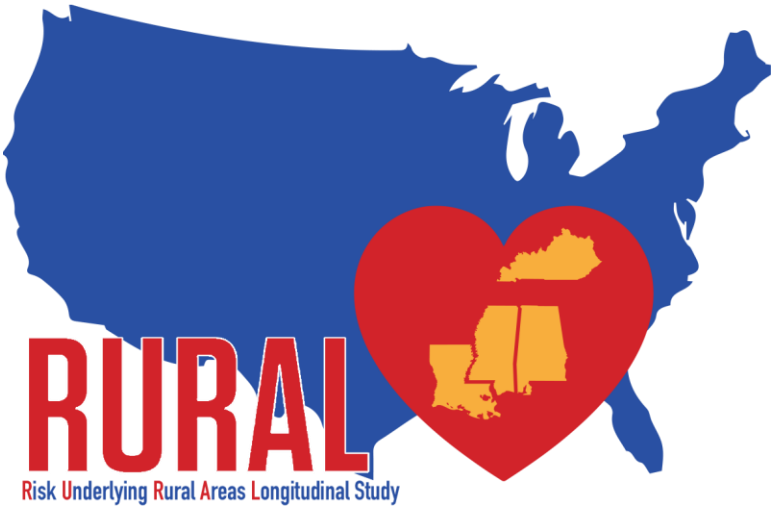

External Data Sharing Policy

December 2020

Contents

1.0 General Principles 3
2.0 What data will be shared 3
3.0 Who will have access to the data 4
4.0 Where data will be available 4
5.0 When data will be shared 5
5.1 Phenotypic Data 5
5.2 Genetic Data 5
6.0 How will researchers locate and access the data? 5
7.0 Sharing RURAL Cohort Study DNA and Biosamples 6
7.0 Dissemination of Results for Clinical and Public Health Practice 6
8.0 NIH Public Access Policy Compliance 7
9.0 Protection of Participant Privacy 7
10.0 Policy Changes 7
11.0 RURAL Cohort Study Contact Representative 7

External Data Sharing Policy

1.0 General Principles

The RURAL Cohort Study is committed to sharing data from the RURAL Cohort Study participants according to the FAIR principles, i.e., Fair, Accessible, Interoperable, and Reusable. As such, the RURAL Cohort Study will share all research data as broadly and as soon as possible and for as long as possible, consistent with National Institutes of Health (NIH) policies to maximize the benefits and scientific yield to the community.

The goals of the data sharing policy is to:

- Clarify details of data availability, accessing, and sharing
- Promote data discovery and integration through the development and the utilization of standard, unambiguous terminology
- Highlight creation and the enhancement of data dictionaries and ontologies
- Actively support the development and the application of methodologies to support interoperability of the RURAL Cohort Study data with other relevant data sets
- Allow formatting and depositing of individual-level de-identified data in existing NIH repositories
- Promote the development and implementation of a searchable web-based platform for the RURAL Cohort Study data sharing and integration
- Ensure the linkage and integration of multiple RURAL Cohort Study data sets from different sources such as the baseline examination in the mobile examination unit, mHealth data, data from geospatially-referenced and health records
- Enable the development and implementation of techniques for data visualization and to support big data analytics

2.0 What data will be shared

The RURAL Cohort Study is committed to sharing all finalized, clean, and de-identified data from the RURAL Cohort Study participants consistent with their informed consent provisions. Our phenotyping spans several different domains and includes imaging, questionnaire-based responses and physiological measurements at rest and upon dynamic testing. The RURAL Cohort Study participants will also undergo genotyping using a novel customized array.

The RURAL Cohort Study will share individual-level data and aggregate/summary data from both the core RURAL Cohort Study project and funded ancillary studies with appropriate and detailed documentation consisting of data dictionaries, meta-data, descriptors, and schema. When summary data are shared, appropriate details of analytical methods used to generate these summaries will be provided in the analysis manual (to be developed). (The RURAL Cohort Study will maintain data versioning and provenance.)

The RURAL Cohort Study will use desirable elements including data content, format, and organization), common data elements, and a standardized vocabulary/terminology to assure interoperability of our data to facilitate cross-cohort comparisons and pooling where feasible. This will be built into the RURAL Cohort Study Statistical Data Coordinating Center's (The University of Pennsylvania) protocols.

All the RURAL Cohort Study data is subject to rigorous Quality Assurance and control measures led by our Statistical Data Coordinating Center. Some data that is shared may require additional approvals and appropriate agreements, either because of risk of participant identification, any associated sensitivity as defined by the community and participant advisory boards (CABs and PABs) and the RURAL Cohort Study Ethics Advisory Board. Geospatial data will not be shared

through repositories due to the re-identification risks to participants

The details of types of data that will be shared and steps for accessing the data will be posted on the RURAL Cohort Study website in the future (https://www.theruralstudy.org)

3.0 Who will have access to the data

The RURAL Cohort Study is committed to wide sharing of de-identified data as broadly as possible to the extent consistent with applicable laws, regulations, NIH rules, and policies, including in accordance with the Federal Policy for Protection of Human Subjects (also known as the Common Rule).

The RURAL Cohort Study Statistical Data Coordinating Center and the Principal Investigators of the RURAL Cohort Study will be responsible for ensuring that no personally identifiable information on study participants is made available and that The RURAL Cohort Study always complies with the individual informed consent of our participants. Appropriate (as defined by the NHLBI, IRBs, Federal guidelines, and the Study Ethics Advisory Board) data-sharing agreements may be necessary to restrict the transfer of protected, sensitive, or confidential data and to require that data be used only for research purposes.

The RURAL Cohort Study will make aggregate de-identified data available to the local RURAL Cohort Study communities, community and participant advisory boards (CABs and PABs), local learning collaboratives (LLC), and the local health authorities to facilitate local health planning at the county level. Furthermore, the names and Institutions of persons either given or denied access to the RURAL Cohort Study data, and the basis for such decisions will be summarized in the annual progress report to the National Heart, Lung, and Blood Institute (NHLBI) and the RURAL Cohort Study Observational Study Monitoring Board (OSMB). The exact process for accessing study data will be described on the RURAL Cohort Study website in the future.

4.0 Where data will be available

The RURAL Cohort Study will deposit all appropriate phenotypic data in the NHLBI BioLINCC data repository and all genomic data in dbGaP. The RURAL Cohort Study data will be deposited in these repositories by the Statistical Data Coordinating Center. The RURAL Cohort Study's datasets and documentation will be formatted according to BioLINCC and dbGaP standards prior to submission. Each dataset submission includes a BioLINCC-/dbGaP-formatted dataset, a data dictionary, i.e. the metadata, a coding manual, a data collection protocol, an annotated data collection form, and a funding source reference.

Both BioLINCC and dbGaP have data access policies and procedures consistent with NIH data sharing policies. In general, outside investigators with approved research proposals and Institutional Review Board protocols will be able to access RURAL Cohort Study genomic and phenomic data if they submit: a) a research proposal; b) evidence of a local IRB approved project (including certifications of exemption from IRB/ethics); c) an executed DMDA – data materials and distribution agreement; d) a computer security plan. All research protocols will be posted on the RURAL Cohort Study website. All RURAL Cohort Study data will be made accessible immediately upon publication or one year from the date that the database is locked for analysis discovery, whichever comes first. There will be the period as defined by the Statistical Data Coordinating Center between when the database is locked for analysis and when the data is accessible in the wider RURAL Cohort Study database to allow the Statistical Data Coordinating Center to review the data for potential issues before it is widely distributed. The exact location and processes for accessing shareable study data will be posted on the RURAL Cohort Study website in the future.

In addition to dbGaP and BioLINCC, another venue for data sharing is through the RURAL Cohort Study (rural@bu.edu).The RURAL Cohort Study Ancillary Studies Subcommittee, and Steering Committee will evaluate and approve each submitted research proposal. Ancillary study proposals submitted by external and internal RURAL Cohort Study investigators will be processed by the RURAL Cohort Study Statistical Data Coordinating Center, which will handle data distribution for approved ancillary study proposals.

In general, the RURAL Cohort Study will recommend usage of public repositories (BioLINCC or dbGaP for non-genetic and genetic data, respectively) when proposals originate from investigators at other institutions (beyond the Core RURAL Cohort Study institutions).

5.0 When data will be shared

5.1 Phenotypic Data

Our phenotyping spans several different domains and includes both questionnaire-based responses, imaging, and physiological measurements at rest and upon dynamic testing. For phenotypic data, the RURAL Cohort Study will adhere strictly to NIH policy and share data as soon as possible, and no later than within one year of completion of the initial phase of the RURAL Cohort Study or any funded ancillary study. As longitudinal surveillance of the RURAL Cohort Study accrues, increments in data will be released consistent with NIH/NHLBI policy in this regard for cohort studies.

Images will be shared according to NIH/NHLBI policies. In addition, the RURAL Cohort Study's images will only be shared through the Ancillary Study process.

Accordingly, some existing datasets will be routinely updated. When an updated dataset is submitted and posted to BioLINCC or dbGaP, the previous older dataset will be retired and no longer displayed. Each database version will be appropriately cataloged and archived. The exact schedule for sharing and release of study data will be posted on the RURAL Cohort Study website in the future.

5.2 Genetic Data

Consistent with NIH Genomic Data Sharing policy, data submission into dbGaP is expected to begin after data cleaning and quality control. This will generally occur within 3 months after the data has been generated but will vary by project. Following the initiation of data submission, there is then a 6-month period of exclusivity before the data would be released by dbGaP to external users.

6.0 How will researchers locate and access the data?

The study website will clearly describe how researchers can learn about, locate, and access the study data.

The RURAL Cohort Study Statistical Data Coordinating Center will create an online training module and lists, tables, charts, and figures of study exams and data on the structure and storage of the RURAL Cohort Study datasets. Additionally, the RURAL Cohort Study will apply and seek additional funding to facilitate face-to-face sessions with groups of investigators interested in actively working with the RURAL Cohort Study datasets.

The online training modules will be covering the following topics:

- Structure of datasets
  - The RURAL Cohort Study exam datasets, including medical history, questionnaire data, physical exam and laboratory measures such as lipids, glucose, and novel

biomarkers, genetic tests, and other tests such as pulmonary function tests, vascular test, CT, dynamic tests, etc.;
    - Sequence of events datasets, containing dates of mortality and HLBS (Heart, Lung, Blood and Sleep) disease events;
    - Death and HLBS events survival datasets, containing last known contact dates and follow up time by event;
    - Summary datasets, containing commonly used risk factors collected over multiple exams over the course of the RURAL Cohort Study.
- Analysis tips to appropriately account for the probabilistic sampling design in commonly used statistical modeling such as Poisson, logistic, GEE, and survival regressions in common statistical packages such as SAS, SUDAAN, and R.
- An explanation of the dataset archive layout to aid users in locating the file with their variables of interest.
- Instructions regarding how to request and access the RURAL Cohort Study data (directly from the Study and/or from BioLINCC or dbGaP).

7.0 External Data and Materials Sharing Policy

Requests for DNA or biosamples of the RURAL Cohort Study participants will require a more extensive research proposal than requests for existing phenotypic and genotypic data. The RURAL Cohort Study DNA and Laboratory subcommittees will review proposals for accessing DNA and non-DNA biosamples, respectively. The NHLBI Project Officials and the RURAL Cohort Study Coordinating Center at Boston University will constitute these two committees. The RURAL Cohort Study will outline the criteria for DNA/Laboratory committee proposals at the RURAL Cohort Study website. Generally, these criteria will include scientific merit, genotyping/assay performance quality of the requesting laboratory, available stockpile of DNA/biosamples on participants, and statistical power for the proposed research as detailed in the submitted research proposal. Additional criteria may include the potential value to the rural participant and/or the community. The DNA and biosamples of study participants will be stored at our Central BioRepository at the University of Vermont. Aliquots of DNA/biosamples will be shipped from the Central Biorepository following requisite steps (typically, proposal approval by the DNA/Laboratory committee, institutional IRB approval, and a signed data and material transfer agreement between the requestor's institution, and Boston University)

7.0 Dissemination of Results for Clinical and Public Health Practice

The RURAL Cohort Study investigators are confident that the RURAL Cohort Study will disseminate our findings widely through various means:

- Results will be presented at scientific meetings in the form of posters and oral presentations.
- Results will be published promptly in high impact, peer-reviewed journals.
- As noted above, any non-genetic data will be made available within one year of adequate data cleaning and QC before the end of the award. Genetic data will be released as per the current NIH GWAS policy.
- The RURAL Cohort Study can also provide web-post key programming codes and macros to facilitate replication and extension of our results by other scientists.

De-identified aggregated data will be shared with local RURAL Cohort Study communities, CABs and PABs, LLCs, and with local departments of health so that these data can be used for planning health services in the respective counties. The Recruitment and Retention Core will return the examination actionable results to the participant.

8.0 NIH Public Access Policy Compliance

Manuscripts based on the RURAL Cohort Study data that are accepted for publication by peer-reviewed journals will be submitted to PubMed Central via the NIH Manuscript Submission System, according to the NIH Public Access Policy.

9.0 Protection of Participant Privacy

The protection of the rights and privacy of the RURAL Cohort Study participants is a priority at all times. Sensitive data will be de-identified and anonymized. This includes restricting access to information regarding sensitive data and restricting sensitive data in the data set. Sensitive data would include various domains dependent on context, sub-population, and time.

Measures that may be used to prevent the breach of data include:

- Anonymity contracts
- Data encryption
- Password-protected portals
- Two-factor authentication
- Audit trails
- Security plans for data breach
- Locked and limited access storage facilities
- Firewall programs
- Cloud-based computing platforms

10.0 Policy Changes

The policies and processes of the data sharing policy is a foundational layout of the RURAL Cohort Study and is subject to change based on the growth and needs of the RURAL Cohort Study, NHLBI requests, and internal input.

11.0 RURAL Cohort Study Contact Representative

RURAL Cohort Study Coordinating Center
Boston University
Email: rural@bu.edu