

Data Management and Sharing Plan

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.

This study aims to collect a comprehensive dataset from renal dialysis patients and matched healthy controls. The types of data include demographic information (age, gender, ethnicity), laboratory results (blood work, urine analysis), clinical observations (symptoms, medical history), and clinical disposition (treatment outcomes, follow-up data). We anticipate collecting data from 250 affected participants and 250 matched healthy controls, resulting in a substantial dataset that will inform our understanding of renal dialysis outcomes and potential improvements to care.

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.

All collected data (demographic, laboratory results, clinical observations, and clinical disposition) from both affected participants and healthy controls will be preserved and shared. The rationale for sharing these data is to contribute to the broader scientific community's understanding of renal dialysis, facilitate research collaborations, and accelerate discoveries that could improve patient outcomes. Sharing these data aligns with the principles of transparency, reproducibility, and advancing public health.

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.

To ensure the interpretability and usability of our dataset, we will also make available metadata (data descriptors, collection methods), study protocols, data collection instruments (questionnaires, survey tools), and any other relevant documentation (informed consent forms, ethical approvals). This comprehensive approach to data sharing will enable other researchers to understand the context, quality, and potential limitations of our data.

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.

Specialized tools and software may be needed for the analysis and manipulation of shared scientific data. Specifically, statistical analysis software (e.g., R, SPSS) and data visualization tools (e.g., Tableau, Power BI) will be necessary. Additionally, programming languages like Python, with libraries such as Pandas and NumPy, could be used for data manipulation and analysis. These tools are widely available in the academic and research communities.

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.

To ensure interoperability of datasets and resources, we will apply common data standards to our scientific data and associated metadata. Specifically, we will adhere to the Clinical Data Interchange Standards Consortium (CDISC) standards for clinical trial data, which include operational data models (ODM) for data capture and dataset structures for analysis. For demographic and laboratory data, we will use standardized vocabularies such as SNOMED CT and LOINC, respectively. These standards facilitate the aggregation of data from different studies, enhance data quality, and support more effective data sharing and reuse.

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.

The scientific data and metadata arising from this project will be archived in the National Institutes of Health (NIH)'s National Library of Medicine (NLM) database, specifically through the Database of Genotypes and Phenotypes (dbGaP) for sensitive data or the NIH's Data Repository for non-sensitive data. These repositories are designated by the NIH for the sharing of scientific 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.

To make our scientific data findable and identifiable, we will assign a persistent unique identifier (e.g., DOI) to the dataset upon publication. The data will also be indexed in standard databases and made discoverable through search interfaces provided by the repositories.

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 the time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

The scientific data will be made available no later than the time of publication of the main findings or at the end of the performance period, whichever comes first. The data are expected to remain accessible for a minimum of 10 years after the project period ends, as per the repository's retention policies and our institutional agreements.

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.

While we aim to maximize data sharing, certain factors may limit the extent of sharing. For instance, to protect participant privacy, identifiable information will be removed, and measures such as data aggregation will be used where necessary. Additionally, considerations related to intellectual property and potential commercial applications may influence how some aspects of the data are shared.

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).

Access to sensitive portions of the dataset, particularly those that could potentially identify individual participants, will be controlled through mechanisms provided by the designated repository (e.g., dbGaP). Researchers wishing to access these data will need to apply and receive approval, demonstrating their intention to use the data only for research purposes and outlining measures they will take to protect participant confidentiality.

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).

To safeguard the privacy, rights, and confidentiality of our human research participants, we will implement several protective measures: (1) de-identification of datasets by removing direct identifiers; (2) use of Certificates of Confidentiality to further protect against compelled disclosure; and (3) adherence to HIPAA guidelines for handling sensitive health information. Participants will also be fully informed about data sharing practices as part of the consent process.

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).

Oversight of data management practices, including data sharing and repository interactions, will be conducted by our institutional review board (IRB) in collaboration with our data management team. Regular audits will ensure compliance with data sharing policies, repository requirements, and ethical standards for protecting participant confidentiality. This oversight structure ensures that we maintain high standards of integrity and responsibility in managing and sharing our research data.