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 will collect renal dialysis data from multiple clinics. Demographic, laboratory results, clinical observations, and clinical disposition will be acquired from 250 affected participants and 250 matched healthy controls.

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.

Identifiable data will be de-identified prior to repository submission. Participant-level clinical data described in A will be preserved through deposition of the data in a controlled access public repository.

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.

The study protocol, data collection forms/case report forms, data dictionary, manual of operations, and a glossary of domain-specific terms will be submitted.

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 clinical data will be analyzed with custom R code and visualized with the ggplot2 package. R packages are all freely available via R CRAN. All code will be shared via a tagged GitHub repository and a readme.md file for the project describing the workflow, relationship between code, instructions, and parameter choices for selected tools.

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.

Participant age, sex, ethnicity, height, weight, socioeconomic status, and dialysis data will be collected using the common data elements (CDEs) from the National Institutes of Health (NIH) CDE Repository.
(1) Demographics (NLM ID: Xyc4G1BHte)
(2) Standing Height (NLM ID: gaz3k9xh1da)

(3) Weight (NLM ID: lLbYoUaBc)
(4) Socioeconomic Status (NLM ID: 7kpJeKE7P)
(5) Dialysis (NLM ID: 71WP2zp2ox)

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 study data will be submitted to a generalist repository that is participating in the NIH Generalist Repository Ecosystem Initiative, Vivli. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Central Repository currently accepts datasets only from NIDDK-funded, large, multi-site clinical consortia and not R01 Studies. No NIH or domain-specific repositories for "renal disease" or "renal dialysis" were found. This study will not generate genomics data, so data will not be submitted to the database of Genotypes and Phenotypes. Therefore, a generalist repository seems like the most appropriate location to deposit the 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.

We will submit the metadata associated with the datasets to Vivli and the journal Renal Failure. The repository and journal will provide metadata, persistent identifiers, and long-term access for open and controlled access

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.

Scientific data will be shared as soon as possible. Scientific data included in published manuscripts will be available at the time of publication; all other generated scientific data will be shared no later than the end of the award. The study data will be stored in the repository for at least 5 years.

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.

The study datasets will be collected with the following informed consent: Health/Medical/Biomedical - The dataset can only be used for studying health, medical, or biomedical conditions and does not include the study of population origins or ancestry.

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).

To maximize the appropriate sharing of scientific data and protect research participants' privacy and confidentiality, reuse of this dataset should use the following Data Use Limitations (DULs) under Controlled Access that is made available by a data repository only after approval of the request by the Vivli independent review panel process. Health/Medical/Biomedical - The dataset can only be used for studying health, medical, or biomedical conditions and does not include the study of population origins or ancestry. IRB Approval Required (IRB) - The requesting institution's IRB or equivalent body must approve the requested use.

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 protect research participants' privacy and confidentiality, data submitted to the repository will not include personally identifiable information such as names or addresses. Additional protections, such as the approach for managing Health Insurance Portability and Accountability Act identifiers, will be used for de-identification or to provide a limited data set to minimize the risk of participant reidentification.

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).

The Principal Investigator for this project, Dr. ABC, will ensure that this Data Management and Sharing (DMS) Plan is followed. The institutional official (title and role), will be responsible for oversight of compliance with the accepted DMS Plan. Compliance will be evaluated annually during the award period and progress towards the plan's DMS activities will be included in the annual Research Performance Progress Report submitted to the NIDDK Project Officer. At the project conclusion, the final progress report will summarize how the DMS objectives were fulfilled and provide links to the shared dataset(s).