

Diabetic Complications Consortium

Application Title: Automated Identification of Diabetic Individuals with Renal Complications

Principal Investigator: John F. O'Toole and Dana Crawford

1. Project Accomplishments:

We have now completed creation of a relational database that will serve as the CKD registry of Cleveland Clinic patients. The registry currently contains over 200 thousand individuals with CKD or a qualifying diagnosis and the accuracy of the data extraction from the electronic health record is currently being manually reviewed in a subset of the records (Aim 1). In addition we have begun evaluation of the accuracy of medication lists obtained from the initial data pull and our ability to test eGFR slope from clinically obtained data compared to medication lists and annually scheduled creatinine measures collected as a part of the CRIC study (Aim 2). Work toward completion of the specific aims is proceeding under an approved no cost extension.

In order to increase the utility of the registry we have obtained IRB approval to expand the registry to include individuals with a calculated eGFR of <90 ml/min. We expect this to increase the size of the registry to well over 1 million individuals. Future plans beyond the scope of this proposal will also include an assessment of the number of individuals with and without diabetes that have had non-invasive kidney imaging. We anticipate that this data will provide a useful basis for future investigations for the use of these images for prediction of progression using machine learning based image analysis through our collaborators at the Case Western Reserve University Department of Biomedical Engineering. We also anticipate the possibility of combining non-invasive imaging obtained for clinical purposes with molecular phenotypes generated by the KPMP for which we are a recruiting center that will be actively recruiting subjects with diabetic kidney disease.

2. Specific Aims:

Specific Aim 1: Establish a research relational database and verify the accuracy and completeness of the data extracted.

Results: We have now updated CKD registry on a Teradata relational database consisting of 203,540 patients at the Cleveland Clinic. Contains over 18 million ICD codes, 565 million lab results and 66 million prescription orders. In addition the registry now includes ICD-10 codes, and all notes and image and procedural reports. Individuals are entered into the registry if they have one face-to-face visit at the main campus of the Cleveland clinic and two measured creatinine with eGFR <60 ml/min separated by 90 days OR a qualifying diagnosis of glomerular disease OR are seen by the nephrology consult services, OR have donated OR received a kidney transplant. Using computational assessments, 167 thousand individuals have CKD stage 3 or higher based on laboratory values or billing codes, over 71 thousand individuals have diabetes and over 140 thousand individuals have

hypertension. We are now in the process of manual chart review of records to determine the accuracy of basic metrics before proceeding to more sophisticated computational phenotyping in the next aim.

Specific Aim 2: Quantify the accuracy of a rule-based algorithm to identify subjects in the CKD registry with DKD and determine the precision of the research database-derived data in replicating medication lists and a critical kidney disease outcome variable, eGFR slope, from curated registries.

Results: Having established the relational data base we are now constructing the rule-based algorithms for computational phenotyping of diabetic kidney disease. We are also now beginning assessment of electronic capture of medications using the 310 individuals that have been enrolled and followed longitudinally at the Cleveland Clinic. Similarly, we will test our ability to accurately calculate eGFR slope using clinically captured data again using the CRIC cohort which has annually scheduled measurement of creatinine as the gold standard.

3. Publications:

Podium Abstract: (3270034) Understanding Chronic Kidney Disease: Augmenting a Code-based Registry from Clinical Records. To be presented by Dr. Ellen Palmer at the AMIA 2020 Informatics Summit March 23-26 in Houston Texas.