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 involves secondary data analysis on kidney magnetic resonance imaging (MRI) data from the database of Genotypes and Phenotypes (dbGaP) to determine parenchymal kidney volume. The primary dataset produced will be a clinical dataset containing estimated kidney volumes for subjects with available imaging data. Given the existing nature of the dbGaP dataset, we anticipate analyzing data from approximately 1,000 to 5,000 participants, depending on the availability of relevant MRI scans and consent for broad data sharing.

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 clinical dataset of estimated kidney volumes will be preserved and shared. The rationale for sharing this data includes facilitating future research into kidney health and disease, enabling the validation of our findings by other researchers, and contributing to the broader scientific understanding of kidney volume as a biomarker for health and disease. Sharing this data aligns with the principles of transparency, reproducibility, and collaboration in biomedical research.

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 will include participant demographics, MRI scan parameters, and kidney volume calculation methodologies. Other relevant data may encompass genetic information linked to kidney health from dbGaP. Associated documentation will consist of study protocols, data collection instruments (if applicable), and detailed descriptions of the analytical methods used for kidney volume estimation.

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 needed to access or manipulate shared scientific data include MRI analysis software (e.g., 3D Slicer, ITK-SNAP) for kidney volume calculations and statistical packages (e.g., R, Python libraries like Pandas and NumPy) for data analysis. These tools are widely available in the research community, with some being open source.

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 scientific data and associated metadata will apply common data standards to enable interoperability of datasets and resources. Specifically, we will adhere to the NIH-supported data standards for clinical and imaging data, such as those promoted by the National Cancer Institute's (NCI) Clinical Data Interchange Standards Consortium (CDISC) for clinical data and the DICOM standard for MRI scans. These standards ensure that our data set can be easily integrated with other research findings and databases.

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 database of Genotypes and Phenotypes (dbGaP), ensuring compliance with NIH policies on data sharing and accessibility.

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.

The dataset will be made findable through dbGaP's search functionality, utilizing keywords related to kidney volume, MRI, and relevant study identifiers. Each participant's data will be linked via a unique identifier that protects privacy while allowing for longitudinal tracking.

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 clinical dataset of estimated kidney volumes will be made available no later than the time of publication of the primary results or at the end of the performance period, whichever comes first. The data will remain accessible for a minimum of 10 years post-project completion to maximize its utility for future research.

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.

Given that all participants have provided consent for broad data sharing, the primary factor affecting access is adherence to dbGaP's guidelines and requirements for data access and use. There are no anticipated limitations on the extent of data sharing beyond those standard protections.

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 the scientific data will be controlled through dbGaP, requiring users to register and agree to terms of use that protect participant privacy and 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).

Participant privacy, rights, and confidentiality are protected through de-identification of the dataset and adherence to dbGaP's policies on data sharing and access. The use of unique identifiers without direct links to personal identifiable information (PII) further ensures participant confidentiality.

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 Data Management and Sharing Plan will be monitored and managed by the project's Principal Investigator in conjunction with the institution's data management office. Regular oversight meetings (at least bi-annually) will ensure that all aspects of data management, sharing, and protection are adhered to as outlined in this plan. The project team, including a designated data manager, will be responsible for implementing the plan, maintaining records of data access and sharing, and addressing any issues or concerns related to data management and sharing.