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

Name:	Rita Isaac (bigger feasibility Project) Biswajit Paul (embedded PhD)
Project Title:	Prevention, detection and treatment of adult lung disease including lung cancer in a poor, rural population in Tamil Nadu: feasibility (*bigger feasibility project with embedded PhD*) Development and pilot testing of Theory of Planned Behaviour based educational intervention to improve knowledge, attitude and health behaviour in people with Chronic Respiratory Disease: A study in Southern Indian Rural Community (*embedded PhD of Biswajit Paul)*
Institute:	RUHSA Department, Christian Medical College Vellore
Start Date:	Feasibility: 1st October, 2018 PhD: 1st April 2018
End Date:	Feasibility: 31st March, 2020 PhD; 31st March 2021
DMP version number and date:	V1 dated 05/10/2020

Responsibilities & Resources (applicable across the sections below)

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

The Principal investigator (PI) from RUHSA Department, CMC Vellore, Co-investigators, data managers in RUHSA and Department of Biostatistics, Christian Medical College, Vellore are responsible for data management.

The Department of RUHSA and Biostatistics, Christian Medical College, Vellore, principally responsible for data storage, data cleaning, quality checks, data security and data sharing with required permissions during the period of the project.

The University of Edinburgh is advising on data repository for depositing data generated out of the project after removal of the identifiers and complete de-identification of the data at the end of the project period. They will advise on secure data storage; data preservation and data sharing to the scientific community.

1. Data Capture

What data will be generated or reused in this research?

The following types of data and datasets will be generated –

A) Quantitative data

i. Pre and post intervention data on awareness, attitudes and health behaviour - using baseline and end-line questionnaires by research team members at the beginning and end of the study period respectively; such data was collected in paper format, entered into an electronic database by data entry operator and held in RedCap software (database) of Christian Medical College (CMC), Vellore.
ii. Clinical data of patients collected during recruitment and subsequent clinical reviews done every three months are entered into electronic data entry form and held by the computer cell and the data management team of RUHSA department, CMC Vellore.
iii. Data on disease status and impact of the intervention is measured at baseline, at 6 months and at the end of the intervention period using St. George's Respiratory Questionnaire (SGRQ); this is an electronic questionnaire and the data captured by the e-forms are transferred to RedCap software for maintenance.
iv. Data on use of inhalers and adherence is collected through Test of Adherence to Inhalers (TAI) questionnaires by members of research team (health aides); these are in paper format and such data is entered by data entry operator and maintained in RedCap database.

B) Qualitative data

Qualitative data is collected at the baseline, in the middle of the intervention period and at the end of the intervention period using a qualitative topic guide by a qualitative researcher from the research team. Qualitative data include focus group discussions, in-depth interviews and key informant interviews. Eight focus group discussions including 51 participants, 5 in-depth interviews with care-givers and 4 key informant interviews were conducted at baseline whereas 16 in-depth interviews were conducted with patients at midline and 20 in-depth interviews were planned for endline. Interview guide was used by the researcher for conducting these interviews. All such conversations – that is focus group discussions and interviews are audio-recorded besides field notes of the researcher in paper format. The audio files are transferred from the recorder and are held by the computer cell of RUHSA, CMC Vellore with secondary backup updated on a daily basis; the audio files are in .mp3 format. The audio recordings of the recorder are deleted permanently.

C) Other data (Embedded PhD)

i. Handouts and pamphlets of educational materials related to health behaviour in chronic respiratory disease (CRD) patients are in paper format as well as in electronic format. The paper formats were for distribution to the participants and in the community. The electronic forms are also available in .pdf format and will shared with University of Edinburgh freely.
ii. Motivational videos based on theory of planned behaviour – the psychological theory used to develop the intervention are present in .mp4 format
iii. Short videos on talk to your doctor in .mp4 format enacting interviewer asking questions to the doctor related to CRD directly or taking phone-ins from patients
iv. Pictures and videos of puppet shows held in the villages for the patient and community awareness about CRDs and include frequently asked questions through a native way of storytelling which have been captured and in .jpeg and .mp4 format respectively. The puppet show is live while the audio for the puppet show is pre-recorded in .mp3 format.

These other data are held by the computer cell and the data management team of RUHSA department, CMC Vellore in their system with a secondary backup site which is backed up daily.

How much data will be generated?

Approximately 50-250 GB

2. Data Management

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

Both Qualitative and Quantitative data will be documented and maintained by the Computer Cell and Data Management team, RUHSA Department, CMC. Documentation of the data has been done through predesigned structured questionnaires, clinical review forms and project documents. Documentation was also done using field notes and audio recordings of the interviews and focus group discussions. The variable level details are available with data management team of RUHSA department which can be accessed with permission. The blank templates of questionnaires, clinical review form and qualitative topic guides can be deposited into Edinburgh DataShare. The research data after de-identification can be deposited with University of Edinburgh data repository, i.e. the Edinburgh DataVault. The metadata will be deposited in XML format.

Where will the data be stored and backed-up?

Live data during the project period is stored by the Biostatistics department of the Christian Medical College, Vellore with backup of the data at a secondary site. The data is backed up automatically every day. This is a secure platform for personal and sensitive data, it has limited access to the data management team and is password protected. The paper copies and documentation are kept in closed cabinets of the RUHSA department under lock and key and have limited access to the data management team. After the project period, all the identifying information from the data will be removed and the anonymised dataset will be created before analysis which will be transferred securely at University of Edinburgh (DataStore) for backup, recovery or access with permission. Research data will be deposited in the Edinburgh DataVault and can be obtained by permission from the PI or Co-I through the nominated data manager. The data obtained by statistical analysis will be kept in the Edinburgh DataShare with an initial embargo period (to be determined). The audio-visual data of educational materials will be deposited with Edinburgh DataShare without any embargo period and will be available immediately.

3. Integrity

How will you quality assure your data?

Quality of the data is ensured at every step. Quality during the data collection ensured by training and retraining of the data collectors. Field supervisors always cross check for accuracy of the data collected. Quality also be ensured during data entry; the mandatory fields must be filled up before proceeding to the next to avoid missing out of variables; the entries also rechecked by the biostatistics department data management personnel to ensure quality control. New data produced during the project will be derived from existing spatial and statistical datasets. For the qualitative data quality checks will be done at the time of transcription and the transcripts will be checked by another member of the research team for accuracy and completeness. For other data which includes audios, videos and handouts, the health workers are provided with these pre-recorded and pre-printed versions to be distributed to the patients and the public. Initial and periodic trainings are given to these health workers for demonstrating correct techniques of inhaler use and steps of breathing exercises which are checked by supervisors periodically.

4. Confidentiality

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

Information sheet was provided to all the participants and consent form signed by them after going through the information sheet. Such consent of the participant included the use of data collected for the present research and sharing such information with other researchers and dissemination of the information derived from such data through conferences and publications. This also included use of de-identified data in future for analysis or sharing with other partners or researchers after approval from the PI and such data will not require any further consent from the participant. As per our Institutional Review Board protocol, the identity of the participants will be protected and name and address or personal identifiers will be removed before the data is used for analysis. For interviews, the participants name was withheld and addressed as respondent with numbers (e.g. respondent 1). The transcripts of the interviews and the focus group discussions had de-identified data; all identifiers removed during the transcription process.

Publicly available documents and information coming out through conferences, dissemination workshops and research publications will be accessible by all; as such no permissions will be required from the primary researchers/authors except for the publication and access policy of the journal concerned. As regards to the access of metadata, it can be accessed through University of Edinburgh. The achieved data will be the de-identified quantitative and qualitative research data which will require permission from PI or Co-PI for access (depending on which repository it is deposited into).

5. Retention and Preservation

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

The de-identified and the anonymised data generated by this project (research data) will be available for use by the researchers and the policy makers for a minimum period of 5 years. The data will be held by the University of Edinburgh data repository, DataVault and can be obtained by permission. The aggregate data will be freely available to general public and researchers in perpetuity after an initial embargo period (to be determined); this data will be available at Edinburgh DataShare.

How will the data be preserved?

The research data will be preserved by the University of Edinburgh closed data repository, i.e. Edinburgh DataVault and can be accessed by researchers with permission and acceptance of a data sharing agreement. Preservation and access of the data has to be ensured by the University of Edinburgh.

6. Sharing and Publication

Which data will be shared and how?

The aggregate data generated out of this research will be freely shared with public and research community through research publications. The research data will be made available for a period of 5 years via controlled access repository, i.e. Edinburgh DataVault whereby permission of PI or Co-PI will be required; no charges will be levied for the data.

Are any restrictions on data sharing required?

The shared data will be de-identified data and the users seeking data will be provided access to be used only for research purpose and no attempt will be made for deducing the identity of participants. De-identified data on Edinburgh DataVault will require acceptance of the user to any data sharing agreement, once permission to the data is granted. As already mentioned, data can be accessed by request to the University of Edinburgh and with permission of the PI or Co PI.