Pre-coordinating mapping conventions for these data will enable efficient and effective collaboration on studies that have potential to improve outcomes for those individuals with rare kidney diseases.

**Title: Capture and consolidation of renal-specific concepts into a cohesive OMOP dataset**

**BACKGROUND:** The European Rare Kidney Disease Reference Network (ERKNet) was formed by the European Union in 2017 and is one of 24 European Reference Networks (ERNs). ERKNet – made up of 32 pediatric and 20 adult nephrology centres -- aims to improve the quality of patient management by: (1) educating healthcare professionals, (2) establishing best practices, (3) enabling virtual expert consultations for unusual cases, and (4) promoting clinical research activities. Patient registries are a critical component of these aims: beginning in 2017, ERKNet centres were active in more than 80 disease-specific registries with regional or national patient coverage or voluntary participation. At that time, however, no single registry was used across all ERKNet centres and few provided monitoring of relevant disease or treatment specific performance and outcome measures. In an attempt to improve the quantity and quality of this valuable registry information, ERKNet created the European Rare Kidney Disease Registry (ERKReg) [1]. ERKNet is now an active data partner in the European Health Data and Evidence Network (EHDEN) and is working together with edenceHealth to transform their rich registry dataset, ERKReg, into the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). ERKReg is focused on rare kidney diseases, and, despite the low prevalence of these diseases, it represents more than 17,000 patients. This increased patient coverage in comparison with other registries is a direct benefit of being part of the ERKNet collaborative effort. Finally, ERKReg is characterised by its continuous longitudinal patient follow-up, which distinguishes it as unique and valuable resource for the study and treatment of rare kidney diseases.

**Figure 1:** Overview of three-machine installation, with one managed instance hosting the OMOP CDM database and ETL processes, one instance hosting the OHDSI tooling (via Docker) required to interact with the OMOP data, and a user workstation capable of connecting to both instances for updating/launching ETL processes and viewing web applications.

**LIMITATIONS AND DISCUSSION:** ERKNet has transformed more than 10'000 unique patients with approximately 500'000 total records (300'000 Observations, 150'000 Measurements, 35'000 Conditions, 15'000 Procedures) in OMOP format. We have mapped more than 90% of the available source variables to OMOP standards, relying on the Participant Provided Information (PPI) vocabulary for many concepts related to family histories and prior medical records. We also took advantage of the recently released Orphanet-to-SNOMEDCT mappings to capture various elements related to rare kidney diseases [2]. One particular (and yet unmet) challenge we have faced with this data is proper linking of patients through familial relationships. Many of the diseases captured within the ERKReg source data have a genetic component, and we are interested in cataloguing any familial connections between patients within these registries. We are currently using Fact Relationship table, though due to its limited integration with OHDSI tooling we are also in the preliminary phases of designing a rare disease extension to the standard OMOP tables that could handle this type of information, along with other useful data such as patient reported outcomes (PROs) and detailed clinical assessments of disease progression, both of which contain valuable information on the overall quality of life, disease severity, response to treatments, and other relevant factors in the rare disease research environment.

**METHODOLOGY:** We designed the Extract-Transform-Load (ETL) process in a way that is both flexible to accommodate different input files and datatypes, and simple to update. The ETL itself is executed via bash script and references an array of SQL queries that perform the necessary lookups and transformations in a PostgreSQL database. Critical to these processes is the semantic mapping file, which holds both the structural and conceptual logic necessary to convert source variables to associated domain-specific events with proper date references. In total, ERKReg mapped approximately 1500 unique source variables to standard concepts in OMOP CDM, capturing both events as well as the absence thereof. Mapping was done collaboratively by multiple team members within a shared and versioned spreadsheet, and coverage was monitored between versions using the OHDSI Ares application. The solution addressed several constraints related to data security and privacy, as well as resources available: (1) the OMOP CDM data needed to remain in Germany and (ideally) within a managed VM, (2) the managed VM selected could not support native Docker processes, and the proposed alternative uDock was not sufficient to host OHDSI tooling, and (3) ERKReg had already provisioned a VM for other Docker processes, and that server had space available for OHDSI tooling.