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 project will generate clinical research data from 500 individuals (250 renal dialysis patients and 250 matched healthy controls) recruited from multiple clinics. Data types to be collected include:
- Demographic data (e.g., age, sex, race/ethnicity, relevant medical history)
- Laboratory results (e.g., blood and urine tests, kidney function markers)
- Clinical observations (e.g., blood pressure, weight, dialysis modality, session details)
- Clinical disposition (e.g., hospitalization, mortality, end-stage renal disease progression, adverse events)

The estimated volume of structured data is approximately 2 GB in total, stored as tabular files (CSV or similar). No high-throughput sequencing or imaging data will be generated.

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 de-identified individual-level participant data relevant to the study aims—including demographics, laboratory values, clinical observations, and clinical disposition—will be preserved and shared. Sharing these comprehensive datasets will maximize the scientific utility and reproducibility of the research, supporting future meta-analyses, hypothesis generation, and validation studies.

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.

Metadata to be shared will include:

- Data dictionaries describing each variable, coding, and permitted values
- Study protocol and informed consent templates
- Case report forms (CRFs) and data collection instruments
- Descriptions of data processing and quality control procedures

These will facilitate interpretation and reuse of scientific data by other investigators.

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 data will be provided in standard, non-proprietary tabular formats (CSV, TXT), compatible with widely used statistical and data analysis software (e.g., R, Python, SAS, SPSS, Excel). No specialized or proprietary software is required to access or manipulate the data. Any custom code (e.g., for data cleaning) will be shared via a public GitHub repository, with links provided in the data documentation.

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.

Common data element standards recommended by the NIDDK and NIH (such as CDISC and NIH Common Data Elements for kidney research) will be used whenever possible for variable definitions and coding. Data and metadata will be formatted in accordance with FAIR (Findable, Accessible, Interoperable, Reusable) principles. Data dictionaries and metadata will follow the Data Documentation Initiative (DDI) standard to enable interoperability with other datasets and resources. Where no consensus standard exists for a specific data element, the definitions and coding will be clearly documented.

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 will be deposited in the NIDDK Central Repository (https://repository.niddk.nih.gov/), a NIH-supported domain-specific repository for studies of diabetes, digestive, and kidney diseases.

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.

Upon submission, the dataset will be assigned a persistent digital object identifier (DOI) or accession number by the NIDDK Central Repository. Metadata records will be indexed and made searchable through the repository's interface and via NIH's data discovery platforms (e.g., NIH RePORTER, Vivli, DataMed).

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 data will be made available no later than the time of publication of the main study findings or at the end of the NIH performance period, whichever comes first. Data will remain available through the NIDDK Central Repository for at least 10 years after the project completion, in accordance with repository policies.

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.

All participants will be consented for broad data sharing. Data will be de-identified in compliance with HIPAA and other applicable regulations. There are no anticipated restrictions on data sharing beyond those

required to protect participant privacy and confidentiality. Data use agreements will prohibit attempts to re-identify participants or use the data for non-research purposes.

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).

Yes, access to individual-level data will be controlled. Researchers will be required to submit a data access request and sign a data use agreement via the NIDDK Central Repository. Summary-level data and metadata will be openly accessible.

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).

All shared data will be rigorously de-identified prior to release, following the HIPAA Safe Harbor method and expert determination as appropriate. Direct identifiers will be removed, and indirect identifiers will be reviewed to minimize re-identification risk. A Certificate of Confidentiality will be obtained for the project. Access to controlled data will require data use agreements specifying requirements for privacy, security, and prohibition of re-identification.

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).

Compliance with this DMSP will be monitored by the study's Principal Investigator and the designated Data Steward (a PhD-level project manager with data governance experience). Oversight activities will include regular reviews (quarterly) of data collection, de-identification, documentation, and repository submission activities. Compliance will also be reviewed annually by the institution's Office of Research Compliance and the NIDDK program officer, as required by NIH policy.