



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

Name:	Dr. Salahuddin Ahmed
Project Title:	Community use of digital auscultation to improve diagnosis of paediatric pneumonia in Sylhet, Bangladesh
Institute:	Projahnmo Research Foundation (PRF)
Start Date:	1 January 2018 (field activities started from May 2019)
End Date:	30 September 2021
DMP version number and date:	Version-1.0; 11 December 2020

Responsibilities & Resources (applicable across the sections below)

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

Data collection and uploading:

Community Health Workers (CHWs) and Community Health Care Providers (CHCPs) are responsible for collection of data. Trained and standardised Paediatricians will interpret recorded lung sounds and create a data file as well as computer-generated lung sound interpretation will be created using artificial intelligence technique.

Programming and regular troubleshooting and data feedback management:

One Programmer designed the data entry system, and he is responsible for updating the entry system and regular troubleshooting whenever required.

Overall data management:

Data Manager is responsible for the overall data management for the study. He is responsible for preparing analytical data files and archiving data files.

1. Data Capture

What data will be generated or reused in this research?

We are collecting data on general physical health, clinical signs, oxygen saturation and lung sound record from children. We are also collecting household socio-economic data



of enrolled children. We are collecting those data using tablets. Lung sound is being recorded in Waveform Audio Format (.wav). A trained and standardised paediatric listening panel are interpreting lung sounds in five categories – Only wheeze, only crackles, both wheeze and crackles, no wheeze and no crackles, and uninterpretable. A machine learning artificial intelligence technique will classify the lung sound in the same categories. The study will also generate audio recording of four focus group discussions (FGDs) and their transcripts in the native language (Bengali), later transcripts will be translated in English and archive as word document.

Below datafiles will be created from this study:

- a) Two monthly children surveillance data file: CHWs are visiting each child in the study area in every two months and collecting child history of respiratory illness symptoms as well as examine the child's respiratory system and recording signs including respiratory rate
- b) Screening data file: Community Health Care Providers (CHCP) are screening all under-5 children who visit community clinic and record the finding which include history of cough or difficult breathing
- c) Enrolment data file: CHCP examine all enrolled children and record findings in this data file which include respiratory symptoms and signs, temperature, oxygen saturation, weight, height, mid arm circumference, record lung sounds using the Smartscope
- d) Confounding factors data file: CHWs are collecting the socioeconomic status and other confounding data from each enrolled child's parents or carer
- e) Lung sound interpretation by human data file: A trained paediatric listening panel is form. All recorded lung sound files are interpreting by two members and if discordant, then, Dr Eric McCollum interpret the sound file and which is the final interpretation. Those interpretation will create another data file
- f) Lung sound interpretation by machine data file: All sound files will be interpreted by a machine learning algorithm and create another data file
- g) FGD data files: Bangla and English transcript in word/pdf files

How much data will be generated?

We anticipate to generate about 30-50 GB of data.

2. Data Management

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

A codebook for each data files will be prepared to describe the variables and the code of the database. The codebook entails the study design, data collection methodology, details of the variable used to record and collect information intended for the study. For FGD, Bangla transcripts and translated English transcripts will be archived in word/PDF document.

Where will the data be stored and backed-up?

We store data on the Projahnmo Research Foundation (PRF) secured server at Sylhet, Bangladesh with a three-hourly copy made to the password- and user-access protected another server at PRF's Dhaka office in Bangladesh. Anonymised data will also be stored on DataVault of the University of Edinburgh.

The data manager of PRF is responsible for ensuring that the data is regularly backed-up.

3. Integrity

How will you quality assure your data?

Data collection tools were pilot tested before use in the main study. All instruments are calibrated at least once in a month. All staff received training and standardised; they also receive quarterly refresher training. Data entry is conducted at the time of the interview and directly into an electronic format to reduce transcribing errors and loss of information. Key variables are double entered in electronic capture to reduce entry error. Validation rules, including logical checks on the range, consistency, uniqueness, and skipping rules for avoiding the collection of undesired information were built in the data system to prevent errors during data collection or data entry. Collected data is routinely reviewed for missing values, incompleteness, and data inconsistency.

4. Confidentiality

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

In this study, we are collecting data using tablets and after collection of data, the data is transferred to a database on our server at Sylhet using internet connection of each tablet. Both the tablet and the server are password protected. Paediatric listening panel is filling up the lung sound interpretation on an online database which is also transferred to our server real time. Access to collated participant data is restricted to individuals from the research team treating the participants, representatives of the sponsor(s) and representatives of regulatory authorities. Different members of the data management team have different levels of permissions for accessing the database based on their responsibilities. Only the Programmer and Data Manager have access to the central database.

Ownership of the data arising from this study resides with the study team. Salahuddin Ahmed, a RESPIRE PhD student at the University of Edinburgh will use the study data for his PhD thesis. Total six datasets will be created using all quantitative data is being collected in the study. Those datasets will be used by Salahuddin Ahmed for his PhD thesis. Issues of copyright will be dealt with via standard University of Edinburgh procedures for publishing

research papers and PhD thesis. The intention is to publish in high-impact open access journals as and when appropriate, depending on availability of funding.

5. Retention and Preservation

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

De-identified data will be preserved without limit of time in electronic format in PRF server in Bangladesh. All study documentation will be kept for a minimum of 3 years and maximum of 5 years from the protocol defined end of study point. When the minimum retention period has elapsed, study documentation will not be destroyed without permission from the sponsor.

How will the data be preserved?

Data will be stored in password- and user-access protected servers located in two different locations (Sylhet and Dhaka) in Bangladesh. A copy of the database will be submitted to the University of Edinburgh to host for public access after the main results are published. The physical copies of filled up data forms will be retained in a locked cabinet, in a locked room at PRF office in Sylhet, Bangladesh.

6. Sharing and Publication

Which data will be shared and how?

In this study, we are collecting data on general physical health, clinical signs, oxygen saturation from children younger than five years. We are also recording lung sound from the children and a paediatric listening panel is interpreting the sound files as only wheeze, or only crackles or both wheeze and crackles or no wheeze and no crackles or uninterpretable. All those data will be analysed to understand the risk factors of childhood pneumonia, performance of community health care providers to assess children younger than five years and their performance of recording lung sound using the Smartscope. 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 will be shared with Salahuddin Ahmed, a RESPIRE PhD student at the University of Edinburgh to be used for his PhD thesis. Salahuddin was actively involved with the study design and implementation of the study protocol and coordinated the study activities. Additionally, a copy of the anonymised data will also be placed in the DataVault 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.

