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.

The project will generate demographic, clinical, and neuroimaging data from 220 participants (110 affected youth and 110 matched healthy controls). Demographic data will include age, sex, ethnicity, and other relevant variables. Clinical data will include psychiatric diagnoses, symptom ratings, and relevant medical history. Neuroimaging data will comprise structural MRI, functional MRI (fMRI), and proton magnetic resonance spectroscopy (1H fMRS) data. For each participant, approximately 2-5 GB of imaging data will be generated, yielding an estimated total dataset size of 500-1,100 GB.

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 deidentified raw and preprocessed imaging data (structural MRI, fMRI, 1H fMRS), along with associated demographic and clinical data necessary to interpret the imaging, will be preserved and shared. Sharing this comprehensive dataset supports reproducibility, enables secondary analyses by the broader research community, and aligns with NIMH and NIH policies to maximize the value and impact of publicly funded 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 to be shared will include imaging acquisition parameters, scanner characteristics, preprocessing pipelines, and quality control metrics. Study protocols, data dictionaries, case report forms, and codebooks for clinical and demographic variables will also be made available. Documentation will include informed consent language for data sharing, as well as manuals describing clinical assessments and imaging protocols.

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.

Standard neuroimaging file formats (e.g., NIfTI for MRI/fMRI, and LCModel/GE for 1H fMRS) will be used. Data can be accessed and analyzed using widely available software such as FSL, SPM, AFNI, and LCModel. Links to freely available software and any custom code or preprocessing scripts will be provided in the repository. No proprietary software is required for basic data access.

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.

Data will comply with the Brain Imaging Data Structure (BIDS) standard for organizing and describing neuroimaging and related data, which facilitates interoperability and reuse. Clinical and demographic data will be harmonized using the NIMH Data Archive (NDA) Common Data Elements where applicable. Metadata standards published by the NDA will be applied to ensure consistency and interoperability.

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.

Scientific data and metadata will be archived with the NIMH Data Archive (NDA), a domain-specific, NIH-approved repository for mental health-related human subject 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.

Data will be indexed and discoverable via the NDA, with each dataset assigned a persistent unique identifier (DOI or NDA Study ID). Each subject will be associated with a Global Unique Identifier (GUID) in accordance with NDA requirements.

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.

Data will be submitted to the NDA no later than the time of first publication of findings or at the end of the performance period, whichever comes first, in accordance with NIH policy. Data will remain available for at least 5 years after completion of the project, or as required by NDA 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 provide informed consent for broad data sharing. Data will be deidentified prior to sharing. Access and reuse may be governed by NDA Data Use Certification Agreements to protect participant confidentiality. There are no anticipated additional limitations on data sharing beyond those required for privacy protection.

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 data will be controlled through the NDA. Researchers must submit a Data Access Request and agree to NDA terms, including data use certification and compliance with participant privacy protections.

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 data will be deidentified in compliance with HIPAA and NDA requirements prior to deposit. Only data elements necessary for generating the NDA Global Unique Identifier (GUID) will be retained in a secure, separate location and not shared. Data sharing will align with the terms outlined in participants' informed consent documents. Institutional Review Board (IRB) approval and oversight will ensure continued protection of participant rights and 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).

The Principal Investigator (PI) will be responsible for ensuring compliance with this Data Management and Sharing Plan, in collaboration with the study's Data Manager and Institutional Data Steward. Oversight will include quarterly internal reviews of data management practices, annual reporting to the IRB, and verification of timely data submission to the NDA. The Office of Research Compliance will provide additional oversight and support as needed.