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

Name:	Professor Samir K Saha
Project Title:	Estimating the Burden of Chronic Respiratory Diseases (CRD) in Bangladesh Using the 4 Country Chronic Respiratory Disease [4CCORD] Study Tools
Institute:	Child Health Research Foundation
Start Date:	January 2021
End Date:	April 2021
DMP version number and date:	DMP_CRD_1.1_19/12/20

Responsibilities & Resources (applicable across the sections below)

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

The following personnel will be involved with data management

1. Data entry operator: They will enter the collected data into EpiInfo platform
2. Data supervisors: They will check the collected data to ensure that all data collected properly.
3. Data managers: They will develop and maintain the data entry software, provide guidance to the data entry operators whenever needed and perform the trouble shooting in the data entry software
4. Statisticians: They will 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 will collect various data using structured questionnaires during the clinic visit. Data collected by research assistants include demographics, risk factor, Chronic respiratory symptoms, work productivity assessment, healthcare associated costs due to Chronic Respiratory Disease, screening questionnaire relevant to asthma and/or COPD, Healthcare associated costs. Data collected by physicians include notes on clinical assessment.
2. Data collected from electronic devices: All participant will attend a spirometry test. Data generated through this test procedures will be directly transfer to the data capture system.

How much data will be generated?

The data of this 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 will be prepared to describe the variables and the code. The codebook will 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?

The data form will be kept under lock and key in the data vault at CHRF. No data forms will be transferred outside the CHRF Office. Data collected during the interview through questionnaires will be entered in a computer using Epi Info. The data will be codified. Data will be transfer directly to the computer. The data will be encrypted and will be preserved in CHRF data archive which is password protected and only data manager have the access the server. A back up copy of the encrypted data will be kept at the secure location in the cloud (One Drive) and local hard drive. A copy of the database will be submitted to the University of Edinburgh to host for public access after the main results are published. A copy of the database will also be stored in the secured cloud system (in the data vault of University of Edinburgh). Any data sharing with external partners such as KEM Hospital within the RESPIRE Network for the auditing of the spirometry data will be shared after the removal of identifier with an unique ID under the auspices of the data sharing agreements

3. Integrity

How will you quality assure your data?

The data quality shall be ensured through routine review of the collected data for missing values, incompleteness, and data inconsistency. Data entry shall be carried out by two persons for avoiding errors and comparison.

These rules will be built in the data system for prevention of errors. Data quality checking will be done through double blinded entry and comparative analysis. Furthermore, the 10% of the spirometry data shall be sent for quality audit to external evaluators (KEM Hospital) through secured platform such as One Drive.

Spirometers has auto-calibration function in places. Metadata will contain information about the information of the variables collected for evaluation and consistency. Log-recording of the individual access of the data shall be kept for audit trails from the central server.

4. Confidentiality

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

The study is committed to follow the GDPR's guideline during data collection, storage and transmission. All data collected shall be protected with password and kept in a secure storage area with limited access. This will entail only the relevant study personnel with specific access to the database as per requirement.

For external sharing, 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 shall not disclose 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 collected data form will be preserved under lock and key until the main result is published or a maximum of 7 years. The electronic database will preserve unlimited time for future use especially to monitor the long term outcome of the study participants.

How will the data be preserved?

The collected data will be kept under lock and key at the CHRF Data Vault. Data will be encrypted and preserved in CHRF data archive which is password protected and only data manager have the access the server. The final database will be stored in the encrypted format in the secure platform at CHRF. A backup copy will be submitted at the University of Edinburgh named Edinburgh DataVault in the anonymized form for preservation.
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?

The database will contain various information related to respiratory health, asthma and COPD risk factors, respiratory healthcare associated costs etc. The clinical information from the lung function assessment will also be linked. The data will be kept under lock and key with access to the relevant study personnel. The electronic copies shall be stored with encryption. Any sharing of the data, the personal identifiers will be removed and personal information will be removed as well. These include

a. Names
b. Postal address information

c. Telephone number
d. Medical identification numbers

Are any restrictions on data sharing required?

A copy of the anonymised data will be placed in the Edinburgh DataShare for public use once the main results are published. The collated data will be shared with 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.