UCSF DREAM Cohort
Reynolds/Woodruff
1UG3CA265845
Data Management and Sharing Plan

In alignment with Findable, Accessible, Interoperable, Reusable (FAIR) principles and in accordance with final NIH Data Management and Sharing (DMS) Policy to promote the management and sharing of scientific data generated from NIH-funded or conducted research, we will ensure efficient data sharing and dissemination of DREAM Cohort data. We are committed to making our findings publicly available as soon as data emerge but within the confines of our Institutional Review Board (IRB) regulations. The outline below is based on new guidance from the NIH on the recommended Elements of Data Management and Sharing Plan template (https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#ex).

Data Type:

The proposed research will include data from approximately 10,000 subjects being screened for environmental chemicals during pregnancy and followed over time via medical record review and linkage to cancer, hospital, and vital records registries to determine intermediate cancer risk. This application requests support to collect statewide morbidity and mortality data from sources such as cancer registries, vital records, OSHPD (hospitalization) registries as well as ecologic public use data on neighborhood social and environmental characteristics.

In addition to medical record and registry linkage data, the final dataset will include self-reported demographic and behavioral data from interviews with subjects and laboratory data from blood and urine specimens provided. We will have measurements of 200 chemicals exposures in approximately 1500 biospecimens.

The final dataset will be stripped of identifiers prior to release for sharing. However, we will make the data and associated documentation available to users only under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed.

Related Tools, Software and/or Code:

An account or subscription may be required to access or manipulate shared data, for example SAS, STATA, R. While R is opensource statistical software, SAS and STATA require licenses. The shared data will be available in Excel or CSV file formats, appropriate for all programs, as well as SAS datasets. These tools are likely to remain available for as long as the scientific data remain available.

Standards:

We will follow common data standards for creating data dictionaries and protocols. We will document changes to data files, including statistical code.

Data Preservation, Access, and Associated Timelines:

De-identified data will be preserved and shared utilizing UCSF's Research Analysis Environment (RAE) and/or Dryad, in which UCSF is an institutional member. UCSF's RAE is a secure data hosting and computing service for UCSF researchers and their internal or external collaborators. The Dryad Digital

Repository is a curated resource that makes research data discoverable, freely reusable, and citable. Dryad provides a general-purpose home for a wide diversity of data types.

Data will be findable via standard indexing tools.

The data will be made available to other users as soon as possible, but no later than the time of an associated publication or end of the project period, whichever comes first and in accordance with UCSF RAE requirements, journal policies. The data will be available for as long as it is useful for the larger research community, institutions, and/or the broader public.

Access, Distribution, or Reuse Considerations:

DREAM Cohort consent documents include an explanation of data sharing so subjects are informed that de-identified information and biospecimens may be shared with other collaborators, both within and outside of UCSF.

Data shared with other researchers will be de-identified. This project will have a Certificate of Confidentiality.

All scientific data derived from humans in the DREAM Cohort will only be made available after data use agreements are approved by the UCSF Office of Sponsored Research.

Oversight of Data Management and Sharing:

We will employ rigorous methods to ensure compliance to the Data Management and Sharing Plan outlined in this document. Drs. Reynolds and Woodruff (PIs), and the research team have developed data management strategies and protocols to ensure QA/QC of datasets while protecting confidentiality of study participants. Lynn Harvey, Senior Data Analyst, will review this plan with PIs and the study team each time data is ready to be shared with other researchers, and/or members of the public. Changes to the plan will be documented and old versions archived in our UCSF server.