DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:
Summarize the types and estimated amount of scientific data expected to be generated in the project,

Leveraging existing cloud infrastructure at GPC coordinating center [1], we will establish a novel, robust and curated database (CISTEM2 database), which will integrate transplant registry data (i.e., the Scientific Registry of Transplant Recipients or SRTR registry) with multi-site electronic health records (EHRs), administrative claims (Medicare claims will be obtained from the Center of Medicare and Medicaid (CMS), and social determinants of health data (e.g., American Community Survey, Area deprivation index) for renal transplant patients leveraging the PCORnet and GPC data infrastructure. We estimated about 41,000 kidney transplant patients will be included in this observational cohort.

B. Scientific data that will be preserved and shared, and the rationale for doing so:
Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

The access to CISTEM2 database will be subject to the study-specific Data Use Agreement (DUA) established between the Data Coordinating Center at MU (DCC) and all participating sites (PCORnet sites and SRTR) during the initial period. Except as authorized under the respective DUA or otherwise required by law, governmental regulation, or any governmental entity with jurisdiction, all approved research team members with data access agree to retain control over the data and not disclose, release, sell, rent, lease, loan, or otherwise grant access to the data to any third party, except Authorized Persons, without the prior written consent of data providers. Given the amount of granular information available in the CISTEM2 database, there may still be risks of re-identification even with the de-identified data. Thus, the DUA prohibited public access to the databases in order to better protect human subjects and participating institutions.

C. Metadata, other relevant data, and associated documentation:
Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

All the source data, i.e., the PCORnet EHR CDM [2], Medicare claims [3], and SRTR registry [4] all released publicly accessible data dictionaries. Upon integration of the different data sources, we will also create CISTEM2 data dictionary and will make it publicly accessible.

Element 2: Related Tools, Software and/or Code:
State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

The CISTEM2 dataset will be access and analyzed using Snowflake SQL, R, Python and other approved analytical tools. All development codes will be made available in our project github repository page [5] no later

than when publications are submitted. The github page will be publicly assessable and will be hosted and maintained for at least 5 years after the grant award ends. This project will also result in further improvement of current CISTEM1 complication calculator, which will be made publicly accessible.

Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

The PCORnet EHR and Medicare claims data follows PCORnet CDM data model standards, both CDM models are required to model local clinical concepts to controlled terminologies (e.g., ICD, CPT, LOINC, SNOMED, etc.). For transplant-specific data elements provided in the SRTR registry, we will map them to standard terminologies as much as possible while maintaining the source concepts.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see Selecting a Data Repository).

The CISTEM2 database will be hosted in a HIPAA-compliant cloud-based data lake, which can only be accessed from approved virtual machines in compliance with security and privacy regulations required by hosting sensitive data such as Medicare claims and registry data.

B. How scientific data will be findable and identifiable:

Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Although we are leveraging the Datavant Privacy-Preserving-Linkage software to link patients across multiple resources, all participating PCORnet sites are still expected to keep a local mapping between the masked identifier and the real identifier for each patient. However, a full IRB application and separate study-specific DUA will be required to access that mapping.

C. When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Access to CISTEM2 database will be restricted to the study period plus an additional 5 year after the end of the study period to facilitate subsequent requests to validate and reuse the database for new research projects. While we cannot guarantee external and institutional support in perpetuity, the goal of the GPC, PCORI, and PCORnet is to make PCORnet data available as a reusable, national resource for all investigators (within and outside the network).

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

Per PCORnet policy, except as authorized under the respective DUA or otherwise required by law, governmental regulation, or any governmental entity with jurisdiction, all approved research team members with data access agree to retain control over the data and not disclose, release, sell, rent, lease, loan, or otherwise grant access to the data to any third party, except Authorized Persons, without the prior written

consent of data providers. The CISTEM2 governance body may permit reuse with appropriate application and agreement. Aggregate results within the proposed study scope can be distributed and published following cell-size suppression rule.

B. Whether access to scientific data will be controlled:
State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).)

Per PCORnet policy, except as authorized under the respective DUA or otherwise required by law, governmental regulation, or any governmental entity with jurisdiction, all approved research team members with data access agree to retain control over the data and not disclose, release, sell, rent, lease, loan, or otherwise grant access to the data to any third party, except Authorized Persons, without the prior written consent of data providers.

C. Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

We will leverage PCORnet's selected privacy-protecting linkage solution by Datavant [6] to securely link patient data from all study sites and data sets without exchanging any PHI information. All PCORnet sites have an existing Site License Agreement (SLA) in place with Datavant and already have access to the Datavant De-ID software. All HIPAA identifiers will be masked except the retention of real dates (e.g. date of birth, date of transplant) which are critical for the analyses. Each patient record will include an encrypted token (a hash) that can be reproducibly generated from identifiable patient information from different sites but cannot be reverse-engineered to reveal the original information. This approach was previously successfully used to link patient data from different sources in PCORnet-based and other studies. The final integrated CISTEM2 database for authorized researchers to access will be fully de-identified, and they can only be accessed through approved virtual environments using approved analytic tools.

Element 6: Oversight of Data Management and Sharing:
Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

Uninterrupted monitoring of accesses and activities happen on the CISTEM2 database following established best practices [1] that have been implemented at system level. Researchers' accesses to CISTEM2 database are required to be reviewed on an annual basis according to current data use protocols.

[1] L. R. Waitman, X. Song, D. L. Walpitage, D. C. Connolly, L. P. Patel, M. Liu, et al. Enhancing PCORnet Clinical Research Network data completeness by integrating multistate insurance claims with electronic health records in a cloud environment aligned with CMS security and privacy requirements. Journal of the American Medical Informatics Association 2021. DOI: 10.1093/jamia/ocab269
[2] Common Data Model (CDM) Specification, Version 6.0. Published by: Patient-Centered Outcomes Research Institute. https://pcornet.org/data/
[3] Chronic Conditions Data Warehouse. Published by: Center of Medicare and Medicaid. https://www2.ccwdata.org/web/guest/home/
[4] The Scientific Registry of Transplant Recipients SRTR Data Dictionary. Published by: Chronic Disease Research Group of the Hennepin Healthcare Research Institute. https://www.srtr.org/requesting-srtr-data/saf-data-dictionary/
[5] gpcnetwork github. Published by: The Greater Plain Collaborative. https://github.com/gpcnetwork
[6] D. Kiernan, T. Carton, S. Toh, J. Phua, M. Zirkle, D. Louzao, et al. Establishing a framework for privacy-preserving record linkage among electronic health record and administrative claims databases within PCORnet(®), the National Patient-Centered Clinical Research Network. BMC research notes 2022 Vol. 15 Issue 1 Pages 337-337.