Derivative Data Agreement

1. What is derivative data?

Derivative data refers to new data generated from the processing and analysis of raw data or biological samples obtained within the study and shared with researchers. By definition, derivative data is data that has been enhanced through processing by researchers. Examples of derivative data include a new behavioral scale derived from subject data, values obtained from image processing, and determinations made from biological samples. Changes made to the dataset to facilitate its processing, such as multiple imputation algorithms and their results, are also considered derivative data. Simple calculations performed on the dataset that can be easily reproduced by other researchers do not qualify as derivative data.

2. Data Sharing Commitment

Researchers commit to sharing their derivative data with the study core once their project has concluded. This deadline is considered met when the objective for which the data was requested has been achieved. Researchers agree to submit the data to the statistical core following the detailed procedures provided.

3. Authorship and Ownership of Derivative Data

Derivative data belongs to the creators of the data; however, the consortium is responsible for its storage, management, and data clerking. Upon receiving the data, the data core will make the availability of this dataset public in response to reasonable requests from the entire consortium. To preserve the effort and authorship of the researchers, upon receiving a request for the use of the derivative data, the original data creators will be invited to join the new research team. Regardless of their decision to collaborate with the team, the creator of the derivative data will be obligatorily included as an author on any work that utilizes the data.

4. Protocol for Sharing Derivative Data

Once the primary objective has been achieved, the creators are required to share their derivative data with the study core. The data must be shared in an actionable format within a package containing three files: 1) the data itself, 2) a data dictionary for interpreting the results, and 3) a metadata file.

1. **Data**: The data should be shared in an electronically readable format and identified with the participant_id used in the original database (sharing the original data is not required, but the identifier must be included).
2. **Data Dictionary:** A file that provides definitions for the variables included in the shared data.

3. **Metadata**: The metadata file is crucial. It must include a header with the date of the last update, the author, and contact information. The body of the file should contain a "Methods" section detailing the entire procedure followed by the researchers to obtain the shared data. This section should be sufficiently detailed to allow the reader to replicate the procedure (please include software, versions, equipment details, etc.). The author identified in this file will become the responsible contact for all matters related to data authorship.

An example of the data sharing is included in: https://github.com/icalandri/data_derivative_protocol

---

Date
Researcher