**Data Management and Sharing Plan**

**Element 1: Data Type**

1. **Types and amount of scientific data expected to be generated by this project:**

All data will be obtained from research databases that are created from electronic health records, insurance claims, pharmacy dispensing, insurance coverage, and other administrative data systems in the four participating health systems. We anticipate the final study data base will include approximately 800,000 episodes of antidepressant treatment between 2010 and 2022 for approximately 250,000 health system patients. For each included antidepressant episode, data elements will describe patients’ demographic characteristics, previous and current treatment with antidepressant medications, co-occurring conditions, environmental/neighborhood characteristics, prior and current use of other mental health services, and responses to standard questionnaires regarding mental health symptoms and social needs.

1. **Scientific data that will be preserved and shared, and the rationale for doing so:**

To the greatest extent possible, we plan to preserve and make available the data that support the peer-reviewed publications that result from this research project, although there may be restrictions on specific data elements as outlined under Element 5 below. This includes the specific data elements that are required to carry out the analyses presented in such publications. This will enable replication of the findings in these publications and facilitate inclusion of such data in subsequent projects such as in meta-analyses or pooled analyses of individual-level data. Such data will be made available in deidentified datasets.

1. **Metadata, other relevant data, and associated documentation:**

In addition to the data described above, we will make available data dictionaries, including variable names and formats, and numbers of participants on whom a given data element is available. These metadata may refer to data that are not included in published analyses, so researchers may determine whether analyses or hypotheses of interest may be pursued within this research project.

**Element 2: Related Tools, Software and/or Code:**

The data described above will be made available in a format that can be read by common statistical software (SAS, R). We will also share programming code to generate study variables from data warehouse tables following the HCSRN/PCORnet/Sentinel common data model as well as statistical programming code used for analyses in primary study publications.

**Element 3: Standards:**

The original source data from which study data will be derived follow health care industry standard coding standards, including those to facilitate interoperability across health care systems such as those of Health Level Seven International (HL7). Examples of standard formats under which clinical data are coded include the International Classification of Diseases versions 9 or 10 (ICD9/10) for diagnoses; , Current Procedural Terminology (CPT) and Healthcare Common Procedure Coding System (HCPCS) codes for procedures; and National Drug Codes (NDC) for medications. These codes are generally used in creation of analytic datasets, although they may be combined to define final analytic variables.

**Element 4: Data Preservation, Access, and Associated Timelines**

1. **Repository where the scientific data will be archived:**

We plan to make datasets and associated metadata and programs as described above available through a publicly accessible data repository, the NIMH Data Archive. The data repository will be accessible through a web interface.

1. **How the scientific data will be findable and identifiable:**

Availability of datasets and associated data, and the associated uniform resource locator (URL) will be identified in publications, project-specific websites, and presentations that may refer to published data, as appropriate. We will ensure that the landing page for accessing datasets will be publicly available and searchable through standard internet search procedures.

1. **When and how long the scientific data will be made available:**

Data associated with peer-reviewed publications will be made available through the above repository at the time of acceptance for publication of the final published version of peer-reviewed manuscripts for each of the study aims. As this project may generate multiple publications, there will likely be multiple deidentified datasets that are made available at different times, corresponding to different publication acceptance dates.

We anticipate that data will be available for at least five years beyond the project end date. At that time, we will consult with NIMH Data Archive leadership regarding the value or utility of continuing to host this large database. If a dataset is ‘retired’ from public access, archived versions of the dataset may be available on request for verification of results.

**Element 5: Access, Distribution, or Reuse Considerations:**

1. **Factors affecting subsequent access, distribution, or reuse of scientific data:**

As participating health systems are covered entities subject to the Health Insurance Portability and Accountability Act (HIPAA) and other relevant regulations, data from this project may fall under the review of the health system privacy offices or committees. These offices or committees review proposed data sharing that may expose data from large numbers of health system windows to assess and prevent reuse or redisclosure of data that may have the potential to compromise participant confidentiality or to expose proprietary information about business practices of participating health systems or health systems’ business associates.

Data to be shared will contain no explicit identifiers (names, Social Security numbers, health plan member numbers) and contain no implicit identifiers as specified by HIPAA. Nevertheless, specific combinations of study data elements may create risk of reidentification. This typically occurs when specific health events that may be ascertained from other sources occur among individuals with uncommon combinations of demographic characteristics (e.g. psychiatric hospitalization in 2017 by Native Hawaiian/Pacific Islander and Hispanic female aged 13-17 living in Minnesota). Following procedures developed in our previous research, we will systematically assess risk of reidentification prior to sharing data and execute any redactions or alternations necessary to reduce risk of reidentification. Decisions regarding and redactions or alterations will attempt to retain maximal scientific benefit while protecting individual privacy. It is not possible to specify in advance what redactions or alterations may be necessary as this depends on patterns in the observed data.

1. **Whether access to scientific data derived from humans will be controlled:**

Given the sensitive nature of some data elements, the fact that data were extracted under a waiver of consent, and the potential for reidentification (see 5a above), access via NDA will be limited to a Controlled Access Permission Group. Users requesting access will be required to document institutional approval/sponsorship and provide formal institutional assurance that the user will not re-disclose data or attempt to reidentify individuals.

1. **Protections for privacy, rights, and confidentiality of human research participants:**

This project will be reviewed by a Kaiser Permanente inter-regional Institutional Review Board (IRB), and only activities approved by the IRB will be carried out. We will request and expect to receive a waiver of the usual requirement for informed consent. Members of participating health systems are notified upon enrollment that their data from electronic health records and related databases may be used for research purposes. Individuals may opt out from research use of their data.

As described in 5a above, all data to be shared be deidentified and thoroughly assessed to mitigate risk of reidentifiction.

**Element 6: Oversight of Data Management and Sharing:**

While this research project is active, data management and sharing will be under the oversight of the Principal Investigator and research project team. Once research activities have ended, oversight will continue with the Principal Investigator, or a successor designated by the Director of the Kaiser Permanente Washington Health Research Institute.