Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D1BS7N

Title: Effects of Placental Dysfunction on Brain Growth in Congenital Heart Disease

Creator: Cynthia Ortinau - **ORCID:** 0000-0003-0851-0796

Affiliation: Washington University in St. Louis (wustl.edu)

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-185

Grant: https://grants.nih.gov/grants/guide/pa-files/PA-20-185.html

Template: NIH-Default DMSP

Start date: 09-01-2023

End date: 08-31-2028

Last modified: 07-08-2024

Copyright information:

Effects of Placental Dysfunction on Brain Growth in Congenital Heart Disease

Data Type

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.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

The proposed project will involve human participants and will acquire demographic, clinical, fetal/placental magnetic resonance imaging (MRI), genomic, and placental tissue data from 225 pregnant women with a fetal diagnosis of CHD and 115 pregnant women with a healthy fetus. This study involves eight co-investigators from three sites, Washington University in St. Louis (WUSTL), Boston Children's Hospital (BCH), and The Hospital for Sick Children (SickKids). Table 1 describes the type and volume of data that will be generated at each site.

Table 1. Data Type and Amount of Data by Site

Data Category	Data Type	WUSTL	BCH	SickKids
Demographic	maternal age, race, ethnicity, fetal sex, socioeconomics (i.e. zip code, insurance type)	125 pregnant women/fetuses	113 pregnant women/fetuses	102 pregnant women/fetuses
Clinical	cardiac diagnosis, details of pregnancy and fetal complications, details of infant postnatal course	125 pregnant women/fetuses	113 pregnant women/fetuses	102 pregnant women/fetuses
MRI (2 per participant)	Regional fetal brain MRI measures, placental oxygenation measures, and MR oximetry	250 placental and fetal brain datasets	226 placental and fetal brain datasets	204 placental and fetal brain datasets
Genomic (trio samples)	whole exome sequencing	125	113	102
Placental tissue	qualitative clinical assessment and conofocal laser scanning microscopy	125	N/A	N/A

Demographic and clinical data will be captured by each site into the REDCap secure electronic data capture system (EDC) and will generate approximately 5 GB of data. Each user will be given role specific access to the EDC system, and access will be controlled by granting users individual usernames and passwords. MRI data will be captured by each site into the Central Neuroimaging Data Archive at WUSTL. Genomic data will be individual level data. Analysis of genomic DNA will link genotype to phenotypic information obtained as part of this study. Genomic data will include trio whole exome sequencing (WES) for a total of 20 TB data. Placental tissue will be maintained in the Pathology Department at WUSTL, with additional slides for conofocal laser scanning microscopy maintained in a secure laboratory environment managed by one of the co-investigators. Data generated from these samples will be entered into the REDCap EDC system.

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.

Demographic, clinical, and placental data entered into the REDCap EDC system will be cleaned, de-identified, preserved and shared. Appropriate measures such as assigning a study identification number and removal of all dates will be used for data de-identification and sharing, and informed consent forms will reflect those plans. MRI data will be preserved and shared via OpenNeuro after images are defaced. Raw data will not be shared due to the potential risk of identifying the fetus. For genomic data, the study will be registered to dbGaP. WES and phenotype data will be uploaded using the dbGaP submission guide and templates.

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.

To facilitate the interpretation and reuse of the data, a data dictionary will be generated and deposited into a repository along with all shared datasets. The data dictionary will define and describe all variables in the dataset. A README file will also be generated and deposited into the repository and will include instructions for data analysis.

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.

MRI data will be analyzed using methods developed by our research team. The software and algorithms used for fetal brain analysis is distributed as open source software and is shared through a GitHub repository (bchimagine). The placental MRI methods are also readily available and can be provided by contacting our research team.

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 be standardized whenever possible, shared data will be deidentified, and original data will be maintained at the investigator's institution. This research project will standardize all fetal brain MRI data using the Brain Imaging Data Structure (BIDS) recommendations. We will adopt similar methods for the placental MRI data. Genomic data will be in CRAM format with alignment to the latest reference genome and variant files in VCF format. Common Data Elements will be applied as relevant using the NICHD data standards.

Data Preservation, Access, and Associated Timelines

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; see [Selecting a Data Repository](#))

All de-identified demographic, clinical, and placental tissue data will be shared via the institutional repository Digital Commons@Becker, which provides metadata, persistent identifiers (DOIs), and long-term access. Digital Commons@Becker is a digital repository for presenting and preserving the scholarly work created at Washington University School of Medicine in St. Louis and is administered and maintained by the Bernard Becker Medical Library. Digital Commons@Becker is hosted by bepress and is backed up by S3 cloud storage to address preservation needs. The repository can host multiple file types, conducts regular audits for fixity and authenticity, and plans exist in instances of repository host change. Data will be made publicly available with controlled access under Data Use Agreement to ensure the safety of participants' data.

MRI data will be deposited into OpenNeuro. Digital content ingested to OpenNeuro is replicated multiple times and stored in geo-diverse locations on different media types. Datasets are audited systematically to ensure that the bits are maintained exactly as deposited, and a log of preservation actions is kept to help ensure the content's integrity. The repository is built using widely-adopted and actively maintained open-source data management tools. These tools permit changes to content to be tracked and "snapshots" to be made that uniquely identify specific points in the lifetime of each dataset. After an optional 36-month embargo period, all datasets are published into the public domain. Prior to being made public, access to a dataset is controlled through strict authentication policies and an isolated storage backend to further guard against unintended access. Metadata describing each dataset snapshot is indexed for searching, and copies of ingested content are provided via persistent DOIs minted for each version of a dataset.

The study will be registered in dbGaP at the time of data cleaning and quality control measure initiation. Genomic data will be shared and deposited into dbGaP according to GDS policy, which requires that genomic data must be submitted within 3 months following data generation and released within 6 months of data submission to the repository or at acceptance of the publication, whichever is first. We will work with our Genomic Program Administrator, John Ilekis (ilekisj@mail.nih.gov) to ensure adherence to processes for genomic data.

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.

DigitalCommons@Becker provides searchable study-level metadata for dataset discovery. DigitalCommons@Becker assigns DOIs as persistent identifiers, and has a robust preservation plan to ensure long-term access. Data will be discoverable online through standard web search of the study-level metadata as well as the persistent pointer from the DOI to the dataset. OpenNeuro also provides persistent DOIs minted for each version of a dataset. Datasets in dbGap will be findable through the study accession number which is a unique, stable, and versioned identifier (ID) that can be used in publications.

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 time of an

associated publication or end of the performance period, whichever comes first) and for how long data will be available.

All scientific data generated from this project will be made available as soon as possible, and no later than the time of publication or the end of the funding period, whichever comes first. The duration of preservation and sharing will be subject to the retention policies of the repositories mentioned above.

Access, Distribution, or Reuse Considerations

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. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

There are no anticipated factors or limitations that will affect the access, distribution, or reuse of the scientific data generated by the proposal. All research participants will be consented for broad data sharing.

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 request access to the data, researchers will use the standard process for DigitalCommons@Becker, OpenNeuro, and dbGAP. For Digital Commons@Becker, data will be available via controlled access using a Data Use Agreement to ensure the data requestor has a legitimate reason for requesting data and will not attempt to re-identification. The dataset in OpenNeuro will become publicly available under a Creative Commons CC0 license after a grace period of 36 months counted from the first successful version of the dataset. For dbGap, summary level data is open but credentialed user must apply for access to individual level data.

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).

In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing of data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local laws, regulations, and policies will be followed. Data will be deidentified by removing all 18 HIPAA identifiers prior to sharing, and the study will have a Certificate of Confidentiality from NIH.

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).

Lead PI Cynthia Ortinau, ORCID: 0000-0003-0851-0796, and the eight co-investigators from the three sites mentioned above who are directly engaged in the research will be responsible for day-to-day oversight of data management activities and data sharing. Dr. Ortinau will meet monthly with key study personnel to ensure the timeliness of data entry and review data to ensure the quality of data entry. Dr. Ortinau will ensure that the metadata are sufficient and appropriate and that the data management and sharing plan follows the FAIR data principles. Dr. Ortinau will report the DMS-related activities as outlined in this DMS plan in RPPR and request approval for a revised plan if there is any deviation from the approved DMS plan. At the project conclusion, the final progress report will summarize how the DMS objectives were fulfilled and provide links to the shared dataset(s).