Optimizing the Electronic Health Record for Clinical Research: Has the Time Come?

Michelle N. Rheault

The routine care of patients in nephrology settings generates an enormous amount of information, including demographics, laboratory results, anthropomorphic data, treatment information, pathology results, genomic reports, and more. But 30 years ago, if a researcher wanted to identify all the patients with nephrotic syndrome from their clinic who may be eligible for a clinical trial, or perform a retrospective research study on this population, they would have to hand-search each individual patient’s paper file to identify those with the correct diagnosis. This took a lot of time and effort, and limited the amount of research that could be completed in these real-world settings. Could there be a better way? Since the 1990s, the use of the electronic health record (EHR) has grown exponentially, partially driven by United States government incentives for use (1). Although the use of EHR data for research holds great promise, the practical implementation of this tool has been underutilized and has not lived up to its potential. In this edition of Kidney360, Oliverio et al. (2) provide a much-needed framework and proof of concept for optimally using EHR data to identify patients for nephrology research.

The EHR can easily be utilized to generate lists of patients with specific International Classification of Diseases 9th revision (ICD-9) or ICD-10 codes; however, the accuracy of these lists is limited by recognition of the diagnosis and correct coding by the provider during the encounter. Prior studies have shown that reliance on ICD-10 coding alone during a single encounter fails to identify half of patients with AKI or CKD (3). Even for patients with rare diseases such as membranous nephropathy (which nephrologists should be including in their coding documentation), ICD-9 coding was only 86% sensitive and 76% specific for identifying a patient with an actual biopsy-proven diagnosis of membranous nephropathy (4). Coding by providers is further complicated by the presence of multiple specific and nonspecific codes for nephrotic syndrome that may have some overlap. At the end of a busy clinic, did you give your patient with membranous nephropathy a code of N04.0 (Nephrotic syndrome) or N05.2 (Unspecified nephritic syndrome with diffuse membranous glomerulonephritis)? It may not matter for actual patient care or billing, however, when these data are used for research, the codes we are entering may make a big difference in how the data we are collecting reflect the actual population of patients with a specific kidney disease.

Oliverio et al. developed a “computable phenotype” to identify patients of interest from within the EHR, which aims to improve the sensitivity and specificity of these searches (2). A computable phenotype is a definition of a condition that is solely on the basis of routinely collected and stored data elements from the EHR. Patients were identified as having primary nephrotic syndrome if they had two or more nephrotic syndrome ICD-10 codes and were excluded if they had a code for a common cause of secondary nephrotic syndrome, such as diabetes mellitus, systemic lupus erythematosus, viral hepatitis, and others. The requirement for at least two codes for nephrotic syndrome aimed to reduce false positives. These computable phenotype criteria were then applied to PCORnet (the National Patient-Centered Clinical Research Network) data at three participating institutions to identify a random sampling of 50 patients and 50 nonpatients from each institution. Patients and nonpatients were reviewed individually by a nephrologist to confirm the diagnosis of primary nephrotic syndrome the old-fashioned way—through individual chart review. Sensitivity was a remarkable 99%; however, specificity was only 79%, with false positives primarily due to patients with secondary FSGS and membranous lupus nephritis (2). Interestingly, specificity varied by institution, suggesting there may be some variability among coding practices in various centers that may limit the application of this computable phenotype more broadly.

Why are computable phenotypes useful in the field of nephrology? First, they can be utilized to identify patients for clinical trials (5). Clinical trials in glomerular disorders have exploded over the past decade, with 36 currently enrolling trials for nephrotic syndrome alone listed on ClinicalTrials.gov (6). Identifying patients using an accurate computable phenotype can save coordinator time for screening, and potentially reach more patients than if a trial waited for a patient to be invited by their nephrologist. Second, computable phenotypes can be used to identify...
patients for comparative effectiveness research and epidemiologic studies. Finally, computable phenotypes can be beneficial in identifying patients with rare kidney diseases to document natural history data (7). As more and more drugs are in development for rare kidney disorders, such as Alport syndrome or hyperoxaluria, it is vital for clinicians and researchers to know the natural history of disease and who is at risk for progression.

There are limitations to the use of computable phenotypes, however, primarily the specificity for certain disorders on the basis of ICD-10 coding alone. Some computable phenotypes may include additional EHR data, such as a particular laboratory value (albumin <2.5 mg/dl, for example) or natural language processing to identify certain words in a physician note (“membranous nephropathy,” for example) to improve accuracy. These future iterations will be a valuable next step as we move forward with adopting computable phenotypes for use in clinical research for a multitude of nephrology diagnoses. Missing data are also a potential concern. For example, if a computable phenotype for nephrotic syndrome required an albumin value of <2.5 mg/dl, a patient who is truly positive may be missed if they were referred from an outside center where the diagnosis was originally made. Finally, there may be patient privacy concerns regarding how an individual’s EHR data is utilized and who may have access to those data (8). Most centers ask patients to “opt in” or “opt out” to research during the initial system intake, to allow patients to have a say in their data utilization; however, the counseling about the ramifications of opting in or out is relegated to front-desk staff, who may not have the expertise to counsel these patients appropriately.

Has the EHR finally come of age to be utilized as the vital clinical research instrument that was promised years ago? This study by Oliverio et al. provides one step toward optimizing the identification of patients with primary nephrotic syndrome for clinical research. Future work to refine the computable phenotype and apply to additional diagnoses will be welcome.

Disclosures
M. Rheault reports having consultancy agreements with Visterra; reports receiving research funding from Chinook, Reata, Sanofi, and Travere; and reports being a scientific advisor or member of the Alport Syndrome Foundation Medical Advisory Board; NephJC (501c3) Board of Directors; Peds Nephrology Research Consortium (501c3) Steering Committee, and Women In Nephrology (501c3).

Funding
None.

Acknowledgments
The content of this article reflects the personal experience and views of the author(s) and should not be considered medical advice or recommendation. The content does not reflect the views or opinions of the American Society of Nephrology (ASN) or Kidney360. Responsibility for the information and views expressed herein lies entirely with the author(s).

Author Contributions
M. Rheault conceptualized the study and wrote the original draft.

References
6. Available at: https://clinicaltrials.gov/. Accessed November 2, 2021
7. US Department of Health and Human Services FDA: Rare diseases: Natural history studies for drug development, Guidance for Industry, Rockville, MD, 2019

Received: November 3, 2021 Accepted: November 22, 2021