The Development and Use of an EHR-Linked Database for Glomerular Disease Research and Quality Initiatives

Richard N. Eikstadt[a]   Hailey E. Desmond[a]   Clare Lindner[a]   Liz Yao Chen[b]
Cheryl D. Courtlandt[c]   Susan F. Massengill[c]   Elaine S. Kamil[d]
Richard Lafayette[e]   Anne Pesenson[f]   Matthew Elliott[g]   Patrick E. Gipson[h]
Debbie S. Gipson[a]

[a]Division of Nephrology, Department of Pediatrics and Communicable Diseases, C.S. Mott Children's Hospital, University of Michigan, Ann Arbor, MI, USA; [b]The Lundquist Institute at Harbor-UCLA Medical Center, Los Angeles, CA, USA; [c]Division of Pediatric Nephrology, Department of Pediatrics, Levine Children's Hospital, Charlotte, NC, USA; [d]Division of Pediatric Nephrology, Department of Pediatrics, Cedars-Sinai Medical Center, David Geffen School of Medicine at the University of California, Los Angeles, CA, USA; [e]Division of Nephrology, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA; [f]The Polyclinic, Seattle, WA, USA; [g]Metrolina Nephrology Associates, Charlotte, NC, USA; [h]Division of Nephrology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

Abstract

Background and Objective: The use of electronic health record (EHR) data can facilitate efficient research and quality initiatives. The imprecision of ICD-10 codes for kidney diagnoses has been an obstacle to discrete data-defined diagnoses in the EHR. This manuscript describes the Kidney Research Network (KRN) registry and database that provide an example of a prospective, real-world data glomerular disease registry for research and quality initiatives. ***Methods:*** KRN is a multicenter collaboration of patients, physicians, and scientists across diverse health-care settings with a focus on improving treatment options and outcomes for patients with glomerular disease. The registry and data warehouse amasses retrospective and prospective data including EHR, active research study, completed clinical trials, patient reported outcomes, and other relevant data. Following consent, participating sites enter the patient into KRN and provide a physician-confirmed primary kidney diagnosis. Kidney biopsy reports are redacted and uploaded. Site programmers extract local EHR data including demographics, insurance type, zip code, diagnoses, encounters, laboratories, procedures, medications, dialysis/transplant status, vitals, and vital status monthly. Participating sites transform data to conform to a common data model prior to submitting to the Data Analysis and Coordinating Center (DACC). The DACC stores and reviews each site's EHR data for quality before loading into the KRN database. ***Results:*** As of January 2021, 1,192 patients have enrolled in the registry. The database has been utilized for research, clinical trial design, clinical trial end point validation, and supported quality initia-

Correspondence to:
Richard N. Eikstadt, eikstadt@umich.edu

tives. The data also support a dashboard allowing enrolling sites to assist with clinical trial enrollment and population health initiatives. ***Conclusion:*** A multicenter registry using EHR data, following physician- and biopsy-confirmed glomerular disease diagnosis, can be established and used effectively for research and quality initiatives. This design provides an example which may be readily replicated for other rare or common disease endeavors.

Introduction

Real-world data generated from within health-care systems have been leveraged to accelerate broad-spectrum and disease-specific research. Use of real-world data for glomerular disease research has historically been limited by the lack of glomerular disease diagnosis code spec-

ificity [1]. The resulting paucity of data leads to or impairs our ability to understand the patient-disease course, clinical practice patterns, and treatment responses using real-world data. In 2015, we launched a glomerular disease research and quality initiative to include real-world data while assuring accurate kidney disease diagnosis classification [2].

The objective of this report was to share an example of a prospective, real-world data, glomerular disease patient research and quality initiative, including the description of the development, operation, and use of the database.

Materials and Methods

The Kidney Research Network (KRN) is a multicenter collaboration of patients, physicians, and scientists with a focus on improving treatment options and outcomes for patients with glomerular disease. The KRN consists of the following cores: Clinical Tri-
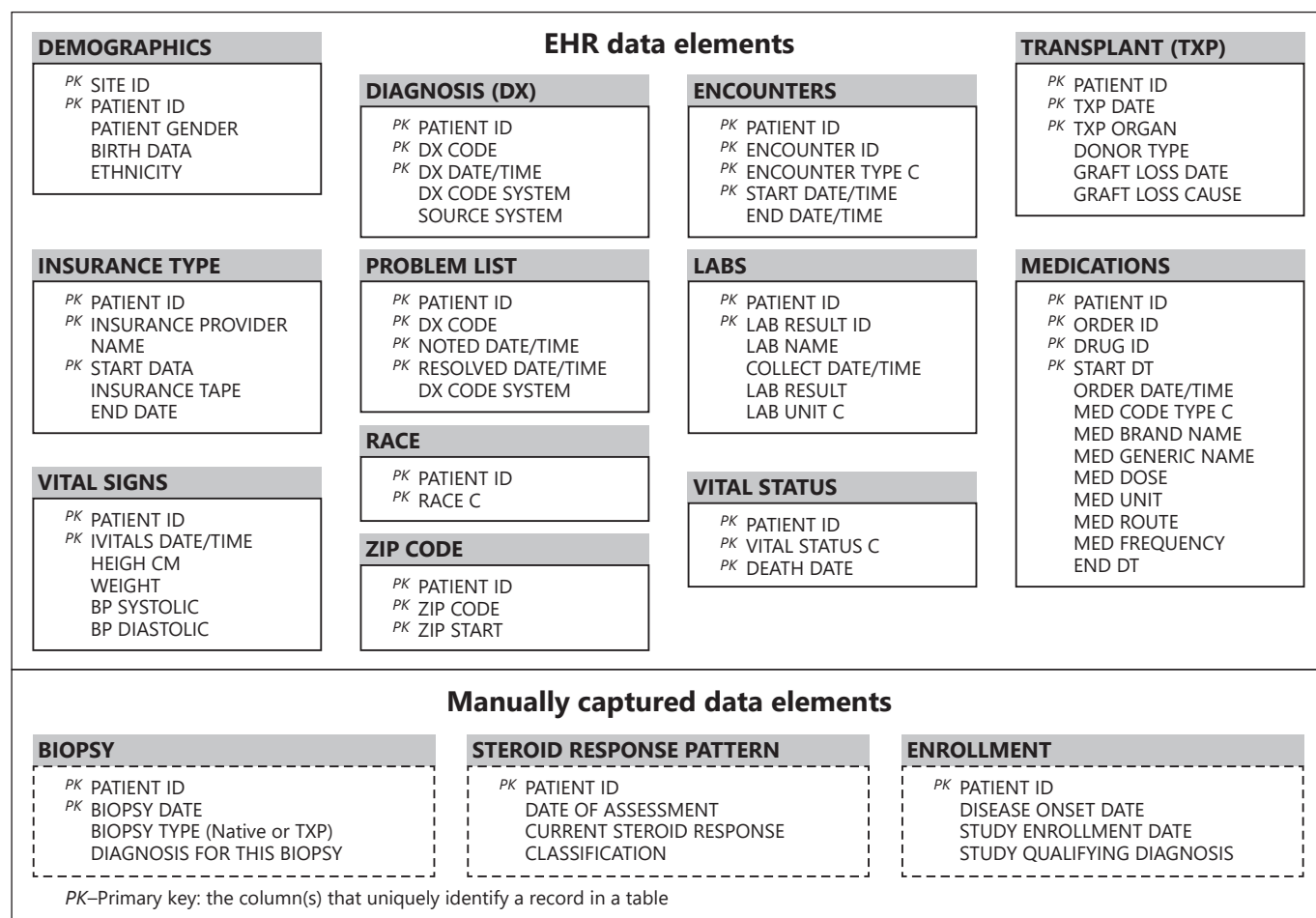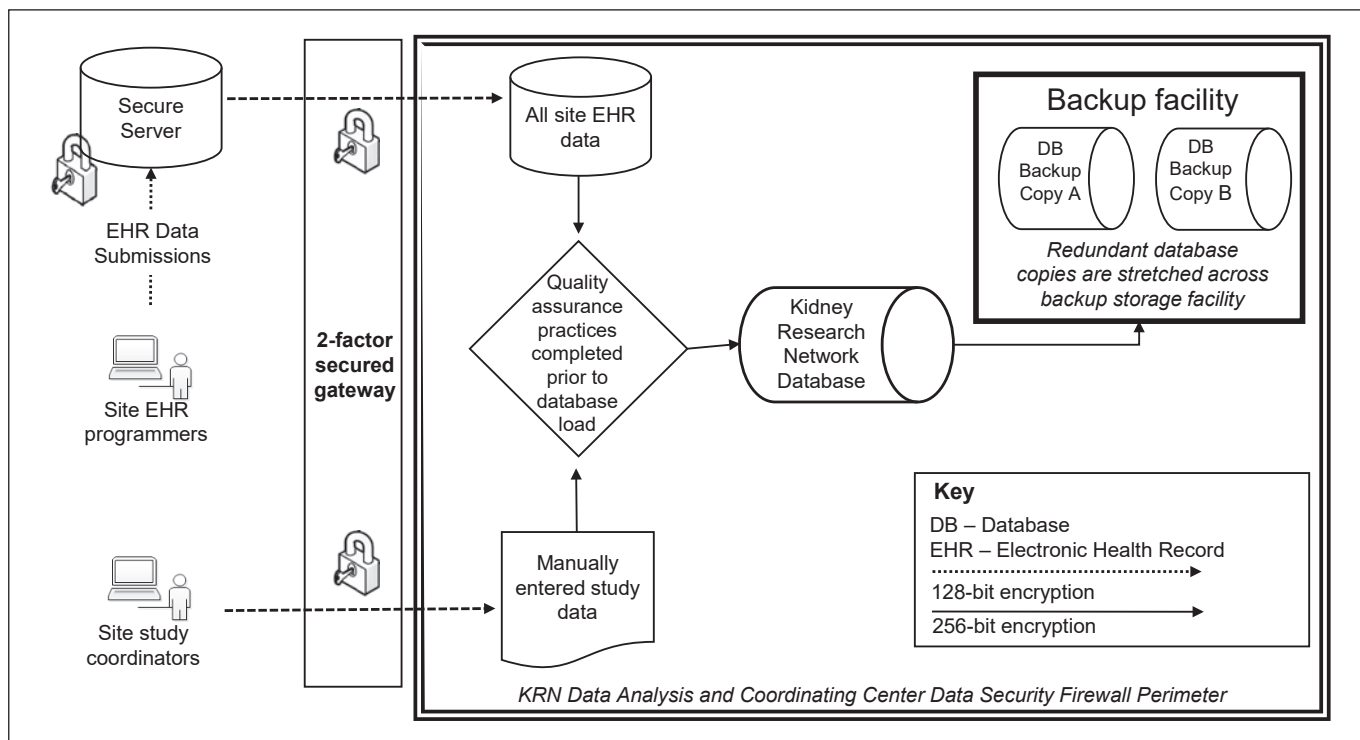

Fig. 1. KRN registry contextual data model. KRN, Kidney Research Network.

Eikstadt et al.

Fig. 2. KRN data storage and security environment. KRN, Kidney Research Network.

als Network, Clinical Trials Consulting, End points and Outcomes, Quality Initiatives, Patient Registry and Data Warehouse, and Data Analysis and Coordinating Center (DACC) [2].

The KRN Patient Registry and Data Warehouse is an ongoing data collection and storage initiative which includes a combination of retrospective and prospective data. Data sources include KRN patient registry data, active research study data, completed clinical trials data sets, patient reported outcomes data, and other relevant KRN data.

Seven enrolling sites participate in the real-world data patient registry and include a diverse selection of large, integrated health systems, academic medical centers, and private nephrology practices. Sites use different electronic health record (EHR) vendors including Epic (Verona, WI, USA), Cerner (Kansas City, MO, USA), and Allscripts (Chicago, IL, USA). The University of Michigan Institutional Review Board (IRB) provides ethics oversight for all enrolling sites and the DACC as a single IRB.

Registry participants are recruited from participating nephrology practices. Following informed consent and assent as appropriate, site study coordinators enter the date of consent, nephrologist-confirmed primary kidney disease diagnosis, and redacted kidney biopsy report into a web-based electronic data capture system (OpenClinica, LLC, Waltham, MA, USA). Patients with a diagnosis of a proteinuric kidney disease are eligible for participation, regardless of age, eGFR, or proteinuria status at the time of consent.

The study is ongoing in that the registry prospectively collects EHR data, and participating sites are continuing to enroll. There is no maximum registry enrollment target. New sites may be added to the registry in the future.

Data Domains and Duration

The KRN registry collects EHR data from demographics, insurance type, zip code, diagnoses, encounters, laboratories, procedures, medications, dialysis/transplant status, vital signs, and vital status (shown in Fig. 1). Following enrollment, EHR data are extracted from the earliest date available for each participant (retrospective) and prospectively on a monthly basis.

Data Extraction, Transformation, and Load in Secure Environments

Each enrolling site programming lead was provided with a data dictionary defining the specific expected data domains and variables with a defined common data model (CDM). Lookup tables were provided allowing programmers to code descriptive variables such as laboratory units or medication routes. These tables were provided in data definition language, a commonly recognized syntax for creating database objects. This allowed sites to efficiently create the defined tables in their local EHR extraction programming environment.

Enrolling sites extract EHR data by running monthly queries. These queries were generally embedded within automated reporting programs to reduce manual work and minimize error. No specific data extraction method was required. However, the extracted data undergoes transformation to the CDM before submission to the DACC. In practice, this means that each variable must be of the correct data type (numeric, character, date, etc.), of the correct length and format, and variables such as laboratory result units, medication frequency, etc. must be consistent with the formatting in the lookup tables provided in the data dictionary.
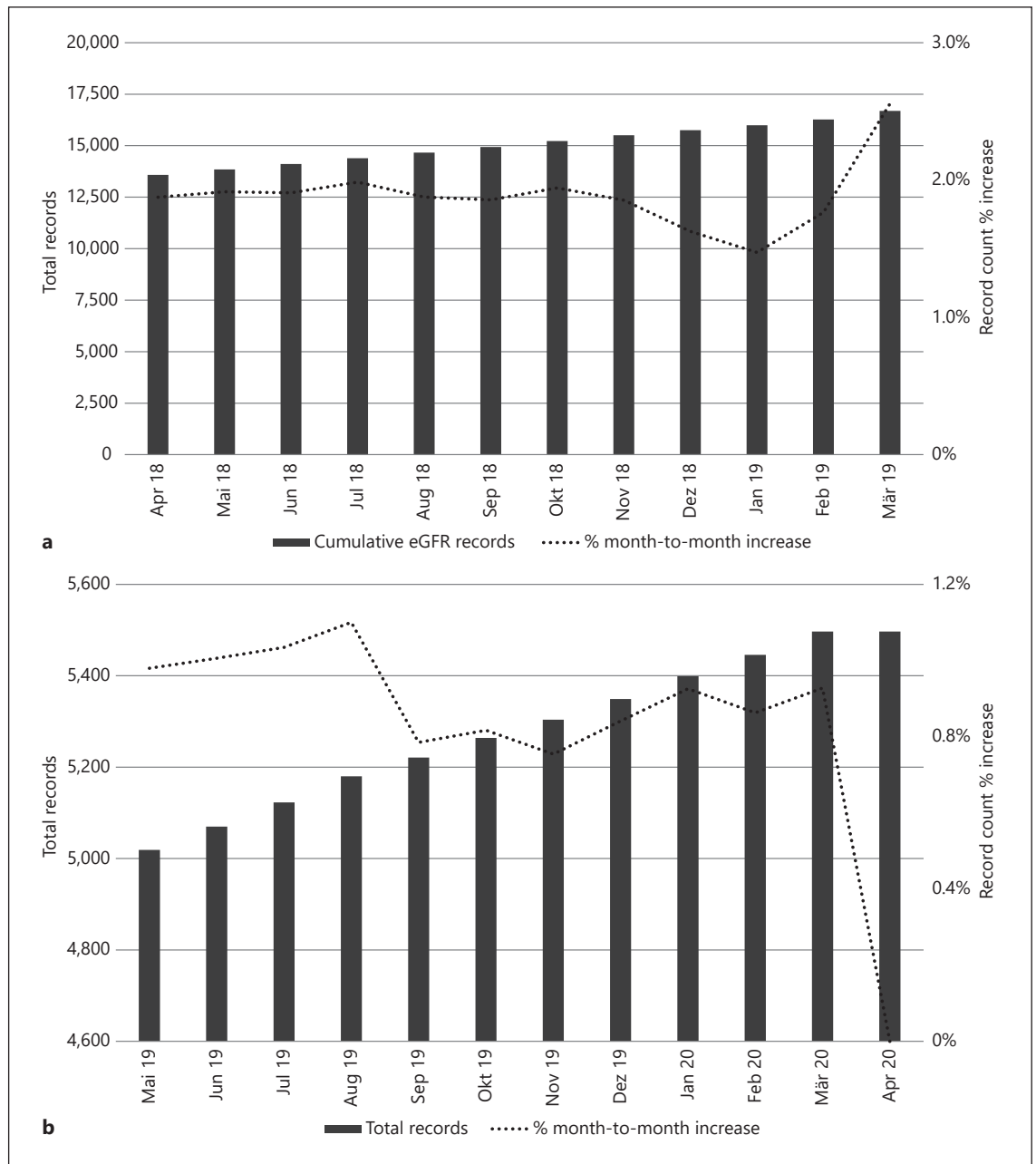
Fig. 3. KRN registry data quality assessment example plots. **a** EHR submission record count and % volume month-to-month change for eGFR laboratory. Overall data volume consistently increases. Month-to-month % increase in records is consistently within a band of 1.5–2.5% and never approaching 0%. **b** EHR submission record count and % volume month-to-month change for hemoglobin laboratory. The general % increase range is between a high of 1% in August 2019 and a low of 0.8% in November 2019. The 0% record count increase in April 2020 warrants inquiry. KRN, Kidney Research Network; EHR, electronic health record.

At monthly intervals, the DACC gathers the data file submissions from secure servers and securely transfers them to servers within the DACC firewall (shown in Fig. 2). Each domain dataset is reviewed for formatting errors and appropriate data volume increases upon receipt. Queries are returned to collaborating sites to determine the cause of any irregularities, and, if necessary, the site resubmits the data for that month. Once a full set of EHR data for a site passes the data quality checks, it is loaded into the registry database.

Eikstadt et al.
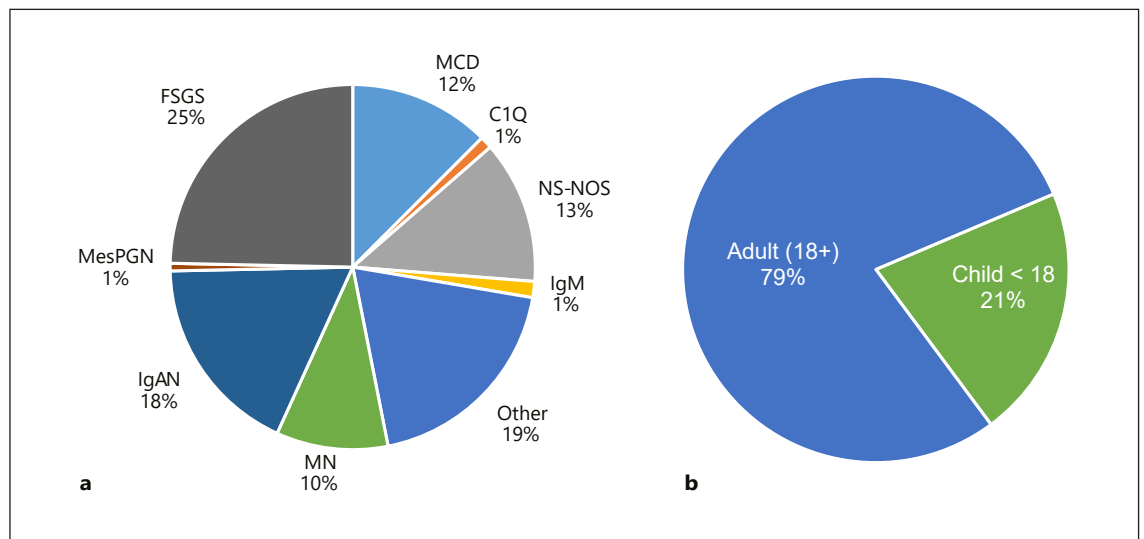
Fig. 4. KRN registry diagnosis and age distribution. **a** Physician-confirmed primary kidney diagnosis at enrollment. **b** Patient age at study enrollment. FSGS, focal segmental glomerulosclerosis; IgAN, IgA nephropathy; NS-NOS, nephrotic syndrome not biopsied/not otherwise specified; MN, membranous nephropathy; MesPGN, mesangial proliferative glomerulonephritis; IgM, IgM nephropathy; C1Q, C1q nephropathy; KRN, Kidney Research Network.

Data Quality and Security
Increasing and Historically Concordant EHR

Monthly EHR data submissions are checked to assure appropriate record count increases within each EHR data domain if appropriate. For instance, the overall laboratory EHR file for a site's set of study patients has historically increased at a rate of 0.5–1%. This same rate of increase holds true for other large volume data domains (approximately 10,000 or more records for each site's study patients) such as patient diagnosis, medications, and vital signs. Comparisons also assure that previously submitted records are present in each subsequent submission. Sites are queried when month-to-month differences in the quality assessments are not within an expected range.

Cumulative data are submitted by the site to the DACC each month. Incremental submissions capturing only newly created EHR records were attempted. However, it was discovered that sites could more easily extract and submit the full EHR record than program for new records alone. Also, data quality processes were more easily handled using the full EHR model. Consequently, there is greater efficiency at both the site level and at the DACC level when managing queries, correcting errors, or updating old records with new information.

Key Data Elements and Timing

EHR data are reviewed for the presence and appropriate volume increase of important, disease-related laboratories. For instance, patients in this registry are assured to have particular laboratory measurements, encounter types, diagnosis information, etc. appropriate to glomerular disease. Data are reviewed to assess presence of data relative to when a patient's EHR record begins, as well as to verify the ongoing collection of expected, contemporary data in a patient record. Month-to-month comparisons are made for the overall number and the percentage increase in records based on both initial data review and historical submission norms (shown in Fig. 3).

Site-Level EHR System Changes

It is not unusual for EHR systems to be replaced at a site. When an EHR system is replaced, it may be impractical to load a complete set of historical data into the new system. In these cases, the participating site must provide the earliest historical date captured for information loaded into the new system. This will allow the DACC to retain all historical EHR data prior to the transition date within the KRN registry. Future EHR submissions from the site, extracted from their new EHR system, will be combined with the retained information housed at the DACC to create a complete historical EHR within the registry.

Data Security

The KRN EHR database is an Oracle database housed on secure servers at the University of Michigan and maintained by the Health Information and Technology Services team. The servers are designed to prevent unauthorized access to data and to prevent data loss due to equipment failure or catastrophic events. The KRN EHR database and file servers are backed up daily to limit data loss. During backup, redundant copies of those data are created and stretched across a limited-access, off-site backup facility. Data are encrypted while in transit from participating sites and while stored at the DACC. All study data are stored behind a network firewall (shown in Fig. 2).

Access Controls

Access to the KRN EHR database and programming environment requires 2-factor authentication using a confidentially assigned and administered role-based username and password. Ac-

cess also requires the user possess an enrolled device. Additionally, off-site access requires the use of a virtual private network in addition to the security requirements outlined above. These procedures ensure that only authorized personnel may view, access, and modify study data.

Patient Confidentiality
Patients are identified with a unique study identification number, replacing the medical record number locally, prior to data submission to the DACC.

Results

As of January 2021, 1,192 patients have enrolled in the KRN patient registry. The diagnosis and demographics distribution are presented in Figure 4. Enrolled patients are of the following races: White 51%, Black 14%, Asian 13%, and other 22%. Overall, the median observation time for enrolled participants is 83 months (range: 1 month, 249 months), and the median number of serum creatinine results is 27 results per patient (range: 1, 784).

The database has been utilized for research, clinical trial design, end point validation [3], and quality initiatives [4–6]. Research studies related to hypertension, immunizations, medication-related adverse events, and psychiatric disorders have been presented in scientific conferences and published and have led to quality improvement initiatives to provide structured support to optimize the management and health outcomes of patients with glomerular disease.

New research findings often take >17 years to be adopted at the bedside [7]. A method to speed this process is data-driven quality improvement initiatives and implementation science. A clear advantage of this database with the submission of aggregate monthly data is that it can be used in quality improvement initiatives both locally and for joint projects over numerous sites. Rapid reporting and feedback can inform accurate decision-making in the delivery of quality patient care, allow benchmarking of care delivery, and foster collaboration among sites.

Studies vary but have shown that <60% of patients with elevated blood pressures may be identified and receive timely appropriate treatment [8]. Using the database, both adult and pediatric sites were able to measure the gap in identification and treatment of elevated blood pressure [4]. To accelerate improvement, a general change package was given to each individual site to customize and adapt to their individual setting. At the end of the project, providers identified and subsequently addressed >88% of patients with elevated blood pressures, with most sites sustaining their improvement.

In addition, the database demonstrated an increased incidence of mood disorders among patients with glomerular disease, the gap identified was the absence of universal screening for anxiety and treatment. One site, recognizing the gap in care of their patients has performed anxiety screening on >90% of their patient population and has intervened in 100% of those identified with elevated anxiety scores. This single quality improvement initiative is being disseminated to other sites within the KRN. The data have also been used to support a web-based secure dashboard accessible to enrolling sites to assist with identification of site-specific registry patients eligible for individual clinical trials; for overview of the site population by diagnosis, blood pressure control status, kidney function, and proteinuria to support population health initiatives; and to provide individual participant longitudinal health metrics.

Discussion

In this study, we present a glomerular disease registry which has been developed to aid in research and quality initiatives. This registry has the advantage of physician-reported primary glomerular disease diagnosis with and enriched by redacted (for patient identifiers) kidney biopsy reports as well as inclusive, comprehensive longitudinal data from the EHR inclusive of renal and nonrenal diagnoses [2]. This can be especially important in rare diseases such as glomerular disease, where the disease and treatments may subject the patient to intermediate and long-term disease progression and nonrenal adverse events difficult to document in single center or short term observational studies. Furthermore, the juxtaposition of research and quality initiatives supports data driven discovery and quality improvement initiatives.

The registry has limitations. Monthly data updates are from the enrolling site and do not include real-world data from care environments outside of the KRN enrolling sites, unless manually entered in the registry's electronic data capture system. Second, while prescribed medications and doses are captured, there is limited ability to measure adherence to prescribed medications beyond assessment of drug levels from the laboratory data. Finally, clearly rich information is captured in the text of patient health records. At the present time, the KRN registry does not include natural language processing methods.

Despite these limitations, the KRN patient registry is a robust resource for discovery, quality initiatives, and clinical trial design and recruitment. Data extraction, trans-

formation, and load procedures provide an efficient and effective method to gather broad-spectrum, well-curated, longitudinal data on a large population with rare glomerular diseases and enable an opportunity for expansion and collaboration while serving as an example for other therapeutic areas.

Statement of Ethics

This study was approved by the Institutional Review Board of the University of Michigan (IRB #HUM00099659). Informed consent was obtained following institutional policies from participants or legal guardians of minor participants before enrollment in the study, and consent covered future analyses of registry data.

Conflict of Interest Statement

Though there are no conflicts of interest or disclosures related to these data, the authors would like to report the following financial relationships: S.F.M. serves on the Travere Advisory Board. E.S.K. has research funding from Pfizer, has served as a onetime consultant for Mallinckrodt, and serves on the Board of NephCure Kidney International. D.S.G. has research funding through the UM with Travere, Goldfinch Bio, Novartis, and Reata Pharmaceuticals and consults through UM with Vertex and AstraZeneca; the remaining authors have disclosed that they do not have any potential conflicts of interest.

Author Contributions

Conceptualization of manuscript: D.S.G. and R.N.E.; KRN study design, conduct, and enrollment of participants: H.E.D., S.F.M., R.L., M.E., E.S.K., A.P., P.E.G., and D.S.G.; manuscript drafting: R.N.E., C.L., H.E.D., D.S.G., and L.Y.C. All the authors reviewed, revised, and approved the manuscript for submissions. Each author contributed important intellectual content during manuscript drafting and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate.

Data Availability Statement

Kidney Research Network (KRN) registry data may be requested by submitting a KRN Ancillary Proposal available in the Research section of https://www.kidneyresearchnetwork.org. Data access is governed by the KRN Steering Committee. Further inquiries can be directed to kidneyresearchnet@med.umich.edu.

References

1 Moxey-Mims MM, Flessner MF, Holzman L, Kaskel F, Sedor JR, Smoyer WE, et al. Glomerular diseases: registries and clinical trials. Clin J Am Soc Nephrol. 2016 Dec 7;11(12):2234–43.

2 Gipson DS, Selewski DT, Massengill SF, Modes MM, Desmond H, Lee L, et al. NephCure Accelerating Cures Institute: a multidisciplinary consortium to improve care for nephrotic syndrome. Kidney Int Rep. 2018 Mar; 3(2):439–46.

3 Troost JP, Trachtman H, Nachman PH, Kretzler M, Spino C, Komers R, et al. An outcomes-based definition of proteinuria remission in focal segmental glomerulosclerosis. Clin J Am Soc Nephrol. 2018 Mar 7;13(3):414–21.

4 Weaver DJ, Waldo A, Oh GJ, Kamil ES, Elliott M, Adler S, et al. Time to initiation of antihypertensive therapy after onset of elevated blood pressure in patients with primary proteinuric kidney disease. Kidney Med. 2020 Apr;2(2):131–8.

5 Oh GJ, Waldo A, Paez-Cruz F, Gipson PE, Pesenson A, Selewski DT, et al. Steroid-associated side effects in patients with primary proteinuric kidney disease. Kidney Int Rep. 2019 Nov;4(11):1608–16.

6 Tran CL, Selewski DT, Oh GJ, Troost JP, Massengill SF, Al-Akash SI, et al. Pediatric immunization practices in nephrotic syndrome: an assessment of provider and parental knowledge. Front Pediatr. 2020;8:619548.

7 Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med. 2011 Dec;104(12):510–20.

8 Khanna RR, Victor RG, Bibbins-Domingo K, Shapiro MF, Pletcher MJ. Missed opportunities for treatment of uncontrolled hypertension at physician office visits in the United States, 2005 through 2009. Arch Intern Med. 2012 Sep 24;172(17):1344.