SPECIFIC AIMS

Kidney transplant (KT) care faces unprecedented challenges resulting from increased patient complexity, the need to expand acceptance of organs from non-standard donors, and challenges from broader organ sharing including longer cold ischemic times. Despite refinement in immunosuppression (IS) management over several decades, nearly all KT patients experience progressive loss of allograft function, leading to graft failure while still risking life-threatening IS-related complications. U.S. KT practice benefits from the unique availability of a national registry capturing some information on all U.S. recipients, but development of evidence-based IS management practice to optimize *long term KT function* has been limited by the need for detailed, longitudinal clinical data for large representative populations. Our previous NIDDK-supported R01 grant (Choosing IS regimens in kidney Transplant by Efficacy and Morbidity; **CISTEM**), leveraged integrated transplant registry, Medicare claims and national pharmacy clearinghouse data were used to assess the impact of *initial IS regimen* selection on key post-KT events: infections (pneumonia, sepsis, urinary tract infections), malignancy, new onset diabetes, as well as traditional metrics (acute rejection rate, allograft survival and patient death).[1] We developed a free web-based interface to assist transplant professionals and patients in shared-decision making about IS choice at the time of KT.[2] However, with IS being a *lifelong* requirement requiring dynamic adjustments, we seek to extend the CISTEM tools to improve *long-term* outcomes by applying state-of-the-art machine learning (ML) algorithms using both the CISTEM data set and a novel robust longitudinal dataset incorporating multicenter electronic medical records. To inform patient management in the current complex transplant ecosystem, we propose the following investigation to enhance our prior work (**CISTEM2)** with the following **Specific Aims**:

Aim 1: Recognizing the data continuum gaps in CISTEM including the lack of longitudinal and granular clinical observations and outcome measures, such as serum creatinine levels (for computation of estimated glomerular filtration rate, eGFR), tacrolimus drug levels, measures of viremia and viruria, biopsy reports (for defining acute rejection (AR) and other forms of graft injury), malignancy diagnoses, rehospitalization events, we will establish a novel, robust and curated dataset (CISTEM2 Dataset) integrating transplant registry data with multi-site electronic medical records (EMRs), administrative claims, and social determinants of health data for renal transplant patients leveraging the PCORnet infrastructure.

Methods: The Great Plains Collaborative (GPC) is one of the largest PCORnet networks, covering 13 healthcare systems (including 12 large KT centers with 12,448 KTs) across the United States Midwest region. GPC has established a centralized data model (CDM) infrastructure integrating electronic medical records from all 13 sites which can be linked to transplant registry, administrative claims and various open social determinants of health databases. Utilizing granular clinical data captured in EMRs and linked transplant registry data, we will expand the statistical analysis performed in CISTEM (multivariable propensity and Cox proportional hazard models with time-varying covariates) to more responsive and clinical meaningful endpoints such as percentage drop in eGFR and the validation of computed phenotypes for key clinical events.

Aim 2 We will extend CISTEM by developing longitudinal machine learning algorithms to dynamically suggest IS strategies that optimize renal allograft function (at 1-, 3- and -5-years post-KT), reduce cost, and limit IS comorbidities identified by patients as contributors to diminished quality of life.

Methods: Machine learning models will incorporate baseline characteristics and initial ISx management to improve prediction of allograft survival at 1,3,5 years posttransplant. We will then develop multiple machine learning models incorporating temporal information (structural modeling, landmark boosting model and reinforcement learning model with sparse reward) for a) rolling prediction of allograft survival in the next 1 to 5 years; b) identification of optimal IS management strategies yielding the highest likelihood of allograft survival; c) posttransplant patient focus groups will identify outcomes of highest interest for model development.

Aim 3: Validate the predictive models refined in Aim 1 and the ML models developed in Aim 2, using two additional PCORnet sites (Total Sample: 40535 KTs). Using computable phenotypes developed using the CDM model, our temporal-aware machine learning models will be evaluated to determine reliability in predicting long-term graft function in independent populations. This aim will additionally incorporate data from three PCORI networks in a distributed learning model to refine the first dynamic clinical decision tool for longitudinal IS management after kidney transplant.

Significance: The ability to accurately balance the risk of acute rejection, patient and graft survival, and risk of IS related complications after kidney transplant, based on highly granular, multicenter, longitudinal clinical from real-world patient experience, will allow patients and physicians to optimize IS choices in a more personalized, patient focused, cost-effective, and dynamic manner.