



RESPIRE Data Management Plan (DMP): Template (adapted from the University of Edinburgh)

Name:	Professor Samir K Saha
Project Title:	The Consequence of Respiratory Syncytial Virus (RSV) Infection in Young Infants
Institute:	Child Health Research Foundation
Start Date:	June 2019
End Date:	June 2021
DMP version number and date:	DMP_RSV_FU_1_23/02/20

Responsibilities & Resources

Who will be involved in the data management of this research?

The RSV follow-up study is a multi-centre study where four centres in three South Asian countries are recruiting study participants. Two centres are located in Pakistan, among them, one is located in the urban area of Karachi city, and the other centre is located in Matiari district of Pakistan. The Bangladesh centre is located in Sylhet district, and the India centre is located in Odisha province. At each centre, data will be collected from children and their caregivers through interview and physical assessment. An embedded PhD is included with this overarching research involving the data collected from the Sylhet site. Data will be collected in a different format at different centres. Therefore, different personnel will be involved in the data management process in different centres. In general, the following personnel are involved with data management:

Projahnmo Research Foundation (Sylhet, Bangladesh):

1. Research Assistants: They collect household information and children medical history from the caregivers by using structured questionnaires.
2. Physician: They perform the clinical assessment of the children.
3. Data entry operators: They enter the collected data into the database.
4. Data supervisors: They audit the collected data to ensure that all data collected properly and check the data consistence.



5. Data managers: He maintain the database and provide guidance to the data entry operators whenever needed, also perform the trouble shooting in the data entry software.
6. Statisticians: They audit the data and check the data quality

Aga Khan University (Matiari, Pakistan):

1. Research Assistants: They collect household information and children medical history from the caregivers by using structured questionnaires.
2. Physician: She performs the clinical assessment of the children.
3. Data entry operators: They enter the collected data into the database.
4. Data supervisors: They audit the data to ensure that all data are collected properly. They also check the data consistency.
5. Data manager: He maintains the database and guides the data entry operators whenever needed, also perform the troubleshooting in the data entry software.
6. Statistician: She audits the data and checks the data quality

Aga Khan University (Karachi, Pakistan):

1. Research Assistants: They collect household information and children medical history from the caregivers by using structured questionnaires.
2. Physician: She performs the clinical assessment of the children.
3. Data supervisors: They audit the data to ensure that all data are collected properly. They also check the data consistency.
4. Data manager: He maintains the database and guides the data entry operators whenever needed, also perform the troubleshooting in the data entry software.
5. Statistician: She audits the data and checks the data quality

Asian Institute of Public Health (Odisha, India)

1. Research Assistants: They collect household information and children medical history from the caregivers by using structured questionnaires.
2. Physician: She performs the clinical assessment of the children.
3. Data supervisors: They audit the data to ensure that all data are collected properly. They also check the data consistency.
4. Data manager: He maintains the database and guides the data entry operators whenever needed, also perform the troubleshooting in the data entry software.
5. Statistician: She audits the data and checks the data quality

Child Health Research Foundation (CHRF, Bangladesh)

1. Data managers: He maintains the database and provide guidance to the centre-specific data entry operators whenever needed, also perform the troubleshooting in the data entry software.
2. Statisticians: They audit the data and check the data quality.

1. Data Capture

What data will be generated or reused in this research?

Data will be generated from two sources.

1. Questionnaire based data: Research assistants collect various data using structured questionnaires during the household visits. Data collected by research assistants include socio-economic status, demographic and physical health. Data collected by physicians include physical health and medical care.
2. Data collected from electronic devices: All children participating in the study are attending spirometry test. Data generated through this test procedure are directly transferred to the data capture system.

How much data will be generated?

The data of this multi-centre study is expected to be in the range 1-10 GB.

2. Data Management

How will the data be documented to ensure it can be understood?

A codebook for each data table is being prepared to describe the variables and the code. The codebook entails the study design, data collection methodology, details of the variable used to record and collect information intended for the study.

Where will the data be stored and backed-up?

This is a multi-centre study where data is collected at four different sites of three south Asian countries-Bangladesh, India and Pakistan. Projahnma Research Foundation (PRF) has developed a SQL based Central Data Management platform to enter the data at a different site. This platform is currently being used at Bangladesh and one Pakistan site (Matiari). The other Pakistan site (Karachi) is using tablet-based data entry system. The Indian site is collecting data in the RedCap system.

The collected data is initially stored at a local server of the respective sites. Sites periodically (once a month) sent the data to the data centre located at the Child Health Research Foundation (CHRF) through one drive. A backup copy is stored at the local level of each centre. At CHRF, the data is stored in a data server, a hard disk and in the cloud (OneDrive). A consolidated database is being created through merging the site-specific dataset. The sharing of the local dataset has been conducted under the auspices of the data-sharing agreement between the local sites and CHRF. The database will be further stored in the Datastore storage facility of the University of Edinburgh for secure storage. Please find the site-specific details for data entry and storage

Projahmno Research Foundation:

Research assistants and clinician collect data using paper-based questionnaires. A supervisor checks the collected data for missing values and data consistency. After supervisory check, the data entry operator enter the collected data into the system. Entered data is stored in the local server. The data manager supervised the overall process and assist in the data entry process for the smooth functioning of the system and ensure that entered data are stored properly. A copy of the database shall be kept with Shahidul's laptop as the same data will be utilized for the embedded PhD within this research

Aga Khan University (Matiari, Pakistan): A mobile team consist of research assistant and clinicians collect data using paper-based questionnaires. A supervisor checks the collected data for missing values and data consistency. After the supervisor check, a data entry operator enters the data in a SQL based platform that contain logical and range check. Entered data is stored in the site-local server and periodically shared with data centre

located at CHRF under the agreed data-sharing agreement. A data manager supervises the overall process and assist the data entry operator for a smooth function of the system and ensure that entered data are stored properly.

Aga Khan University (Karachi, Pakistan): Research assistant and clinicians collect data using tablet-based questionnaires. A supervisor checks the collected data for missing values and data consistency. Entered data is stored in the site-local server and periodically shared with data centre located at CHRF. A data manager supervises the overall process and assist the field team for the smooth functioning of the system and ensure that entered data are stored properly.

Asian Institute of Public Health (Odisha, India): Research assistant collect data using REDCap system. A supervisor checks the collected data for missing values and data consistency. Entered data is stored in REDCap and will shared periodically with the data centre located at CHRF. A data manager supervises the overall process and assist the field team for smooth functioning of the system and ensure that entered data are stored properly.

3. Integrity

How will you quality assure your data?

The data quality is assured through routine review of the collected data for missing values, incompleteness, and data inconsistency. Sites that are collecting data in paper form are entering twice (double data entry) for avoiding punching error into an electronic database. The self-reported data collected is being validated through logical checks on the range, consistency, uniqueness and skipping rules for avoiding the collection of undesired information. Furthermore, 10% of the spirometry test data will be sent for quality audit to external evaluators within the RESPIRE network through restricted access of a secured platform such as One drive in encrypted pdf format.

The spirometer has the auto-calibration function and does not need manual calibration. Still, all sites have planned to perform biological control to ensure the quality of the test. The performance of the pulse oximeters is also planned to check using a biological control. All data are linked through the participant identification number of the parent study (ANISA study) to ascertain the association of RSV status with the expected outcome. Metadata will contain information about the variables collected for evaluation and consistency. Log-

recording of the individual access of the data shall be kept at the individual sites for audit trails from the respective servers.

4. Confidentiality

How will you manage any ethical and Intellectual Property Rights issues?

This study is committed to maintaining GDPR's (General Data Protection Regulation and Data Protection Act 2018) guideline during data collection, storage and transmission. All collected data forms are kept under lock and keys, and only authorised persons have access to the review the collected data. The data collection application used in Karachi and Odisha sites are password-protected and only selected personnel to have access to the system who are prior authorised. The study have taken the following data safety measures at different sites:

Projahmno Research Foundation (Sylhet): Each individual participant has received a child ID and are tracked using that ID. The mothers of the child has also received a current and a permanent ID for tracking purpose. The collected data are kept under lock and key, and only authorized personnel can review the data. All the collected data containing personal information (e.g. household address) are codified, and the codes are preserved separately.

Aga Khan University (Matiari): Each individual received a child ID and tracked using that ID. The mother of the child also received a current and a Permanent ID for tracking purpose. The collected data are kept under lock and key, and only authorised personnel can review the data. All collected data are containing personal information (e.g. household address) are codified, and codes are preserved separately.

Aga Khan University (Karachi): Each individual received a child ID and tracked using that ID. The mother of the child also received a current and a Permanent ID for tracking purpose. Data are collected using a mobile application which is password protected. All personal information is codified in the database, and codebook are kept separate from the data system. All identifiers are removed before sharing the data with the data centre.

Asian Institute of Public Health (Odisha): Each individual received a child ID and tracked using that ID. The mother of the child also received a current and a Permanent ID for tracking purpose. Data are collected using a mobile application which is password

protected. All personal information is codified in the database, and codebook is kept separate from the data system. All identifiers are removed before sharing the data with the data centre.

Child Health Research Foundation: All study sites share data removing the personal identifiers with CHRf using OneDrive which are encrypted and password protected and accessed to the relevant study personnel. Site-specific data are merged for developing the consolidated database. The consolidated database is stored at the central server facility at CHRf. The server is password protected with only authorised access. The data analyses shall be done with the removal of any identifiable information collected for the identification of the eligible study children.

For data sharing with the sponsor, all data shall be anonymous, and any identifiable information shall be removed or scrambled to prevent any ethical breach of the data of the collected household.

The study staff and investigators involved in the study are committed not to disclose the data collected in this study or use for any purpose other than reporting performance and findings of the study. All analysis shall be carried out in line to protect the confidential information of the study participants. Compliance with the appropriate data protection legislation (including General Data Protection Regulation and Data Protection Act 2018). Prior written agreement from the sponsor or its designee must be obtained for disclosure of any confidential information to other parties.

From this study, we do not aim to produce any IP product.

5. Retention and Preservation

Which data do you plan to keep and for how long?

All data collection forms will be preserved under lock and key until the main result is published or up to a maximum period of 7 years. The electronic database will preserve unlimited time for future use, especially to monitor the long term outcome of the study children as mentioned in the consent form.

How will the data be preserved?

The paper based collected data are kept under lock and key. Data will be stored at the site-specific server locally, which will be shared with the study coordination team at Child Health Research Foundation (CHRF) through OneDrive. The data from the study sites are collated and encrypted. The encrypted data will be preserved in CHRF data archive, which is password-protected, and only the data manager has access to the server. The final database will be stored at the secure platform at CHRF. A copy of the database will be submitted to the University of Edinburgh to host for public access after the main results are published.

6. Sharing and Publication

Which data will be shared and how?

In this study, we are collecting the disease history of the participants, as well as assessing the lung function and blood eosinophil count. Additionally, specific asthma risk factor data also being collected. All this data will be analysed to measure the impact of RSV infection on child health. All the data generated in the study must be kept in a restricted manner with secured access only to the relevant study staff. While sharing the data with any institute (mainly sponsor) or individual (investigators and data analyst), all personal identifiers and personal information will be removed. These includes:

- (a) Names;
- (b) Postal address information (other than town or city)
- (c) Telephone numbers;
- (d) Medical identification number

Are any restrictions on data sharing required?

The study data collected at Bangladesh (Sylhet site) Centre will be share with Mohammad Shahidul Islam, a RESPIRE PhD student at the University of Edinburgh to be used for his PhD thesis. Shahidul was actively involved with the study design and implementation of the study protocol at different Centres and coordinated the study activities. Additionally, a copy of the anonymised data will also be placed in the DataShare for public use once the main results are published. The collated data will be shared with the sponsor and will be preserved in DataVault. Any sharing of the preserved data other than sponsor will be documented through a written data-sharing agreement entailing a commitment for the usage of data for research purpose only in a secure manner which will be destroyed upon completion of the analysis.