2)	124.3 (231.2)	129.7 (258.0)

<sup>a</sup>eGFR: estimated glomerular filtration rate (ml·min<sup>-1</sup>·m<sup>-2</sup>).

<sup>b</sup>ICD-10: International Classification of Diseases, Tenth Revision.

<sup>c</sup>CKD: chronic kidney disease.

<sup>d</sup>VA: Veterans Affairs.

**Table 3.** Prevalence of advanced chronic kidney disease (CKD) based on different criteria.

Prevalence subpopulation definition	Users, n	VA <sup>a</sup> users with creatinine lab measurement within last 12 months (n=93,216), % (95% CI)	Total VA users (N=133,756), % (95% CI)
Total VA users with at least one creatinine measurement within the last 12 months	93,216	100.0 (100-100)	69.7 (69.4-69.9)
Veterans with ICD-10 <sup>b</sup> code N18.4 or N18.5 within last 12 months	1102	1.2 (1.1-1.3)	0.8 (0.8-0.9)
Veterans with index eGFR <sup>c</sup> <30	1391	1.5 (1.4-1.6)	1.0 (1.0-1.1)
Veterans with index eGFR <30 and 90 days prior eGFR <60	1346	1.4 (1.4-1.5)	1.0 (1.0-1.1)
Veterans with high risk of advanced CKD	928	1.0 (0.9-1.1)	0.7 (0.6-0.7)
Veterans with high and intermediate risk of advanced CKD	1408	1.5 (1.4-1.6)	1.1 (1.0-1.1)
Cumulative prevalence of advanced CKD (6 months) based on high- and intermediate-risk groups	1840	2.0 (1.9-2.1)	1.4 (1.3-1.4)

<sup>a</sup>VA: Veterans Affairs.

<sup>b</sup>ICD-10: International Classification of Diseases, Tenth Revision.

<sup>c</sup>eGFR: estimated glomerular filtration rate (ml·min<sup>-1</sup>·m<sup>-2</sup>); index eGFR refers to the latest eGFR measurement at the time of extraction of the cohort.

**Table 4.** Diagnostic accuracy of International Classification of Diseases, Tenth Revision codes for advanced chronic kidney disease (CKD) compared to estimated glomerular filtration rate (eGFR)-based defining criteria.

Accuracy metric	Index eGFR <sup>a</sup> <30 (n=1371), point estimate (95% CI) <sup>b</sup>	Index eGFR <30 and 90 days prior eGFR <60 (n=1346), point estimate (95% CI)	High-risk advanced CKD (n=928), point estimate (95% CI)	High- and intermediate-risk advanced CKD (n=1408), point estimate (95% CI)
Sensitivity	0.55 (0.52-0.57)	0.55 (0.53-0.58)	0.65 (0.62-0.68)	0.57 (0.55-0.60)
Specificity	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.99 (0.99-1.00)	1.00 (1.00-1.00)
Positive predictive value	0.68 (0.66-0.71)	0.68 (0.65-0.71)	0.55 (0.52-0.58)	0.74 (0.71-0.76)
Negative predictive value	0.99 (0.99-0.99)	0.99 (0.99-0.99)	1.00 (1.00-1.00)	0.99 (0.99-0.99)

<sup>a</sup>Index eGFR refers to the latest eGFR measure (ml·min<sup>-1</sup>·m<sup>-2</sup>) at the time of extraction of the cohort.

<sup>b</sup>20 patients were excluded from this column because they were missing a prior eGFR value; the subsequent column criteria/definitions required two eGFR values, and thus patients without two eGFR values were excluded for consistency between column criteria/definitions.

**Table 5.** Probability of remaining in advanced chronic kidney disease stages (4 and 5) based on various electronic health record phenotypes at the 6-month follow-up.

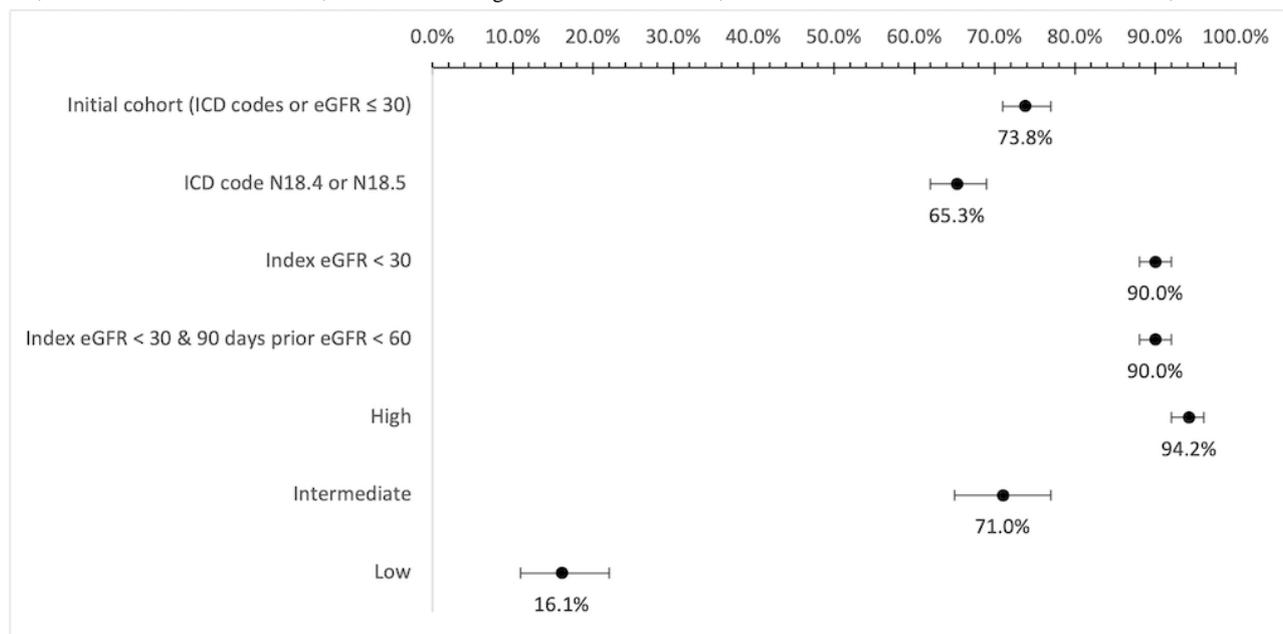
Characteristic	Initial cohort (ICD=10 <sup>a</sup> codes or eGFR <sup>b</sup> ≤30) (N=1759)		ICD-10 code N18.4 or N18.5 (n=1102)		Index eGFR <sup>c</sup> <30 (n=1391)		Index eGFR <30 and 90 days prior eGFR <60 (n=1346)		High risk (n=928)		Intermediate risk (n=480)		Low risk (n=315)	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
eGFR, mean (SD)	26.2 (12.1)	26-27	27.7 (13.7)	27-29	22.0 (6.5)	22-22	22.0 (6.5)	22-22	20.3 (6.6)	20-21	27.4 (5.6)	27-28	42.1 (16.6)	40- 44
Days between stage-defining eGFR values, mean (SD)	263.8 (394.3)	245-282	229.0 (304.9)	211-247	274.3 (399.8)	253-296	260.7 (341.7)	242-279	226.7 (159.3)	216-237	316.0 (526.0)	269-363	296.8 (594.2)	231-363
Subsequent eGFR measurement, n (%)	981 (55.8)	53%-58%	706 (64.1)	61%-67%	753 (54.1)	51%-57%	751 (55.8)	53%-58%	549 (59.2)	56%-62%	238 (49.6)	45%-54%	180 (57.1)	51%-63%
Any subsequent eGFR ≥30 after index eGFR, n (%)	293 (16.7)	15%-19%	271 (24.6)	22%-27%	96 (6.9)	5.7%-8.4%	96 (7.1)	5.8%-8.7%	39 (4.2)	3.0%-5.8%	85 (17.7)	14%-21%	162 (51.4)	46%-57%
Current eGFR ≥30 at 6-month follow-up, n (%)	257 (26.2)	23%-29%	245 (34.7)	31%-38%	75 (10.0)	8.0%-12.0%	75 (10.0)	8.0%-12.0%	32 (5.8)	4.1%-8.2%	69 (29.0)	23%-35%	151 (83.9)	78%-89%

<sup>a</sup>ICD-10: International Classification of Diseases, Tenth Revision.

<sup>b</sup>eGFR: estimated glomerular filtration rate (ml·min<sup>-1</sup>·m<sup>-2</sup>).

<sup>c</sup>Index eGFR: latest eGFR measure at the time of extraction of the cohort.

**Figure 2.** Probability of remaining in advanced CKD stages (4 and 5) based on various EHR phenotypes at 6-month follow-up. CKD: chronic kidney disease; EHR: electronic health record; eGFR: estimated glomerular filtration rate; ICD-10: International Classification of Diseases, Tenth Revision.



## Discussion

### Principal Findings

Accurate identification of an advanced CKD cohort within a clinical database can allow large health care organizations to provide targeted evidence-based clinical care, conduct system-wide needs assessment studies, and facilitate clinical and epidemiological outcome studies. Several EHR-based models to identify CKD using ICD codes and laboratory values have been published [12,21,22]. While there is a reasonable consensus regarding the EHR-based strategies to define CKD within a clinical database, no targeted study has examined the feasibility of extracting an advanced CKD cohort in such databases. Exploring the clinical database of one of the largest regional Veterans health care systems in the country, we identified several coding, identification, and accuracy-related concerns in extracting an advanced CKD cohort.

Researchers have conventionally used the provider diagnosis codes to identify and stage patients with CKD in clinical databases. Using the more accurate eGFR-based definitions, several investigators have shown that identifying CKD cohorts purely by diagnostic codes underestimates its true prevalence [23]. For example, Diamantidis et al [24] showed that the clinical recognition of CKD utilizing diagnostic codes was only 11.8% among Medicare beneficiaries. In a systemic review of studies primarily conducted on non-VHA health care databases, Grams et al [23] found that the coding accuracy for CKD varies widely between 8% and 83%, depending on providers' awareness, and rises with the comorbidity burden and severity of CKD.

Few investigators have evaluated the use and accuracy of CKD diagnosis codes in the VHA clinical database. In a recent analysis of the national VHA database, Saran et al [16] estimated the burden and cost of CKD care on VHA among over 6 million VHA-registered Veterans. While the investigators did not

examine the coding accuracy, they found its overall use to be very low (3.2%) compared to much higher estimates (8.02%-27%) obtained using laboratory values [16]. Similar results were recently obtained by Bansal et al [19] in a selective cohort of Veterans with diabetes/hypertension at Veteran Integrated Service Network 17. They found that the laboratory-based prevalence of CKD was approximately 36%, but only 44% of them had diagnosis codes for CKD [19]. Similarly, Norton et al [25] found that 63% of entries lacked CKD codes in a military health system. In conjunction with these reports, our analysis showed that the sensitivity and PPV of diagnosis codes, when compared to the eGFR-based phenotypes, to identify advanced CKD is low, in the range of 55%-65% and 55%-74%, respectively. Our study further shows that when prospectively followed, nearly one-third of the cohort defined by diagnosis codes had an eGFR value over 30 ml·min<sup>-1</sup>·1.73 m<sup>-2</sup> at the end of 6-month study. Overall, our findings confirm that the utility and accuracy of diagnosis codes for identifying advanced CKD cohorts in the VHA clinical database is poor.

There are also concerns about using an eGFR-based staging system in clinical databases. EHR-based phenotypes require laboratory measurements of creatinine; however, the regular and periodic availability of creatinine may be inconsistent in the clinical databases. For example, Norton et al [14] showed that only 55% of the study sample had eGFR measurements while validating their CKD EHR phenotype. Similarly, a study examining the VA database showed that only 65% of the VA users had any measurements of eGFR during the study period [16]. This lack of availability of eGFR measures can generate errors in the measurement of disease burden. Further, while the definition of CKD requires the demonstration of a persistent reduction of renal function, many studies report CKD staging statistics using a single eGFR value, with a significant fraction of the cohort lacking the second reported eGFR value. For example, in an analysis performed by the National Kidney

Disease Education Program Workgroup, 31% of patients with stage-4 CKD and 36% of patients with stage-5 CKD did not have a prior eGFR  $<60 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  value available [14]. Similarly, in the analysis by Saran et al [16] examining the burden of CKD in the VA database, only approximately 27% of Veterans had two eGFR measurements more than 90 days apart, raising concerns about the accuracy of the disease burden. However, in our analysis, focusing on the advanced stages of CKD, we found that over 1723 (98%) of Veterans had two eGFR values reported for the initial source cohort, substantially increasing the reliability of screening for advanced CKD. Additionally, we noticed that over 55% ( $n=981$ ) of the source cohort had subsequent measurements of eGFR over the prospective 6 months (Table 5), further providing a more robust overall reliability of our advanced CKD estimates.

While using eGFR-based phenotypes improves the identification of CKD, staging CKD into stages 3, 4, and 5 can be complex in a clinical database due to physiologic variability in creatinine levels, performance of biochemical tests, frequency of measurements, and intercurrent illness and volume status [13]. Examining such variations in repeat estimations over 3-6 months in the VHA database, Shahinian et al [26] reported that nearly 30% of patients with stage-4 CKD and 6% of patients with stage-5 CKD had eGFR values  $\geq 30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  in the repeat measurements, thus misclassifying as advanced CKD instead of CKD stage 3 [26]. These inaccuracies can lead to the misidentification of patients with advanced CKD, creating misappropriations of clinical resource allocation or errors in research outcomes for studies that target a specific advanced CKD population.

Considering these inherent limitations of eGFR and diagnostic codes, we sought to refine the predictive accuracy of isolating an advanced CKD cohort for TEACH-VET by categorizing our EHR-derived source cohort into high-, intermediate-, and low-risk advanced CKD cohorts using the two latest eGFR values obtained 90 days apart. Assessing the cohort prospectively for 6 months, we found a very high and graded level of stability with our tiered approach, with 94% and 71% of Veterans in the high-risk and intermediate-risk groups having a eGFR less than  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  at the study end point, thus remaining in an advanced CKD stage. These findings suggest that such an operational definition can significantly improve clinical and research decision-making and optimize resource allocations, which is currently used to prioritize and enroll Veterans in a clinical study targeting advanced CKD [17]. At the same time, we show that approximately 16% of those with a low risk for advanced CKD had an eGFR below  $30 \text{ ml}\cdot\text{min}^{-1}\cdot 1.73 \text{ m}^{-2}$  at the 6-month follow-up, highlighting the high-risk individuals even among those with apparent inaccuracies in diagnosis codes.

Our study explored various available methods to provide a more optimal method to obtain the population statistics for an advanced CKD burden and stratified this cohort based on their longitudinal probability of requiring stage-specific care. Examining real-time data and accurately determining the denominator to only those with an available eGFR estimation within the prespecified 12-month period, we found that the

prevalence of advanced CKD (high and intermediate risk) was 1.5%, which is 2-3 times higher compared to the US general population estimates (0.5%) derived from National Health and Nutrition Examination Survey (NHANES) enrollees [1,27], but is less than VHA estimates (1.62%) provided by Saran et al [16]. Even based on the conservative estimates and accounting for all the VA users as the denominator, the prevalence of advanced CKD seems to be higher than that of the general population (Table 3). Recently, VHA has implemented a clinical tool for identifying a CKD cohort based on a single eGFR measurement [28]. Further refinements in the tool by implementing the proposed tiered risk approach to identify an advanced CKD population can allow the VHA to implement judicious allocation of care and resources to those in the highest need. A manual chart review showed an error rate of 11%, mainly attributed to the Veterans being on dialysis. Although the VA database can be linked to the United States Renal Data System (USRDS) database and help exclude dialysis patients, there is a lag in the USRDS data and hence this might not be helpful when the need for identification of advanced CKD in real time arises, as intended in our study for enrollment into a clinical trial [8,17]. In the VHS system, using the community care dialysis list can further increase the sensitivity of the screened list and reduce the error rate by excluding the Veterans who are currently receiving dialysis.

### Limitations

Our study has a few limitations. In recent times, investigators have described advanced EHR algorithms to identify patients with CKD [29]. However, such phenotypes require complex machine-learning algorithms and validation for the target population, and their application in staging CKD is even further away. This study aimed to explore a pragmatic model for identifying Veterans with high, intermediate, and low risk of advanced CKD in real time that can be easily implemented in routine practice and across a large health care system. Second, we did not incorporate the presence or severity of albuminuria within our parsimonious risk model. However, we believe that it is unlikely to improve upon the model for several reasons. Measurement of albuminuria or even proteinuria is uncommon in clinical databases, including the VHA database, and frequently requires the use of proteinuria categorization on routine urinalysis. The risk for complications and adverse outcomes is significantly high for advanced CKD, as highlighted in the KDIGO classification, irrespective of the degree of albuminuria. Considering the unreliable availability of urine protein measurement, it is likely to be of limited additional value, if any [10]. We acknowledge that the true significance of our parsimonious approach will require studies examining longitudinal clinical outcomes. Third, our eGFR values are based on the creatinine values and utilizing the MDRD equation, according to the then-prevalent practices of the VA CDW at the time of the study. Since the overall intention of the study was to evaluate the methodologies for identifying advanced CKD cohorts within a health care system such as VHA, this is unlikely to change the outcome of the study. Future analyses will need to consider the updated CKD-Epidemiology Collaboration equations incorporating creatinine and cystine values for more accurate staging of CKD. Finally, it needs to

be mentioned that our results are applicable only among the active VHA users rather than all VHA-registered Veterans, and thus may misrepresent the true burden of advanced CKD among the entire Veteran population. EHR phenotypes, in general, may exclude people with reduced access to care.

### Conclusion

We found that the prevalence of advanced CKD at NF/SG VHS is higher than that in the general population as per various EHR phenotypes, including our EHR model. There is significant discordance between coding and laboratory parameters for the identification of advanced CKD, consistent with other studies.

EHR phenotypes based on CKD diagnosis codes alone are insufficient for identification of an advanced CKD cohort in a clinical database. We report a simplified and pragmatic EHR-based model to identify advanced CKD within a regional VHS in real time with a tiered approach that allows allocation of resources to the groups requiring immediate attention and are at risk of progression to ESKD. Further testing of this model is needed to determine its broader applicability across the VHA. If validated, similar models can be tested across the non-VHA databases to identify the true burden of advanced CKD and target clinical care in real time.

### Acknowledgments

This study was supported by funding from the Department of Veterans Affairs, Office of Rural Health FY21 Awards, and Health Service Research and Development Awards (I01HX002639). AMS additionally receives ongoing grant support from the Department of Veterans Affairs, Clinical Science Research and Development Merit Grant (I01CX001661).

### Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

ICD codes used for deriving the source cohort and stratifying the advanced CKD cohort (Tables S1) and source data for Figure 2 (Table S2).

[\[DOCX File , 16 KB-Multimedia Appendix 1\]](#)

### References

1. Chronic Kidney Disease Surveillance System-United States. Centers for Disease Control and Prevention. URL: <http://www.cdc.gov/ckd> [accessed 2021-12-06]
2. Smart NA, Titus TT. Outcomes of early versus late nephrology referral in chronic kidney disease: a systematic review. *Am J Med* 2011 Nov;124(11):1073-1080 [doi: [10.1016/j.amjmed.2011.04.026](https://doi.org/10.1016/j.amjmed.2011.04.026)] [Medline: [22017785](https://pubmed.ncbi.nlm.nih.gov/22017785/)]
3. Fung E, Chang TI, Chertow GM, Thomas I, Asch SM, Kurella Tamura M. Receipt of nephrology care and clinical outcomes among veterans with advanced CKD. *Am J Kidney Dis* 2017 Nov;70(5):705-714 [FREE Full text] [doi: [10.1053/j.ajkd.2017.06.025](https://doi.org/10.1053/j.ajkd.2017.06.025)] [Medline: [28811048](https://pubmed.ncbi.nlm.nih.gov/28811048/)]
4. Shukla AM, Easom A, Singh M, Pandey R, Rotaru D, Wen X, et al. Effects of a comprehensive predialysis education program on the home dialysis therapies: a retrospective cohort study. *Perit Dial Int* 2017;37(5):542-547 [doi: [10.3747/pdi.2016.00270](https://doi.org/10.3747/pdi.2016.00270)] [Medline: [28546368](https://pubmed.ncbi.nlm.nih.gov/28546368/)]
5. Shukla AM, Hinkamp C, Segal E, Ozrazgat Baslanti T, Martinez T, Thomas M, et al. What do the US advanced kidney disease patients want? Comprehensive pre-ESRD Patient Education (CPE) and choice of dialysis modality. *PLoS One* 2019 Apr 9;14(4):e0215091 [FREE Full text] [doi: [10.1371/journal.pone.0215091](https://doi.org/10.1371/journal.pone.0215091)] [Medline: [30964936](https://pubmed.ncbi.nlm.nih.gov/30964936/)]
6. Easom AM, Shukla AM, Rotaru D, Ounpraseuth S, Shah SV, Arthur JM, et al. Home run-results of a chronic kidney disease telemedicine patient education study. *Clin Kidney J* 2020 Oct;13(5):867-872 [FREE Full text] [doi: [10.1093/ckj/sfz096](https://doi.org/10.1093/ckj/sfz096)] [Medline: [33123362](https://pubmed.ncbi.nlm.nih.gov/33123362/)]
7. Shukla AM, Bozorgmehri S, Ruchi R, Mohandas R, Hale-Gallardo JL, Ozrazgat-Baslanti T, et al. Utilization of CMS pre-ESRD Kidney Disease Education services and its associations with the home dialysis therapies. *Perit Dial Int* 2021 Sep 01;41(5):453-462 [FREE Full text] [doi: [10.1177/0896860820975586](https://doi.org/10.1177/0896860820975586)] [Medline: [33258420](https://pubmed.ncbi.nlm.nih.gov/33258420/)]
8. Saran R, Li Y, Robinson B, Abbott KC, Agodoa LYC, Ayanian J, et al. US Renal Data System 2015 Annual Data Report: epidemiology of kidney disease in the United States. *Am J Kidney Dis* 2016 Mar;67(3 Suppl 1):Svii, S1-Svii,305 [FREE Full text] [doi: [10.1053/j.ajkd.2015.12.014](https://doi.org/10.1053/j.ajkd.2015.12.014)] [Medline: [26925525](https://pubmed.ncbi.nlm.nih.gov/26925525/)]
9. Shukla AM, Cavanaugh KL, Wadhwa A, Crowley ST, Fried L. Basic requirements for improving home dialysis utilization: universal access to specialty nephrology care and comprehensive pre-ESKD education. *J Am Soc Nephrol* 2023 Jan 01;34(1):21-25 [doi: [10.1681/ASN.2022060685](https://doi.org/10.1681/ASN.2022060685)] [Medline: [36283810](https://pubmed.ncbi.nlm.nih.gov/36283810/)]
10. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Working Group. KDIGO 2012 Clinical practice Guideline for the Evaluation and Management of Chronic Kidney disease. *Kidney Int Suppl* 2012:19-62

11. VHA directive 1053. Chronic kidney disease prevention, early recognition, and management. Department of Veterans Affairs. 2020. URL: <https://www.va.gov/vhapublications/publications.cfm?pub=1> [accessed 2021-12-06]
12. Drawz P, Archdeacon P, McDonald CJ, Powe N, Smith K, Norton J, et al. CKD as a model for improving chronic disease care through electronic health records. *Clin J Am Soc Nephrol* 2015 Aug 07;10(8):1488-1499 [FREE Full text] [doi: [10.2215/CJN.00940115](https://doi.org/10.2215/CJN.00940115)] [Medline: [26111857](https://pubmed.ncbi.nlm.nih.gov/26111857/)]
13. Shang N, Khan A, Polubriaginof F, Zaroni F, Mehl K, Fasel D, et al. Medical records-based chronic kidney disease phenotype for clinical care and "big data" observational and genetic studies. *NPJ Digit Med* 2021 Apr 13;4(1):70 [doi: [10.1038/s41746-021-00428-1](https://doi.org/10.1038/s41746-021-00428-1)] [Medline: [33850243](https://pubmed.ncbi.nlm.nih.gov/33850243/)]
14. Norton JM, Ali K, Jurkovitz CT, Kiryluk K, Park M, Kawamoto K, et al. Development and validation of a pragmatic electronic phenotype for CKD. *Clin J Am Soc Nephrol* 2019 Sep 06;14(9):1306-1314 [FREE Full text] [doi: [10.2215/CJN.00360119](https://doi.org/10.2215/CJN.00360119)] [Medline: [31405830](https://pubmed.ncbi.nlm.nih.gov/31405830/)]
15. Tummalapalli SL, Peralta CA. An electronic CKD phenotype: a step forward in improving kidney care. *Clin J Am Soc Nephrol* 2019 Sep 06;14(9):1277-1279 [FREE Full text] [doi: [10.2215/CJN.08180719](https://doi.org/10.2215/CJN.08180719)] [Medline: [31492804](https://pubmed.ncbi.nlm.nih.gov/31492804/)]
16. Saran R, Pearson A, Tilea A, Shahinian V, Bragg-Gresham J, Heung M, VA-REINS Steering Committee, VA Advisory Board. Burden and cost of caring for US Veterans with CKD: initial findings from the VA Renal Information System (VA-REINS). *Am J Kidney Dis* 2021 Mar;77(3):397-405 [doi: [10.1053/j.ajkd.2020.07.013](https://doi.org/10.1053/j.ajkd.2020.07.013)] [Medline: [32890592](https://pubmed.ncbi.nlm.nih.gov/32890592/)]
17. Shukla AM, Hale-Gallardo J, Orozco T, Freytes I, Purvis Z, Romero S, et al. A randomized controlled trial to evaluate and assess the effect of comprehensive pre-end stage kidney disease education on home dialysis use in veterans, rationale and design. *BMC Nephrol* 2022 Mar 30;23(1):121 [FREE Full text] [doi: [10.1186/s12882-022-02740-8](https://doi.org/10.1186/s12882-022-02740-8)] [Medline: [35354430](https://pubmed.ncbi.nlm.nih.gov/35354430/)]
18. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 1999 Mar 16;130(6):461-470 [doi: [10.7326/0003-4819-130-6-199903160-00002](https://doi.org/10.7326/0003-4819-130-6-199903160-00002)] [Medline: [10075613](https://pubmed.ncbi.nlm.nih.gov/10075613/)]
19. Bansal S, Mader M, Pugh JA. Screening and recognition of chronic kidney disease in VA Health Care System primary care clinics. *Kidney360* 2020 Sep 24;1(9):904-915 [FREE Full text] [doi: [10.34067/KID.0000532020](https://doi.org/10.34067/KID.0000532020)] [Medline: [35369564](https://pubmed.ncbi.nlm.nih.gov/35369564/)]
20. R: A language and environment for statistical computing. URL: <https://www.R-project.org/> [accessed 2021-12-06]
21. Mendu ML, Ahmed S, Maron JK, Rao SK, Chaguturu SK, May MF, et al. Development of an electronic health record-based chronic kidney disease registry to promote population health management. *BMC Nephrol* 2019 Mar 01;20(1):72 [FREE Full text] [doi: [10.1186/s12882-019-1260-y](https://doi.org/10.1186/s12882-019-1260-y)] [Medline: [30823871](https://pubmed.ncbi.nlm.nih.gov/30823871/)]
22. Mendu M, Schneider L, Aizer A, Singh K, Leaf D, Lee T, et al. Implementation of a CKD checklist for primary care providers. *Clin J Am Soc Nephrol* 2014 Sep 05;9(9):1526-1535 [FREE Full text] [doi: [10.2215/CJN.01660214](https://doi.org/10.2215/CJN.01660214)] [Medline: [25135764](https://pubmed.ncbi.nlm.nih.gov/25135764/)]
23. Grams ME, Plantinga LC, Hedgeman E, Saran R, Myers GL, Williams DE, CDC CKD Surveillance Team. Validation of CKD and related conditions in existing data sets: a systematic review. *Am J Kidney Dis* 2011 Jan;57(1):44-54 [FREE Full text] [doi: [10.1053/j.ajkd.2010.05.013](https://doi.org/10.1053/j.ajkd.2010.05.013)] [Medline: [20692079](https://pubmed.ncbi.nlm.nih.gov/20692079/)]
24. Diamantidis CJ, Hale SL, Wang V, Smith VA, Scholle SH, Maciejewski ML. Lab-based and diagnosis-based chronic kidney disease recognition and staging concordance. *BMC Nephrol* 2019 Sep 14;20(1):357 [FREE Full text] [doi: [10.1186/s12882-019-1551-3](https://doi.org/10.1186/s12882-019-1551-3)] [Medline: [31521124](https://pubmed.ncbi.nlm.nih.gov/31521124/)]
25. Norton JM, Grunwald L, Banaag A, Olsen C, Narva AS, Marks E, et al. CKD prevalence in the Military Health System: coded versus uncoded CKD. *Kidney Med* 2021 Jul;3(4):586-595 [FREE Full text] [doi: [10.1016/j.xkme.2021.03.015](https://doi.org/10.1016/j.xkme.2021.03.015)] [Medline: [34401726](https://pubmed.ncbi.nlm.nih.gov/34401726/)]
26. Shahinian VB, Hedgeman E, Gillespie BW, Young EW, Robinson B, Hsu C, CDC CKD Surveillance System. Estimating prevalence of CKD stages 3-5 using health system data. *Am J Kidney Dis* 2013 Jun;61(6):930-938 [FREE Full text] [doi: [10.1053/j.ajkd.2013.01.018](https://doi.org/10.1053/j.ajkd.2013.01.018)] [Medline: [23489675](https://pubmed.ncbi.nlm.nih.gov/23489675/)]
27. Myers O, Pankratz V, Norris K, Vassalotti J, Unruh M, Argyropoulos C. Surveillance of CKD epidemiology in the US - a joint analysis of NHANES and KEEP. *Sci Rep* 2018 Oct 26;8(1):15900 [doi: [10.1038/s41598-018-34233-w](https://doi.org/10.1038/s41598-018-34233-w)] [Medline: [30367154](https://pubmed.ncbi.nlm.nih.gov/30367154/)]
28. Chronic Kidney Disease Dashboards. powerbigov.us. US Department of Veterans Affairs. URL: <https://app.powerbigov.us/groups/me/reports/1033dde8-98b0-4d13-9117-0553a6213576/ReportSection3c3f482e02a663a111c7> [accessed 2022-09-23]
29. Ilyas H, Ali S, Ponum M, Hasan O, Mahmood MT, Iftikhar M, et al. Chronic kidney disease diagnosis using decision tree algorithms. *BMC Nephrol* 2021 Aug 09;22(1):273 [FREE Full text] [doi: [10.1186/s12882-021-02474-z](https://doi.org/10.1186/s12882-021-02474-z)] [Medline: [34372817](https://pubmed.ncbi.nlm.nih.gov/34372817/)]

## Abbreviations

- AKI:** acute kidney injury
- CDW:** Corporate Data Warehouse
- CKD:** chronic kidney disease
- CPT:** Current Procedural Terminology
- eGFR:** estimated glomerular filtration rate
- EHR:** electronic health record

**ESKD:** end-stage kidney disease

**ICD:** International Classification of Diseases

**KDIGO:** Kidney Disease: Improving Global Outcomes

**MDRD:** Modification of Diet in Renal Disease

**NF/SG:** North Florida/South Georgia

**NHANES:** National Health and Nutrition Examination Survey

**PPV:** positive predictive value

**TEACH-VET:** Trial to Evaluate and Assess the effects of Comprehensive pre-ESKD education on Home dialysis among Veterans

**USRDS:** United States Renal Data System

**VA:** Veterans Affairs

**VHA:** Veterans Health Administration

**VHS:** Veterans Health System

**VINCI:** Veterans Affairs Informatics and Computing Infrastructure

*Edited by A Mavragani; submitted 10.10.22; peer-reviewed by C Yang, J Norton, R Kotagiri; comments to author 21.02.23; revised version received 14.04.23; accepted 17.04.23; published 24.07.23*

*Please cite as:*

*Chamarthi G, Orozco T, Shell P, Fu D, Hale-Gallardo J, Jia H, Shukla AM*

*Electronic Phenotype for Advanced Chronic Kidney Disease in a Veteran Health Care System Clinical Database: Systems-Based Strategy for Model Development and Evaluation*

*Interact J Med Res 2023;12:e43384*

*URL: <https://www.i-jmr.org/2023/1/e43384>*

*doi: [10.2196/43384](https://doi.org/10.2196/43384)*

*PMID: [37486757](https://pubmed.ncbi.nlm.nih.gov/37486757/)*

©Gajapathiraju Chamarthi, Tatiana Orozco, Popy Shell, Devin Fu, Jennifer Hale-Gallardo, Huanguang Jia, Ashutosh M Shukla. Originally published in the Interactive Journal of Medical Research (<https://www.i-jmr.org/>), 24.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.i-jmr.org/>, as well as this copyright and license information must be included.