# 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.*

Type of study: Interventional, Clinical Trial. A Phase 3, Randomized, Placebo-controlled, Double-Blind. Study of 180 mother-infant pairs in patients with early onset preeclampsia with primary outcome: days of maintaining pregnancy. 360 participants (3 sites) will be enrolled in the study upon diagnosis of preeclampsia but not beyond 34 weeks gestational age. The subjects will be followed during antepartum management and assigned to either the treatment or placebo groups. Infants will return for follow-up visits at 3, 6, and 12 months of age.

Clinical and laboratory data: Includes medical history, demographic information, and patient characteristics at time of enrollment. Other data will be extracted from each subsequent clinical encounter leading to delivery, this includes lab tests consistent with clinical management, physical exams (presence of headache, swelling, vision issues, etc), and blood pressure measurements. Fetal monitoring will include assessments of estimated fetal weight, and non-stress tests. Blood and urine will be collected at each visit to generate research laboratory data including results of urinalysis (proteinuria/creatinine ratio), analysis of angiogenic factors, uric acid, and glycemia. Pregnancy duration (prolongation days of pregnancy) and neonatal outcomes (birth weight, Apgar scores, head and chest circumference at birth, short term prognosis of neonate, NICU admission) will also be recorded. Infant follow-up studies will include anthropometric measurements, and neurodevelopment assessments such as the Bayley scale of infant development at 3, 6, and 12 months. These values will be reported as tabular data and stored in a REDCap secure electronic data capture system.

Clinical imaging: 2D images generated from Doppler studies of uterine, umbilical and middle cerebral arteries will be collected and de-identified raw images will be stored as .jpeg files. The presence of uterine artery notching will also be recorded. The pulsatility index for each image will be extracted as tabular data and stored in .csv files.3D images of the fetal thigh will be obtained and stored as de-identified .vol files. These images will then be used to measure total thigh, lean mass, and subcutaneous fat volumes. Volume measurement will be stored in .csv files. Real-time grayscale 2D clips of the fetal heart will be acquired to measure global sphericity and cardiac area percentile. Deidentified DICOM files of raw data and .csv files of cardiac measurements and percentiles will be stored.

### 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 subject-level clinical, laboratory, and clinical imaging data described in 1A will be preserved and shared. Shared data, including ultrasound images, will be deidentified, and original data will be maintained at the investigator's institution. Recruitment progress and final results will be documented at clinicaltrials.gov.

### 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.*

Protocols, informed consent forms to participate in the trial and for biological sample collection, data dictionary, and code book will be shared with the data through the repositories. For data submitted to DASH, variable-level metadata will be provided using the DASH Codebook, which is a templated data dictionary, and will include details of Common Data Elements, definitions, and standards used for data collection and sharing. For data submitted to MIDRC, the data will be mapped to the Medical Imaging and Data Resource Center Commons dictionary. Other study-associated documentation including image acquisition parameters and image analysis workflow diagrams will also be submitted to MIDRC as supplementary files associated with the study to facilitate interpretation of the scientific data.

**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.*

Clinical and laboratory data will be collected in the electronic data capture system (REDCap) and analyzed using open-source statistical software packages. R software will be used for statistical analyses and graphics. Doppler examinations will be performed using RM6C matrix 4D convex probe (Voluson E10, GE Healthcare) and V4-8 4D convex probe (Medison V20 Prestige, Korea) with the high-pass filter at 60 Hz. Spectral Doppler analysis of flow velocity waveforms in uterine and fetal blood vessels will be performed on the ultrasound machine (Voluson E10). For imaging, 2D .jpeg images will be directly exported from either the ultrasound machine or EMR image storage. 3D volumetric assessment will be performed with the 4DView 7.0 licensed software (Voluson, GE Healthcare). Analysis of real-time cardiac clips will be performed using the open source DICOM viewer ITK-SNAP, an application supported by Insight toolkits (Kitware Inc., Clifton Park, NY, USA).

**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.*

Raw data will be standardized to CDISC (Clinical Data Interchange Standards Consortium) format whenever possible. Medical laboratory data will be standardized using LOINC whenever possible. This research project will use relevant Common Data Elements from the NIH Common Data Elements Repository for data collection and relevant Medical Imaging and Data Resource Center (MIDRC) Commons dictionary for imaging data. File formats for each data type are described in Element 1A.

**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.*

DASH (clinical and laboratory data) Medical Imaging and Data Resource Center (MIDRC) (clinical imaging)

### 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.*

DASH creates a DOI for each study. We will submit a request to MIDRC to create a persistent identifier for the study and will list the other study persistent identifier on the repository study pages. Potential new users from other organizations can find out about our data and identify whether it could be suitable for their research purposes through summary information (metadata) that will be readily available on our study website, the repository websites, and clinicaltrials.gov.

### 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.*

Baseline data (defined as any data collected prior to any intervention), such as demographics, tabular clinical data (i.e. blood and urine tests), and images that require no additional analyses will be submitted within 4 months after enrollment is completed and released through the repositories 4 months after that. After the study is complete and unblinded, the study team will submit all remaining scientific data to the data repositories (DASH and MIDRC) and will update the RCT status to "complete" in clinicaltrials.gov. According to the proposed project timeline, the release of the remaining scientific data will \approximately coincide with the submission of results to clinicaltrials.gov, as mandated by FDAAA, and will occur by the end of award or publication, whichever comes sooner. Under the current repository policies, data will be preserved and available for the wider research community in perpetuity.

## 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.*

Data will be shared as allowed by the informed consents and institutional certification.

### 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).*

Clinical and laboratory data will be shared with controlled access in DASH for general research use. Deidentified data shared with MIDRC will be available via registered access.

### 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).*

For a prospective study like this one, consent procedures will include provision for data sharing to maximize the value of the data for wider research use, while providing adequate safeguards for participants. As part of the consent process, proposed procedures for data sharing will be set out clearly and current and potential future risks associated with this will be explained to research participants. Language informing the participant or legally authorized representative that residual biological specimens may be stored in a biorepository for other scientific investigations will be used in the consent forms. The informed consents will contain language permitting secondary use with broad data sharing under controlled access for general research use through established data repositories such as DASH (clinical and laboratory data) and MIDRC (imaging). Patients will be informed that they will not be contacted or re-consented for future sharing or accessing data through repositories. Privacy and confidentiality protections will be consistent with applicable federal, Tribal, state, and local laws, regulations, and policies. As the data is collected it will be digitally preserved in our secure institutional cloud storage environment and the PI's computer under strict password access. As this research data includes personal data relating to human participants, special measures will be followed to safeguard the security of data. We will follow robust policies for managing confidentiality and data security, consistent with legal, good practice and NIH policy requirements. This study will be compliant with ISO 27001 IT security standards. Our institution is ISO compliant (registration number 2374573.) Data will be deidentified before sharing in DASH and MIDRC by removing all 18 direct identifiers according to HIPAA Privacy Rule's Safe Harbor.

**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 contact PI for the project is Dr. Responsible at University of Pregnancy. Dr. Responsible will meet monthly with members of the project team, Dr. Doppler (University of City, clinical imaging data collection), Mr. Organized (University of Pregnancy, data manager), and Dr. Laboratory (University of State, processing of laboratory samples) to ensure that data collection, management and submission to repositories are compliant with this DMS Plan. Progress on this Plan will be communicated annually in the RPPR.