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

Name:	Professor Dr Ee Ming Khoo
Project Title:	Process for Estimating Chronic Respiratory Disease (Asthma and COPD) burden in adults in Asian low and middle-income countries: A feasibility pilot. [4CCORD Study - 4 Country Chronic Respiratory Disease Study] - The 4CCORD study is coordinated and managed from King Edward Memorial Hospital and Research Centre (KEMHRC), PUNE, India with the support of the RESPIRE team, led from Edinburgh.
Institute:	University of Malaya, Kuala Lumpur, Malaysia
Start Date:	3rd September 2018
End Date:	31st August 2019
DMP version number and date:	DMP Version 1 and 01.12.2020

Responsibilities & Resources (applicable across the sections below)

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

Person in charge:
Principal investigator, principal investigator at the site and research manager of the project are listed below:

1. Principal Investigator: Professor Dr Ee Ming Khoo
2. Principal Investigator at the site: Associate Professor Dr Nik Sherina Hanafi
3. Research Manager: Jayakayatri Jeevajothi Nathan

Resources:
Hardware and software used in the management of data are listed below:

1. Laptop (encrypted and secured with username and password and kept in a locked cabinet)
2. Thumb drive (encrypted, password protected and kept in a locked cabinet)
3. Solid-state drive (encrypted, password protected and kept in a locked cabinet)
4. DataStore (to back up active research data)
5. Microsoft Word
6. Microsoft Excel
7. Locked cabinet for any hard copies (research protocol, participant information sheets, informed consents and questionnaires)

1. Data Capture

What data will be generated or reused in this research?

Data generated are as follows:

1. Quantitative - sociodemographic and personal profiles such as gender, ethnicity, date of birth, years of schooling and level of education; health profile; health care use and cost; smoking status; occupational exposure; biomass exposure; spirometry and peak flow rate.
2. Physician clinical assessment - clinical history, clinical examination (if necessary, repeat spirometry) - based on this assessment, the physician will diagnose the participant as having either asthma, COPD or 'other chronic respiratory diseases (CRD)'.

How much data will be generated?

2GB – 50GB

2. Data Management

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

Source documents were documented in Microsoft Excel. A codebook containing information about each of the variable in the dataset were created to ensure that the data is understood and interpreted properly.

Where will the data be stored and backed-up?

The electronic data were stored in two RESPIRE laptops, a thumb drive, a solid-state drive and will be backed-up in DataStore. Hard copies (research protocol, participant information sheets, informed consents, questionnaires and codebook) were stored in locked cabinet.

3. Integrity

How will you quality assure your data?

The completed questionnaires were checked by researcher for completeness of the data before it was entered into Microsoft Excel. Data entered were checked for errors by running a frequency distribution on each of the variables. Participants would be contacted for clarification if any missing data or error traced. Spirometry results were checked by researchers that includes investigators and pulmonary physician by looking at components such as the forced vital capacity (FVC), first forced expiratory volume (FEV1), forced expiratory time (FET) and good start, good peak, good

curve and good inspiratory loop on the graph before determining if the test results were acceptable. Spirometry tests were repeated if the initial test results were not acceptable.

4. Confidentiality

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

1. All Investigators and study site staff involved with this study will comply with the requirements of the General Data Protection Regulation (GDPR) 2018 with regard to the collection, storage, processing and disclosure of personal information and will uphold the Act's core principles. Access to collated participant data will be restricted to individuals from the research team treating the participants, representatives of the sponsor(s) and representatives of regulatory authorities.
2. Computers used to collate the data will have limited access measures via usernames and passwords.
3. All the information that is collected about the participant (such as name, age, sex, address, contact information) during the course of the research would be kept confidential. Published results will not contain any personal data that could allow identification of individual participants.
4. The data of each study participant will be identified with the help of unique identifier and it will be completely anonymized and scrambled before sharing. There will be no such information in the shared data which will disclose the identity of the study participant. Standard and recommended security measures and confidentiality with data sharing agreements will be in place with access control at every stage and audit trails maintained for all access and changes in data.

5. Retention and Preservation

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

Identifiable Data
All sociodemographic forms and informed consents of the participants will be kept for five years.

Non-identifiable Data
De-identified electronic data will be preserved for long term.

How will the data be preserved?

The hard copies of these identifiable data will be kept in a locked cabinet and will be shredded and disposed in secure bins after five years. The non-identifiable data will be preserved in two RESPIRE laptops, a thumb drive, a solid-state drive and on DataVault.

6. Sharing and Publication

Which data will be shared and how?

The final de-identified analysed data, conference abstract presentations, journal articles and dissemination event will be shared in DataShare.

Are any restrictions on data sharing required?

Yes, prior to release for sharing, there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, for data archived in Edinburgh DataVault, we will make the data available to users only under a data-sharing agreement that provides for:

1. A commitment to using the data only for research purposes and not to identify any individual participant;
2. A commitment to destroying or returning the data after analyses are completed.