RURAL

Data Request Policy

Prepared by

The Statistical and Data Coordinating Center (SDCC)

V.1.0.20240605

Contents

1. **Overview**

The purpose of the policy established herein is to encourage and facilitate the sharing of RURAL data while providing guidelines that ensure its appropriate use.

Data definitions:

➢ Aggregate: Provides summary measures of data. This type of data requires additional biostatical support from the SDCC. Table shells may be requested.

➢ Completely de-identified individual: Provides de-identified data (no geographic measures smaller than county, no dates, no names).

➢ Restricted use: Contains individual data that may have identifiers. Appropriate IRB approval will be necessary.

Of note, RURAL PIs will be provided access to an analysis-ready de-identified data set which will be updated on a scheduled or as needed basis. These datasets will include an array of derived study variables and individual data elements. This data request policy is pertaining to additional data that is not yet available to RURAL PIs.

Figure 1. Data Request Process flow diagram
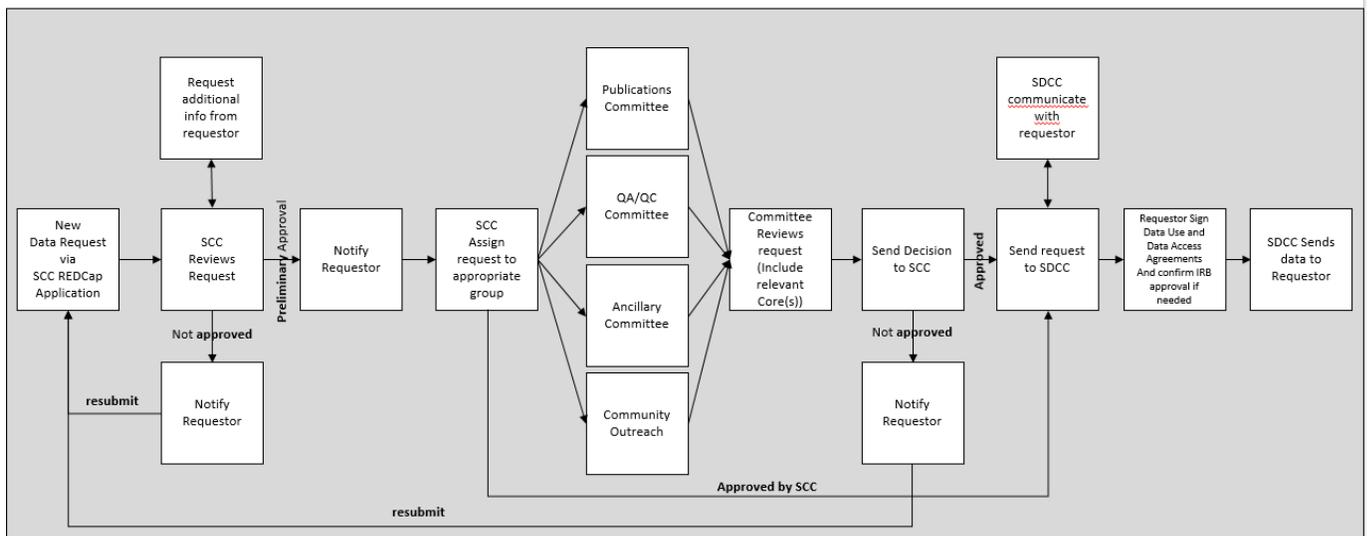
2. **Data Request Submission Process**

To ensure accuracy and consistency of reporting RURAL Study data, all data requests must be submitted to the Study Coordinating Center (SCC) via this REDCap link. All data requests will be assessed by the SCC in collaboration with the appropriate sub-committee.

The data request types include the following:

- Scientific Publications and Abstracts
- Ancillary Study Submissions
- Data for quality control (QC) assessment
- Dissemination of results to CABs/Community Stakeholders
- Other requests

3. **Data Request Types**

3.1. *Scientific Publications and Abstracts*

Data requests related to scientific publications and abstracts will be reviewed and approved by the RURAL Publications Committee. Please see the [Publications and Presentations Policy](#) for guidance on the approval process. Upon approval by the RURAL Publications Committee, the request will be submitted to the RURAL Scientific and Data Coordinating Center (SDCC) to be put in the Biostats queue for faculty and analyst assignments. Contributing factors to the amount of time it may take to receive a Biostats assignment include the order in which proposals are received, the number of manuscripts with active analyses, the number of faculty and analysts available, and the extent to which Biostats is conducting core activities (i.e., non-manuscript). Data requests related to scientific publications and abstracts should include the following information:

- Manuscript proposal with the following details:
  - Introduction (Brief rationale and background)
  - Research Hypothesis (Clear statement of scientific questions to be addressed)
  - Data analysis location (Local or SDCC)
  - Analysis plan and methods in consultation with the SDCC (Detailed description of proposed statistical analyses)
- Proposed mock-up tables and figures
- Completed variable selection sheet (local analysis)

3.1.1. Local Analysis:

For analysis performed at the local institution, data will be provided in de-identified formats with unique identifiers where appropriate. Data will be provided in comma separated values (CSV) and SAS, and R formats. The CSV file format is compatible with most statistical programs. Please note that for investigators that do not reside at a RURAL core, a data use agreement (DUA) between the University of Pennsylvania and the receiving institution is required prior to data sharing.

Any investigator requesting access to RURAL data should agree to the following conditions:

- You will not use or further disclose RURAL data for any purpose other than accomplishing the approved scientific aims for which you will be conducting analyses.
- You will abide by all requirements of the Data Request policy with regard to the use and disclosure of any RURAL data.

- You will implement administrative, physical and technical safeguards that reasonably and appropriately protect the confidentiality and integrity of any RURAL data as well as prevent its inadvertent use or disclosure.
- You agree to return or destroy any RURAL data you receive upon completion of analysis (e.g., defined by the acceptance of the manuscript reporting your findings) or at the end of any mandatory period of data archiving at your institution.

3.1.2. Central Analysis (SDCC)

For analysis performed at the SDCC, data will be provided in summary format based on the mock tables and figures.

3.2. *Ancillary Applications*

Preliminary data can provide additional information that will enable reviewers on the Ancillary Committee to evaluate the proposed ancillary study favorably. For data requests for new ancillary proposals, the SDCC requires 6-8 weeks' notice. The Ancillary Committee will be responsible for approving preliminary data requests for new ancillary proposals. Please see the [Ancillary Study Committee Policy](#) for details regarding the ancillary proposal submission process. Upon approval from the Ancillary Committee, the SCC will connect the investigator with the relevant cores and the SDCC for further discussion and to obtain the necessary data.

3.3. *Quality Control (QC) Assessment*

The SDCC has developed core-specific customizable dashboards to track receipt of data and/or samples and visual data summaries for quality control assessment. Cores may require access to data from another core to evaluate patterns of missingness. The core(s) requesting access to data from another core will be asked to provide a description of how data will be used for QC purposes. The **QA/QC Committee** will be the governing body for these requests. Upon approval by the RURAL QA/QC Committee, the description will be forwarded to the relevant cores for approval and participation to ensure the data are used appropriately.

3.4. *Dissemination of results to CABs/Community Stakeholders*

Dissemination of research findings to the community-at-large is important for the RURAL Study. The **Executive Committee?** will be the governing body for these requests. For these requests, the SCC will work with the SDCC to provide the necessary results to develop a message that is clear, simple, and easy to understand and appropriate for the target audience.

3.5. *Other requests*

We anticipate data requests that may not fall within one of the pre-specified categories. For these requests, the SCC and the SDCC will determine whether is necessary to obtain additional review and or approval by a RURAL sub-committee.

4. Request Timelines

Please allow at least 6-8 weeks from the date of approval of the request to receive the data output from the SDCC. Contributing factors to the amount of time it may take to receive the data include the order in which requests are received, the number of requests in the queue, the number of faculty and analysts available, and the extent to which the SDCC is conducting core activities. Please refer to the appropriate committee's policy for information regarding the approval process.

5. Data Access for RURAL Cores

The RURAL Cores will be given access to an analysis-ready limited datasets at the end of enrollment in each county or as needed. These local datasets will include an array of derived study variables and individual data elements to facilitate exploratory as well as full data analyses. Prior to receiving access to RURAL data, there must be a fully executed Data Use Agreement (DUA) in place between the University of Pennsylvania and the institution which houses the RURAL Core. Additionally, the Core PI will be asked to sign the RURAL Data Access Agreement. The RURAL Data Access Agreement will be renewed every two years.

5.1. Data Transfers

Data will be made available to the RURAL Cores via Box (or other secure data transfer platform). A schedule of uploads by the SDCC to Box will be established. A Box folder labeled "RURAL Data from SDCC" will contain the current version of the datasets; prior versions will be stored in a subfolder labeled Archive, and documentation will be stored in a subfolder labeled Documentation. All datasets will be labeled in a way that includes the date of dataset creation so that different versions of the same dataset can be distinguished. Documentation supplied by the SDCC will include a spreadsheet labeled M-Data Variable List.xlsx. Each row in this spreadsheet will contain one M-Data variable; the columns will provide information such as a short description of the variable and the timepoint at which it is collected.

Each Core PI will appoint a **Data Steward** who will have responsibility for downloading the data from Box and maintaining data security. The Data Steward will be the only individual at the Receiving Location with access to download the datasets from Box. It is the responsibility of the Core PI to notify the SDCC in a timely manner if there is a change in the Data Steward.

5.2. Data Storage and Security

All data obtained from the SDCC, and any datasets derived from those data, must be stored on an encrypted device and regularly backed up. Qualified devices must meet all HIPAA Privacy Rules. It is the responsibility of each Data Steward to ensure that the device is qualified. The Data Steward agrees to retain control over the data and shall not permit the transfer of data outside of the local institution and/or to any unauthorized individuals.

Because the datasets that will be distributed to the RURAL Cores will include some dates, it is possible that the IRBs at some or all of the RURAL Cores will want to have some oversight. Each RURAL Core must

confirm what their IRB requires and will be responsible for obtaining any regulatory approval that may be required by that IRB.

All data downloaded from Box should be kept in a single folder on the server. To meet the requirements of storing multiple copies of the data, the server must have a minimum of 35 gigabytes of storage available.

In addition to the above, all RURAL Cores must follow the data security terms and conditions outlined in the signed Data Transfer and Use Agreement.

5.3. Data Tracking

Each RURAL Core will be responsible for developing a data tracking system. The data tracking system at the RURAL Core will document: (a) the identity of every individual who has access to each folder containing RURAL data, and (b) which datasets are used in each approved analysis, including the original dataset and the final analysis dataset. The data tracking systems must have the capability of producing regular reports. The Data Stewards will have ultimate responsibility for maintaining these systems.

5.4. Intellectual Property

The SDCC shall provide the data to the requestor for the research purposes. The RURAL Study shall retain ownership of any rights it may have in the data, and recipient does not obtain any rights in the data other than as set forth in the signed Data Transfer and Use Agreement and the Data Request Policy.