CJASN ePress. Published on November 4, 2010 as doi: 10.2215/CJN.04230510

Development and Validation of an Electronic Health Record–Based Chronic Kidney Disease Registry

Sankar D. Navaneethan,* Stacey E. Jolly,† Jesse D. Schold,*‡ Susana Arrigain,‡ Welf Saupe,§ John Sharp,‡ Jennifer Lyons,* James F. Simon,* Martin J. Schreiber, Jr.,* Anil Jain,†§ and Joseph V. Nally, Jr.*

**Department of Nephrology and Hypertension, Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, Ohio; †Department of General Internal Medicine, Medicine Institute, Cleveland Clinic, Cleveland, Ohio; ‡Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio; and §eCleveland Clinic eResearch, Cleveland Clinic, Cleveland, Ohio*

Background and objectives: Chronic kidney disease (CKD) is increasing, and outcomes-related research from diverse health care settings is needed to target appropriate efforts and interventions. We developed an electronic health record (EHR)-based CKD registry at the Cleveland Clinic and validated comorbid conditions.

Design, setting, participants, & measurements: Patients who had at least one face-to-face outpatient encounter with a Cleveland Clinic health care provider and (*1*) had two estimated GFR values <60 ml/min per 1.73 m^2 >90 days apart as of January 1, 2005 and/or (2) were patients with International Classification of Diseases-9 (ICD-9) diagnosis codes for kidney disease were included.

Results: Our registry includes 57,276 patients (53,399 patients met estimated GFR criteria and 3877 patients met ICD-9 diagnosis code criteria) as of March 2010. Mean age was 69.5 ± 13.4 years, with 55% women and 12% African Americans. Medicare is the primary insurer for more than one half of the study cohort. The κ statistics to assess the extent of agreement between the administrative dataset extracted from the EHR and actual EHR chart review showed substantial agreement (>0.80) for all conditions except for coronary artery disease and hypertension, which had moderate agreement (<0.60).

Conclusions: Development of an EHR-based CKD registry is feasible in a large health system, and the comorbid conditions included in the registry are reliable. In addition to conducting research studies, such a registry could help to improve the quality of care delivered to CKD patients and complement the ongoing nationwide efforts to develop a CKD surveillance project.

Clin J Am Soc Nephrol 5: ●●●–●●●, 2010. doi: 10.2215/CJN.04230510

The prevalence of chronic kidney disease (CKD) is increasing, with about 13% of the U.S. population affected (1). CKD is a precursor to ESRD that warrants dialysis or transplantation. Medicare spending on ESRD was 30 billion dollars in 2008 up from 7 billion in 1991, and it is projected that this number will continue to rise given the rising number of non–dialysis-dependent CKD patients (2). Extensive literature is evolving on ESRD patients, their quality of care, and outcomes through a national reporting system, the U.S. Renal Data System (USRDS), as well as dialysis group– or company-specific registry data (2–4). The Scientific Registry of Transplant Recipients provides such data for renal transplant recipients (5). However, such a national data reporting system is not available for non–dialysis-dependent CKD patients.

In recognition of the growing CKD population and health care costs related to CKD, several efforts have been made to identify and improve the quality of care delivered to patients with mild to moderate CKD (6,7). Recently, the Centers for Disease Control (CDC) reported the need for a national CKD surveillance system that will help identify and track various aspects of CKD (8). One of the stated objectives of the CDC national CKD surveillance registry is to identify national and regional data sources (*i.e.*, electronic health records based databases or registries) that contain important health information about CKD patients especially from a variety of health care settings. Very few such large CKD registries are available in the United States, whereas such registries have been developed successfully in other countries (9–11).

A commercially available ambulatory electronic health record (EHR) was adopted in early 2001 and was fully operational in nearly all ambulatory settings at the Cleveland Clinic by 2002. Furthermore, an integrated inpatient EHR has been in use since 2006. In 2005, the Cleveland Clinic integrated result reporting from the nine Cleveland Clinic Health System hospitals (CCHS) and 15 community-based health clinics into this EHR starting in 2005. Using demographic, clinical, and laboratory data from the Cleveland Clinic EHR data repository, we developed the Cleveland Clinic CKD registry (*1*) to identify CKD patients in CCHS who have been seen only in primary

Received May 14, 2010. Accepted August 4, 2010.

Published online ahead of print. Publication date available at www.cjasn.org.

Correspondence: Dr. Sankar D. Navaneethan, Department of Nephrology and Hypertension, Glickman Urological and Kidney Institute, 9500 Euclid Avenue, Q7, Cleveland Clinic, Cleveland, OH 44195. Phone: 216-636-9230; Fax: 216-444-9378; E-mail: navanes@ccf.org

ISSN: 1555-9041/512–0001

care so that processes of care for CKD could be examined and quality-of-care programs could be developed and implemented, (*2*) to describe the course of CKD in the outpatient setting through outcomes-related research, and (*3*) to serve as a comprehensive data source for evaluation of intervention programs related to CKD. In this article, we describe the methods used in the development and validation of this registry, overview of its contents, and initial data as of March 2010.

Materials and Methods

Setting

CCHS is a not-for-profit multispecialty academic medical center located in Cleveland, OH. CCHS is comprised of the Cleveland Clinic main hospital and its 15 community-based health centers, along with 8 community hospitals. The Cleveland Clinic uses a common EHR system (Epic Systems Corporation), and all CCHS facilities use a common patient identifier also known as an enterprise master patient index, allowing for integrated data including result reporting regardless of which hospital or clinic the patient seeks care at. The Cleveland Clinic serves an estimated population of >1.5 million people, with 75% of patients coming from the seven counties adjacent to Cleveland, whereas the other 25% of patients come from throughout Ohio, the United States, and other countries.

Use of the Cleveland Clinic EHR is mandated for all ambulatory providers and is routinely used for scheduling, order entry, documentation of progress notes, results review, medication management, and provider-to-provider and provider-patient communication. The eResearch group at Cleveland Clinic maintains a clinical data repository of EHR data and support personnel experienced in the use of EHR data. Both the data and personnel were used for this project in an Institutional Review Board–approved manner.

Identification of Patients with CKD

We used the following criteria to identify CKD patients seen at the Cleveland Clinic (Table 1) (12).

Inclusion Criteria. Patients who had at least one face-to-face outpatient encounter with a Cleveland Clinic health care provider and (*1*) had two estimated (eGFR) values <60 ml/min per 1.73 m^2 >90 days apart as of January 1, 2005 and/or (*2*) were patients with International Classification of Diseases (ICD-9) diagnosis codes (used twice in an outpatient encounter) for CKD, polycystic kidney disease, glomerulonephritis, diabetic nephropathy, hypertensive nephrosclerosis, or renovascular disease as of January 1, 2005.

Exclusion Criteria. Patients <18 years of age and those who were already diagnosed with ESRD needing dialysis or renal transplant were excluded. We also excluded patients with serum creatinine >20 mg/dl. Patients who did not have ESRD as of January 1, 2005 and met the above inclusion criteria, but subsequently developed ESRD or died during the study period, were retained in the registry.

This registry will be updated every 3 months, which will help us to

Table 1. Characteristics of CKD patients included in our registry (stratified by CKD stage on the date of CKD confirmation)

	Stage 3 (n = 48,195)	Stage 4 (n = 3804)	Stage 5 (n = 1400)	ICD-9 Code Diagnosis Only (n = 3877)	Combined (n = 57,276)
Age [mean (SD)]	70.4 (12.5)	70.8 (14.1)	62.2 (14.8)	59.5 (17.0)	69.5 (13.4)
Years in registry [mean (SD)]	2.9 (1.4)	3.3 (1.5)	2.8 (1.7)	2.1 (1.8)	2.8 (1.5)
Gender [n (%)]					
female	27,161 (56.4)	2,091 (55.0)	668 (47.7)	1,694 (43.7)	31,614 (55.2)
male	21,034 (43.6)	1,713 (45.0)	732 (52.3)	2,183 (56.3)	25,662 (44.8)
Ethnic group [n (%)]					
missing	409 (0.8)	40 (1.1)	14 (1.0)	150 (3.9)	613 (1.1)
white	41,903 (86.9)	3,036 (79.8)	792 (56.6)	2,640 (68.1)	48,371 (84.5)
African American	4,791 (9.9)	635 (16.7)	545 (38.9)	939 (24.2)	6,910 (12.1)
Hispanic/Latino	414 (0.9)	35 (0.9)	31 (2.2)	74 (1.9)	554 (1.0)
Asian/Pacific Island	200 (0.4)	18 (0.5)	7 (0.5)	30 (0.8)	255 (0.4)
other	478 (1.0)	40 (1.1)	11 (0.8)	44 (1.1)	573 (1.0)
Insurance group [n (%)]					
missing	1,073 (2.2)	103 (2.7)	66 (4.7)	197 (5.1)	1,439 (2.5)
Medicare	25,710 (53.3)	2,316 (60.9)	902 (64.4)	1,666 (43.0)	30,594 (53.4)
Medicaid	826 (1.7)	119 (3.1)	73 (5.2)	192 (5.0)	1,210 (2.1)
others	20,586 (42.7)	1,266 (33.3)	359 (25.6)	1,822 (47.0)	24,033 (42.0)
Comorbidities at inclusion [n (%)]					
diabetes	9,510 (19.7)	915 (24.1)	350 (25.0)	1,595 (41.1)	12,370 (21.6)
hypertension	39,336 (81.6)	2,862 (75.2)	980 (70.0)	3,039 (78.4)	46,217 (80.7)
coronary artery disease	8,802 (18.3)	740 (19.5)	185 (13.2)	579 (14.9)	10,306 (18.0)
congestive heart failure	3,261 (6.8)	413 (10.9)	109 (7.8)	330 (8.5)	4,113 (7.2)
hyperlipidemia	36,188 (75.1)	2,609 (68.6)	846 (60.4)	2,691 (69.4)	42,334 (73.9)
Cerebrovascular disease	3,693 (7.7)	332 (8.7)	84 (6.0)	209 (5.4)	4,318 (7.5)

Stage 3 CKD, GFR 30 to 59 ml/min per 1.73 m^2; stage 4 CKD, GFR 15 to 29 ml/min per 1.73 m^2; stage 5 CKD, GFR <15 ml/min per 1.73 m^2; SD, standard deviation.

identify new CKD patients who meet these inclusion criteria in the Cleveland Clinic and assess other outcomes such as development of ESRD. In addition, for patients who met the above inclusion criteria, we extracted their previous serum creatinine values and have calculated their eGFR values. This will enable us to look at the time of onset of CKD and slope of eGFR over time.

Renal Function Parameters

Serum Creatinine. All serum creatinine measurements for the study population were performed in the same clinical laboratory, which used integrated database management system–traceable samples to minimize calibration bias.

eGFR. We calculated eGFR using the recommended four-variable Modification of Diet in Renal Disease equation (13) for all patients who had at least two outpatient serum creatinine values. We also identified patients who meet the definition of CKD using the recently released Chronic Kidney Disease Epidemiology Collaboration equation, which may have implications in clinical practice (14).

Covariates

It is important to understand the potential explanatory factors or covariates that may impact CKD and its outcomes. Therefore, we extracted data related to various covariates that can be broadly classified into the following data elements: demographics, comorbid conditions, medications, laboratory, imaging, and clinical measures. The contents of each of these key data elements included in our registry are described in Table 2.

Validation of CKD and Specific Comorbidities

In addition to having abnormal eGFR, it is possible for patients to enter the CKD registry by ICD-9 diagnosis codes related to kidney diseases alone. We wanted to validate some of the conditions for the ICD-9 diagnosis codes used in our inclusion criteria to make sure we accurately captured those patients with CKD who may or may not have had an abnormal eGFR. Next, we wanted to validate specific comorbid disease conditions included in the CKD registry. Given their known influence on CKD, we chose to validate the following six conditions: diabetes mellitus, hypertension, coronary artery disease, cerebrovascular disease, congestive heart failure, and hyperlipidemia.

EHR Validation of Select CKD Registry Content

The EHR validation process was performed in two stages by two authors (S.N. and S.J.) and by discussion with a third author (J.V.N.) when discordance arose. The two reviewers used various sections of the EHR such as problem list, physician notes, laboratory reports, and imaging results based on prespecified validation criteria developed for the select CKD registry condition.

Twenty randomly selected charts were reviewed to assess whether the ICD-9 codes (related to kidney diseases) used to include patients in the registry from ICD-9 codes alone for polycystic kidney disease, diabetic nephropathy, and glomerulonephritis identified patients with CKD. To reduce the chance of misclassification because of "rule out" diagnoses, we required two occurrences of the ICD-9 code for each comorbid disease condition from the "encounter diagnosis" section. The reviewers, who were blinded, used prespecified validation criteria developed by an expert panel of nephrologists.

Six comorbid disease conditions were selected and validated using the definitions reported in Table 3. Definitions and criteria for these comorbid conditions were based on prior definitions and criteria used in the literature (15) from a combination of billing codes, use of relevant medications, laboratory values, and imaging studies (Table 3). Comorbid conditions must also have been present within 180 days of inclusion in the CKD registry to be included as a comorbid condition for that particular patient. An additional 20 random charts were reviewed to exclude the presence of a specific comorbidity that may have been falsely claimed to be present in these patients. A total of 184 charts were reviewed, some for multiple conditions.

Linkage with External Data

We have linked our registry with the National Social Security death index to obtain details about mortality rates. Furthermore, we plan to obtain data from the Center for Medicare and Medicaid Services to validate ESRD onset.

Data Storage

Demographic, clinical, and laboratory data from the EHR was extracted from the eResearch Oracle clinical data repository (Clarity, Epic Systems, Verona, WI) and eResearch database in the Cleveland Clinic Enterprise data center and loaded into the CKD Registry that is housed in an Oracle database in the Quantitative Health Sciences research data center at the Cleveland Clinic. The underlying data are minimally transformed from the source system but may be recoded to provide harmonization across the different practice settings. For example, laboratory data that have already been appropriately coded in the EHR-associated clinical data repository may reflect multiple internal codes for the same blood test. The data are securely maintained, password protected, and are behind the Cleveland Clinic firewall with limited access granted only to study-specific personnel. To protect patient confidentiality, all personal identifiers are deleted before their use for research and reporting purposes.

Statistical Analyses

We report the baseline characteristics of our CKD registry cohort. Patients were classified using the Kidney Disease Outcomes and Quality Initiative guidelines into various stages of CKD (stage 3 CKD, eGFR 30 to 59 ml/min per 1.73 m^2; stage 4 CKD, eGFR 15 to 29 ml/min per 1.73 m^2; stage 5 CKD, eGFR <15 ml/min per 1.73 m^2) (13). We calculated sensitivity and specificity to measure the accuracy of recording presence/absence of ICD-9 codes used for diagnosis related to kidney diseases and the six comorbid disease conditions. The κ statistic to assess the extent of agreement between the administrative dataset derived from the EHR and actual EHR chart review was calculated. We categorized the κ statistic into five groups: 0.81 to 1.00 (near perfect agreement), 0.61 to 0.80 (substantial agreement), 0.41 to 0.60 (moderate agreement), 0.21 to 0.40 (fair agreement), and <0.20 (poor agreement) (16). Unix SAS 9.2 (SAS Institute, Cary, NC) was used for all descriptive statistical analyses shown. The Institutional Review Board of the Cleveland Clinic Foundation approved the inception and ongoing use of this registry.

Results

Selection of CKD Patients

Among patients who were followed in our health system, 516,022 patients had at least one outpatient serum creatinine value, and 292,372 patients had two outpatient serum creatinine values between January 1 2005 and March 2010 (Figure 1). A total of 1,181,800 patients had at least one outpatient visit in our health care system during this time period. Of those with two serum creatinine values, 53,399 patients had two eGFR <60 ml/min per 1.73 m^2 that were >90 days apart. We included an additional 3877 patients who had the ICD-9 code criteria rele-

Table 2. Key data elements included in our registry

Demographics	Comorbid Conditions[a]	Medications	Laboratory Details/Imaging	Anthropometric/BP Details
1. Patient identifier 2. Date of birth 3. Gender 4. Race/ethnicity 5. Zipcode- socioeconomic status 6. Educational status 7. Insurance details	1. Coronary artery disease 2. Congestive heart failure 3. Cerebrovasuclar disease 4. Diabetes mellitus 5. Hypertension 6. Peripheral vascular disease 7. Hyperlipidemia 8. Malignancy 9. Chronic lung disease 10. Hematological disorders 11. Thyroid disorders 12. Psychiatric disorders	1. Antihypertensive agents 2. Hyperlipidemic agents 3. Other cardioprotective agents a. Aspirin b. Clopidogrel c. Nitrates 4. Renal medications a. Phosphate binders b. Vitamin D analogues c. Erythropoeitin agents d. Iron supplements e. Sodium bicarbonate 5. Oral hypoglycemics 6. Vaccination details 7. Miscellaneous agents a. NSAIDS b. Paracetamol c. Opioid analgesics d. Laxatives e. Antihistamines f. Allopurinol	1. Complete blood count 2. Anemia studies 3. Renal function studies 4. Liver function studies 5. Lipid profile 6. Complements, other serology for various types of glomerulonephritis 7. Thyroid function studies 8. Urinary studies—random and 24-hour urinary studies for protein, creatinine clearance and stone panel 9. Renal ultrasound	1. BP—systolic BP and diastolic BP 2. Heart rate or pulse rate 3. Height (cm or inches) 4. Weight (kg or lb) 5. BMI

[a]Several other comorbid conditions are being included in the registry, and definitions of some selected conditions have been described in Table 3. NSAIDS, nonsteroidal anti-inflammatory agents; BMI, body mass index.

Table 3. ICD-9 codes used to identify and the criteria used to validate kidney disease and selected comorbid conditions included in the registry

Comorbid Condition	ICD-9 Codes for Relevant Conditions	Chart Review Validation Criteria
CKD	585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9	a) Estimated GFR b) Nephrology notes
Diabetic nephropathy	250.4, 250.40, 250.41, 250.42, 250.43	a) History of diabetes mellitus b) Micro or macroalbuminuria c) Nephrology notes
Glomerulonephritis	580, 580.0, 580.4, 580.8, 580.81, 580.89, 580.9, 582, 582.1, 582.2, 582.4, 582.8, 582.81, 582.89, 582.9	a) Renal biopsy reports b) Proteinuria c) Nephrology notes
Polycystic kidney disease	753.12, 753.13, 753.14	a) Renal ultrasound or computed tomography or CT to look for cysts b) Nephrology notes
Coronary artery disease	410, 410.0, 410.00, 410.01, 410.02, 410.1, 410.10, 410.11, 410.12, 410.2, 410.20, 410.21, 410.22, 410.3, 410.30, 410.31, 410.32, 410.4, 410.40, 410.41, 410.42, 410.5, 410.50, 410.51, 410.52, 410.6, 410.60, 410.61, 410.62, 410.7, 410.71, 401.72, 410.8, 410.81, 410.82, 410.9, 410.90, 410.91, 410.92, 411, 411.0, 411.1, 411.8, 411.81, 411.89, 412, 413, 413.0, 413.1, 413.9, 414, 414.0, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.1, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.8, 414.9	a) Cardiac catheterization results or procedure notes b) Cardiology notes
Congestive heart failure	428.0	a) Echocardiogram b) Cardiology notes c) Cardiac stress test results d) Relevant medications
Cerebrovascular disease	430.0, 431, 432.0, 432.1, 433.00, 433.01, 433.1, 433.10, 433.11, 433.2, 433.20, 433.21, 433.3, 433.30, 433.31, 433.8, 433.80, 433.81, 433.9, 433.91, 434.0, 434.00, 434.01, 434.1, 434.10, 434.11, 434.9, 434.90, 434.91, 435.0, 435.1, 435.2, 435.3, 435.8, 435.9, 436, 437.0, 437.1, 437.3, 437.4, 437.5, 437.6, 437.8, 437.9, 438, 438.0, 438.10, 438.11, 438.12, 438.19, 438.20, 438.21, 438.22, 438.30, 438.31, 438.32, 438.40, 438.41, 438.42, 438.50, 438.51, 438.52, 438.53, 438.6, 438.7, 438.81, 438.81, 438.82, 438.83, 438.84, 438.85, 438.89, 438.9	a) Computed tomography/magnetic resonance imaging/MRA/carotid doppler results b) Neurology/cardiology notes
Diabetes	250, 250.0, 250.00, 250.01, 250.02, 250.03, 250.1, 250.10, 250.11, 250.12, 250.13, 250.2, 250.20, 250.21, 250.22, 250.23, 250.3, 250.30, 250.31, 250.32, 250.33, 250.4, 250.40, 250.41, 250.42, 250.43, 250.5, 250.50, 250.51, 250.52, 250.53, 250.6, 250.60, 250.61, 250.62, 250.63, 250.7,250.70, 250.71, 250.72,250.73, 250.8, 250.81, 250.81,250.82, 250.83, 250.9, 250.90, 250.91, 250.92, 250.93	a) Use of oral hypoglycemics or insulin b) Relevant laboratory tests ($HgbA_1C$, fasting blood sugar)

Table 3. (Continued)

Comorbid Condition	ICD-9 Codes for Relevant Conditions	Chart Review Validation Criteria
Hyperlipidemia	272.0, 272.1, 272.2, 272.3, 272.4	a) Use of ICD-9 codes at two different visits and/or b) Use of statins or fibrates and/or c) Total cholesterol >200 mg/dl or LDL cholesterol >130 mg/dl)
Hypertension	401, 401.0, 401.1, 401.9, 402, 402.0, 402.00, 402.01, 402.1, 402.10, 402.11, 402.9, 402.90, 402.91, 403, 403.00, 403.01, 403.1, 403.10, 403.11, 403.9, 403.91, 404, 404.0, 404.00, 404.01, 404.02, 404.03, 404.1, 404.10, 404.11, 404.12, 404.13, 404.9, 404.90, 404.91, 404.92, 404.93, 405, 405.0, 405.01, 405.09, 405.1, 405.11, 405.19, 405.9, 405.91, 405.99	a) Use of two or more use of relevant ICD-9 codes during outpatient visits and/or b) BP >130/80 in two different visits and/or c) One outpatient use of ICD-9 code and use of at least one antihypertensive medication

vant to kidney disease who did not meet the eGFR criteria. Our final CKD cohort included 57,276 patients as of March 2010.

CKD Patient Characteristics

Mean age was 69.5 ± 13.4 years, with women making up 55% of the study cohort (Table 1). African Americans and other minority groups constituted 15% of the entire cohort. Medicare was the primary insurer for more than one half of the study cohort. Diabetes, hypertension, and hyperlipidemia were prevalent in 22, 80, and 74% of the cohort, respectively. On average, patients were in the registry for 2.8 ± 1.5 years. The demographic and clinical characteristics of stage 3 and stage 4 CKD were similar. However, the stage 5 CKD cohort was younger and had more men compared with the stage 3 and 4 cohorts. The proportion of African Americans with CKD increased from stage 3 to stage 5 CKD (9.9 to 38.9%).

Death Rate

Based on the first serum creatinine measurement among patients identified with CKD, the annual death rate was 3.7, 6.2, 9.5, and 11.2% for patients with eGFR 45 to 59, 30 to 44, 15 to 29, and <15 ml/min per 1.73 m^2, respectively.

Agreement between Data Sources

The κ statistics to assess the extent of agreement between the administrative dataset derived from the EHR and actual EHR chart review are shown in Table 4. Substantial to near perfect agreement (≥0.61) was noted for all conditions except coronary artery disease and hypertension, which had moderate agreement (0.60). Both sensitivity and specificity were >80% in the majority of conditions, along with similarly high positive and negative predictive values, indicating that EHR-based identification of the conditions that we validated are reliable.

Discussion

We described the development of an EHR-based CKD registry that includes demographic, clinical, and laboratory details of >57,000 patients with CKD in our health care system. In an effort to assess the reliability of our data, we also validated the ICD-9 diagnosis codes for kidney diseases inclusion criteria and six comorbid disease conditions included in the registry. Our dataset's agreement with EHR-extracted data for documentation of the presence and absence of comorbid conditions, other than hypertension and coronary artery disease, ranged from substantial to near perfect agreement. Our CKD registry is a unique data source rich with clinical data, representing a large population of patients with Medicare as primary payer, a large minority population, and a large geriatric population, which lends itself to answering important questions about CKD progression, treatment, and management.

Administrative datasets were originally developed for the purposes of conducting health care operations and internal quality assessment (17). However, large administrative datasets have recently become a rich source for research studies (18–20) in several disciplines of medicine as EHR became more widely used. The combination of electronically extracted EHR clinical data with traditional administrative datasets holds significant

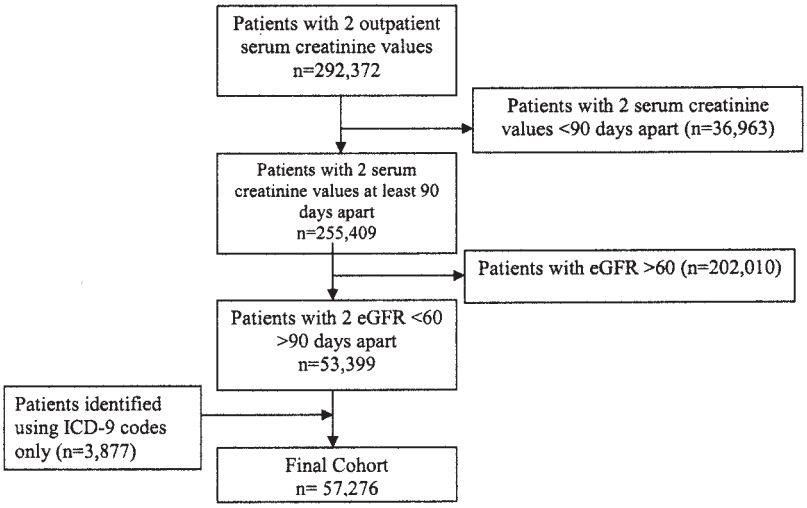

Figure 1. Flow chart showing how patients were selected to be included in the registry.

Table 4. Agreement rates between EHR and chart review for ICD-9 codes related to nephrology diagnosis and selected comorbid conditions

Condition	κ Statistic (95% CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Diabetic nephropathy	0.75 (0.55 to 0.95)	85%	90%	90%	85%
Glomerulonephritis	0.85 (0.69 to 1.00)	95%	90%	90%	85%
Polycystic kidney disease	0.90 (0.77 to 1.00)	100%	90%	90%	100%
Diabetes	0.90 (0.77 to 1.00)	91%	100%	100%	90%
Hypertension	0.45 (0.20 to 0.70)	65%	90%	95%	50%
Hyperlipidemia	0.85 (0.69 to 1.00)	95%	90%	90%	95%
Coronary artery disease	0.60 (0.35 to 0.85)	80%	80%	80%	80%
Congestive heart failure	0.85 (0.69 to 1.00)	87%	100%	100%	85%
Cerebrovascular disease	0.90 (0.76 to 1.00)	95%	95%	95%	95%

promise for providing increased value to researchers without the unnecessary burden of chart abstraction. Such registries are less expensive compared with large prospective clinical cohort studies while providing rich, patient-centric data. In kidney disease, although national datasets provide data relating to dialysis and transplant patients, datasets that contain details about non–dialysis-dependent CKD patients are limited. National datasets such as the Veterans Affairs health care system have a significantly lower proportion of female CKD patients and other limitations that render the conclusions less generalizable to the non-Veterans Affairs population. Also, available datasets often do not contain detailed laboratory information and medication details, and importantly, longitudinal data are lacking in these registries.

Administrative datasets may be criticized because they capture only actual patterns of clinical care, unlike a formal clinical study in which patients must enroll, and the relevant information is collected in a controlled prespecified manner. However, the real world setting information from administrative datasets, such as disease registries derived from EHRs, provides details about individuals who are less frequently enrolled in clinical trials, such as patients from minority race/ethnic groups and lower socioeconomic status. Our registry addressed some of these limitations. For instance, our registry included >7000 minority patients with CKD who could provide valuable information relating to CKD and its complication in these patients. We have a mix of payer source, such as Medicare, Medicaid, private, or no health insurance, among our patients, which facilitates examining the impact of health care access and socioeconomic status on CKD. Additionally data could be merged with neighborhood level information to further evaluate social determinants of health and CKD outcomes.

The median number of serum creatinine values available for this cohort was seven, which provides an opportunity to conduct longitudinal studies such as progression of CKD among different racial/ethnic groups and among patients with different stages or types of CKD. Furthermore, we linked our registry to the Social Security Death Index files that provide mortality data for our CKD patients. We intend to link the registry to other national databases such as U.S. Renal Data System and Scientific Registry of Transplant Recipients that might provide additional details regarding dialysis and transplant. We are

obtaining Centers for Medicare and Medicaid Services files to assess heath care utilization and outcomes for Medicare patients, which make up about 50% of the patients included in our registry. Furthermore, we plan to link our registry with our acute renal failure databases (21) and other local databases in CCHS to obtain details about data relating to cardiovascular disease, such as echocardiography, cardiac catheterizations, or other procedures.

Apart from outcome studies, such large datasets like ours with extensive clinical data may be helpful to identify the detection of infrequent outcomes such as medication toxicities. It could also serve as a venue for identifying patients for inclusion into clinical trials. Additionally, there are 51,926 patients who had one eGFR <60 ml/min per 1.73 m^2 in our health care system. They did not meet our inclusion criteria and therefore we did not include them in the CKD registry reported here. These patients may or may not have a similar prognosis to patients with two eGFR values <60 ml/min per 1.73 m^2, a hypothesis that could be tested in the future using our registry data and available EHR data of these patients. In our health care system, inpatient care occurs within the same vendor-based electronic health record as ambulatory care, and all inpatient data are available and can be integrated into this CKD registry project to answer important questions about hospitalizations and morbidity.

We evaluated the agreement between the administrative data and chart data because accuracy and completeness of these data are important issues given the potential for errors that occur during the development phase of such datasets. We validated the use of ICD-9 diagnosis codes related to kidney disease and six common comorbid disease conditions seen in CKD patients. The high sensitivity, specificity, positive predictive value, and negative predictive values all assured us about the reliability of our data. Quan *et al.* (22) validated several comorbid conditions derived from ICD-9 code–based administrative data and reported similar values for the conditions that we tested. We noted a lower sensitivity and κ statistic for hypertension despite using a widely accepted definition for hypertension. Such lower rates for hypertension have also been noted by other investigators (23). The poor κ statistics for coronary artery disease might be related to the wide variety of conditions (stable and unstable angina, myocardial infarction, etc.) for which physicians might be using the ICD-9 diagnosis code for coronary artery disease. We also required cardiology notes, documentation of prior coronary artery bypass surgical history, or a cardiac catheterization report for chart validation, which may be more conservative. Other comorbid conditions included in our registry will be validated by the investigators during the conduct of relevant research projects.

In addition to conducting research studies, our comprehensive EHR will allow the development of clinical decision support tailored to CKD screening and/or management (*e.g.,* pop-up messages to convey patient-specific guidance or process of care reminders). Provider behavior and patient outcomes can be used to study interventions to improve CKD care. It would also offer an opportunity to assess the quality and process of care delivered to CKD patients, become part of the CKD surveillance that is being implemented across the country, and interface new technologies to improve management of CKD.

Our registry is not without limitations. We used the number of patients with serum creatinine measurement followed in our health care system as a denominator. Misclassification of some patients as normal who have true kidney disease could occur if they did not have two serum creatinine levels measured, which are needed to calculate eGFR during the specified time period. Furthermore, it is not clear whether findings using our patient population are generalizable to the general population or to other health care systems. For example, rather than a population survey, our registry includes only patients who sought care. In addition, the more services received the greater chance of meeting our inclusion criteria, perhaps biasing our population to be older and with comorbidities other than just CKD. Moreover, we identified our registry patients based on eGFR and ICD-9 diagnosis codes only, and therefore, patients with stage 1 and 2 CKD (eGFR > 60 ml/min per 1.73 m^2 with proteinuria and other structural abnormalities) in the absence of documented kidney disease were not included.

Although, our registry is currently primarily focused on CKD stages 3 or higher by the currently classification system, we plan to include patients with albuminuria in the absence of eGFR criteria or ICD-9 diagnosis code criteria. In addition, despite the inclusion of data from the Cleveland Clinic Health System, we recognize that patients may seek care elsewhere and that data will not be available. Thus, we will not be able to evaluate outcomes of patients (other than ESRD and death) that seek care outside of the CCHS other than through claims data from CMS. With the emergence of regional health information exchanges, we hope that the completeness of our data will only improve. Finally, although our registry contains details of when medications are prescribed and stopped by the providers and medication reconciliation efforts at each encounter increase the accuracy, we cannot currently determine whether patients fill these prescriptions nor can we determine their adherence once filled.

Conclusions

We showed that the development of a registry for a chronic condition such as CKD with specific covariates identified is feasible, reliable, and valid in a large open health care system with an integrated EHR. With a diverse population and greater details about patient characteristics, this registry provides a unique opportunity for outcomes research along with improving the care delivered to our CKD patients through quality improvement and innovative projects.

Acknowledgments

Part of the work in this manuscript was presented as a poster at the American Society of Nephrology Renal Week 2009 in San Diego, CA. We thank Ben Varone, RN, MSN, of the University of Pennsylvania and Victoria Konig of the Cleveland Clinic for participation in the development of the registry. S.D.N. is supported by the National Institutes of Health, the National Center for Research Resources, Multidisciplinary Clinical Research Career Development Program Grant RR024990

through CTSA of Cleveland. The creation of the registry was funded by an unrestricted grant from Amgen to the Department of Nephrology and Hypertension Research and Education fund. S.E.J. was supported by a National Institutes of Health grant 3U01HL064244-09S1.

Disclosures

None.

References

1. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, Van Lente F, Levey AS: Prevalence of chronic kidney disease in the United States. *JAMA* 298: 2038–2047, 2007
2. U.S. Renal Data System: *USRDS 2009 Annual Data Report: Atlas of End-Stage Renal Disease in the United States*, Bethesda, MD, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2009, pp 335–343
3. Chan KE, Lazarus JM, Wingard RL, Hakim RM: Association between repeat hospitalization and early intervention in dialysis patients following hospital discharge. *Kidney Int* 76: 331–341, 2009
4. Miller JE, Kovesdy CP, Nissenson AR, Mehrotra R, Streja E, Van Wyck D, Greenland S, Kalantar-Zadeh K: Association of hemodialysis treatment time and dose with mortality and the role of race and sex. *Am J Kidney Dis* 55: 100–112, 2010
5. U.S. Department of Health and Human Services: The U.S. Organ Procurement and Transplantation Network and the Scientific Registry of Transplant Recipients: 2008 OPTN/SRTR Annual Report: Transplant Data 1998–2007. Available at: http://www.ustransplant.org/annual_reports/current/default.htm. Accessed August 25, 2010
6. McCullough PA, Vassalotti JA, Collins AJ, Chen SC, Bakris GL: National Kidney Foundation's Kidney Early Evaluation Program (KEEP) annual data report 2009: Executive summary. *Am J Kidney Dis* 55: S1–S3, 2010
7. National Institute of Diabetes and Digestive and Kidney Diseases: National Kidney Disease Education Program (NKDEP) for Health Professionals. Available at: http://www.nkdep.nih.gov/professionals/index.htm. Accessed August 25, 2010
8. Saran R, Hedgeman E, Plantinga L, Burrows NR, Gillespie BW, Young EW, Coresh J, Pavkov M, Williams D, Powe NR: Establishing a national chronic kidney disease surveillance system for the United States. *Clin J Am Soc Nephrol* 5: 152–161, 2010
9. Hemmelgarn BR, Clement F, Manns BJ, Klarenbach S, James MT, Ravani P, Pannu N, Ahmed SB, MacRae J, Scott-Douglas N, Jindal K, Quinn R, Culleton BF, Wiebe N, Krause R, Thorlacius L, Tonelli M: Overview of the Alberta Kidney Disease Network. *BMC Nephrol* 10: 30, 2009
10. Rutkowski M, Mann W, Derose S, Selevan D, Pascual N, Diesto J, Crooks P: Implementing KDOQI CKD definition and staging guidelines in Southern California Kaiser Permanente. *Am J Kidney Dis* 53: S86–S99, 2009
11. Hanratty R, Chonchol M, Miriam Dickinson L, Beaty BL, Estacio RO, Mackenzie TD, Hurley LP, Linas SL, Steiner JF, Havranek EP: Incident chronic kidney disease and the rate of kidney function decline in individuals with hypertension. *Nephrol Dial Transplant* 25: 801–807, 2010
12. National Kidney Foundation: KDOQI CKD Guidelines. Available at: http://www.kidney.org/professionals/KDOQI/guidelines_ckd/p4_class_g1.htm. Accessed August 25, 2010
13. Levey AS, Coresh J, Greene T, Marsh J, Stevens LA, Kusek JW, Van Lente F: Chronic Kidney Disease Epidemiology Collaboration: Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clin Chem* 53: 766–772, 2007
14. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF III, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J: A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150: 604–612, 2009
15. 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* 351: 1296–1305, 2004
16. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174, 1977
17. Bates DW, Pappius EM, Kuperman GJ, Sittig D, Burstin H, Fairchild D, Brennan TA, Teich JM: Measuring and improving quality using information systems. *Stud Health Technol Inform* 52: 814–818, 1998
18. Wilke RA, Berg RL, Peissig P, Kitchner T, Sijercic B, McCarty CA, McCarty JA: Use of an electronic medical record for the identification of research subjects with diabetes mellitus. *Clin Med Res* 5: 1–7, 2007
19. Fultz SL, Skanderson M, Mole LA, Gandhi N, Bryant K, Crystal S, Justice AD: Development and verification of a "virtual" cohort using the National VA Health Information System. *Med Care* 44: S25–S30, 2006
20. Pantalone KM, Kattan MW, Yu C, Wells BJ, Arrigain S, Jain A, Atreja A, Zimmerman RS: The risk of overall mortality in patients with type 2 diabetes receiving glipizide, glyburide, or glimepiride monotherapy: A retrospective analysis. *Diabetes Care* 33: 1224–1229, 2010
21. Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP: A clinical score to predict acute renal failure after cardiac surgery. *J Am Soc Nephrol* 16: 162–168, 2005
22. Quan H, Parsons GA, Ghali WA: Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Med Care* 40: 675–685, 2002
23. Hanratty R, Estacio RO, Dickinson LM, Chandramouli V, Steiner JF, Havranek EP: Testing electronic algorithms to create disease registries in a safety net system. *J Health Care Poor Underserved* 19: 452–465, 2008