

QUALITATIVE REMIX: UPDATE FOR NIH 2023 DMSP

DATA TYPE

Types and amount of scientific data expected to be generated in the project:

Three types of data will be collected -- survey data, interview data, and documents. Qualitative data will be collected using a semi-structured interview protocol. Audio recordings will be taken of the interviews. Data will be gathered from ClinicalTrials.gov, professional society websites (if publicly available without membership), published trial reports from PubMed, and institutional/practice websites. Code books from the content analysis of the collected documents will be created and added to the study database. This project is estimated to generate 500 Gigabytes of data. Raw data will be transformed by masking method to make all publicly available data will be de-identified. To protect research participant identities, summarized data will be made available for sharing.

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

All data will be retained for up to 5 years following completion of the study. Participants may elect to have data destroyed after completion of the study. After 5 years, all portable media containing recordings will be destroyed before discarding the physical media. Data will be de-identified and, after completion of this study, the email addresses will be deleted from all study records, creating an anonymous dataset for future research. Then, the data may be made available for re-use. Due to ethics and privacy issues associated with the personal, sensitive nature of the qualitative data, all identifiable data will not be shared with users outside of the research team. De-identified, summarized data will be shared in a structured format using comma separates value (.csv files) datasets.

A brief listing of 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.

Each transcript (both clean and coded and preserve as a TEXT file) and audio recording (preserved in .wav files, used and processed in .mp3 files) will be uploaded to the participant study record in REDCap. The final codebook, analytic memos, and field notes will be uploaded to the File Repository in REDCap, then transferred into the generalist repository Dryad for reuse. We will use Dryad's internal metadata schema and provide the necessary information for optimal Findability, Accessibility, Interoperability, and Re-use.

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

REDCap has the capability to export the collected data to comma separated value (.csv) files and SPSS, the software programs that will be used in this study for analysis. Transcription of the audio interviews will be completed by using software for audio segmenting, labeling, and transcription software called NVivo. All transcribed data will be translated into .csv files for re-usability. Transcripts will be uploaded to qualitative software on the local system for analysis. During the qualitative phase analysis, a codebook will be developed. All datasets from the qualitative research portion will be available in Dryad in .csv format.

If applicable, specify how needed tools can be accessed, (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team) and, if known, whether such tools are likely to remain available for as long as the scientific data remain available.

The software licenses for qualitative data coding (NVivo) and statistical analysis (SPSS) will be purchased.

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

Dryad uses the DataCite Schema for optimal data searchability. Our project will follow the DataCite Schema, along with a structured and detailed versioning system to keep records updated, accurate, and flexible.

Versioning of files will be controlled through the following naming system:

[data type]_[data name]_[YYYY-MM-DD]_v[xx.yy]

where [data type] describes the ExportData, Protocol, ICF, DataDictionary, InterviewGuide, CaseBook, etc.

where [data name] describes the file such as POP1, POP2, Case4Caring1, Case4Science2,Qual1, Quant2, etc.

where YYYY-MM-DD is the four digit year, two digit month, and two digit day the file is created

where 'xx' is the sequential number of major revision and 'yy' is the number of times the file is modified

DATA PRESERVATION, ACCESS, AND ASSOCIATED TIMELINES

Repository where scientific data and metadata will be archived:

All data sets that can be shared will be deposited in Dryad, which provides metadata, persistent identifiers (i.e., DOIs), and long-term access. Dryad is the institutional data repository supported by the University of California and all data is shared under a CC0 waiver, which makes the dataset(s) publicly available. Data will be made available as soon as possible or at the time of associated publication. Dryad datasets are backed up to Merritt, the UC's CoreTrustSeal-certified digital repository, for long-term storage and accessibility.

How scientific data will be findable and identifiable:

Dryad provides metadata, persistent identifiers (DOI), and long-term access. This repository is supported by the University of California and all data is shared under a CC0 waiver, which makes the dataset(s) publicly available.

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

Data will be made available as soon as possible or at the time of associated publication.

ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

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

Exclusive use of the data is needed to support a student researcher's dissertation project. Following completion of the dissertation project, secondary use of the qualitative research data is possible through the data repository.

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).

Access to scientific data can occur as soon as it is uploaded in the data repository.

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).

The participant name is never collected in Phase 1 of the study for the online questionnaires. During Phase 2, the participant will be referred by his/her name during the course of the interview and this will be recorded via audio. However, in the study report, an alias will be assigned to the cases and the participant's name will not be associated with the data in the study report.

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 student researcher, Jane Doe, will be responsible for data management of this project including reviewing, revising, and implementing the data management plan. This researcher will be required to review data on a weekly basis.

PLANNED RESEARCH OUTPUTS

PLANNED RESEARCH OUTPUT DETAILS

| Title | Type | Anticipated release date | Initial access level | Intended repository(ies) | Anticipated file size | License | Metadata standard(s) | May contain sensitive data? | May contain PII? |
|--------------------|---------|--------------------------|----------------------|--------------------------|-----------------------|----------------------------|--------------------------------------|-----------------------------|------------------|
| patient responses | Sound | 2028-01-24 | Open | DRYAD | GB 20 | CC NC SA 4.0 International | DataCite Metadata Schema Dublin Core | Yes | Yes |
| dates of interview | Dataset | 2028-01-24 | Open | DRYAD | GB 2 | CC NC SA 4.0 International | Dublin Core | No | No |
| time of day | Dataset | 2028-01-24 | Open | DRYAD | GB 1 | CC NC SA 4.0 International | Dryad Metadata Application Profile | No | No |