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

Name:	Dr. Shams El Arifeen Dr. Ahmed Ehsanur Rahman
Project Title:	***Assessing the feasibility and effectiveness of introducing pulse oximetry in Integrated Management of Childhood Illness (IMCI) services to manage acute respiratory infections at the first level health facilities of Bangladesh***
Institute:	icddr,b
Start Date:	*September 1; 2018*
End Date:	*May 31, 2021*
DMP version number and date:	v.1 December 8, 2020

Responsibilities & Resources (applicable across the sections below)

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

Si no	Name	Designation	Role in data management
1	Dr Shams El Arifeen	Senior Scientist and Senior Director, Maternal and Child Health Division, icddr,b	**Principle Investigator-Bangladesh:** overall leadership and guidance
2	Dr Ahmed Ehsanur Rahman	Associate Scientist, Maternal and Child Health Division, icddr,b PhD student, University of Edinburgh	**Co-Principal Investigator-Bangladesh:** day-to-day operations and coordination regarding the development of data collection tools, data collection, data quality monitoring, data analysis and data sharing
3	Mr Qazi Sadeq-ur Rahman	Senior IT Development Specialist, Maternal and Child Health Division, icddr,b	**Data manager-Bangladesh:** overall coordination regarding data entry, data cleaning, data, data store and data sharing.

1. Data Capture

What data will be generated or reused in this research?

We employed both quantitative and qualitative methods of data collection to address the research questions and study objectives (described below). The proposed data collection method/methods for each of the research questions are outlined below:

Quantitative:

Research Question 1: Are the facilities ready to provide IMCI services, including the use of pulse oximetry?

- **Indicator/Thematic area of exploration:** Proportion of facilities having essential items related to IMCI services including pulse oximeter
- **Means of verification:** Health facility assessment checklist
- **Frequency of data collection:** Three times; once before introduction of the pulse oximetry (already done) and once at the middle and once at the end of implementation

Research Question 2: Do the IMCI service provider have knowledge regarding the use of pulse oximetry?

- **Indicator/Thematic area of exploration:** Proportion of IMCI service providers having knowledge regarding the use of pulse oximetry in clinical severity classification of pneumonia
- **Means of verification:** Structured interviews with IMCI services providers
- **Frequency of data collection:** Three times; once before introduction of the pulse oximetry (already done) and once at the middle and once at the end of implementation

Research Question 3: Do the IMCI service providers routinely use pulse oximetry in IMCI services?

- **Indicator/Thematic area of exploration:** Proportion of children clinically classified as 'Pneumonia' by IMCI service providers whose SpO2 was measured by the IMCI service providers using PO
- **Means of verification:** -Data extraction from IMCI Registers
- **Frequency of data collection:** Once every week (2-3 weeks of data extraction already done) and continuous throughout the implementation period

Research Question 4: Do the IMCI service provider appropriately use pulse oximetry in IMCI services?

- **Indicator/Thematic area of exploration:** Proportion of children receiving IMCI services whose SoP2 was measured as per the Standard Operating Procedure (SoP) by the IMCI service providers using pulse oximetry

- **Means of verification:** Structured observation of IMCI services and use of pulse oximetry
- **Frequency of data collection:** Twice, once at the middle and once at the end of implementation

Research Question 5: How much time does it take to obtain a stable SpO2 reading with pulse oximetry by IMCI service providers?

- **Indicator/Thematic area of exploration:** Mean/median time taken for obtaining a stable SpO2 reading with pulse oximetry by IMCI service providers
- **Means of verification:** Structured observation of IMCI services and use of pulse oximetry
- **Frequency of data collection:** Internal piloting (already done with four assessor). Twice, once at the middle and once at the end of implementation

Research Question 6: Is the SpO2 status measured with PO by the IMCI service providers valid?

- **Indicator/Thematic area of exploration:** Level of agreement between IMCI service providers and gold standard assessor regarding SoP2 status of children measured with pulse oximetry
- **Means of verification:** Reassessment of SpO2 by an independent assessor
- **Frequency of data collection:** Twice, once at the middle and once at the end of implementation

Research Question 7: What proportion of children clinically classified as 'pneumonia' have hypoxaemia (SpO2<90%)?

- **Indicator/Thematic area of exploration:** Proportion of children clinically classified as 'pneumonia' have hypoxaemia (SpO2<90%)
- **Means of verification:** Data extraction from IMCI Registers
- **Frequency of data collection:** Once every week (2-3 weeks of data extraction already done) and continuous throughout the implementation period

Research Question 8: Do the IMCI service provider use information related to SpO2 status of the children in decision making?

- **Indicator/Thematic area of exploration:** Proportion of children clinically classified as 'pneumonia' and have hypoxaemia (SpO2<90%) who are referred to higher level facilities
- **Means of verification:** Data extraction from IMCI Registers
- **Frequency of data collection:** Once a week and continuous throughout the implementation period

Research Question 9: What is the acceptability of introducing PO in IMCI service as reported by the parents and guardians of sick children?

- **Indicator/Thematic area of exploration:** Acceptable, perceived importance, perceived benefit, etc.
- **Means of verification:** Exit interviews with caretakers of the sick children receiving IMCI services
- **Frequency of data collection:** Twice, once at the middle and once at the end of the implementation period

Qualitative:
Research Question 10: What are the barriers and challenges related to the use of PO in IMCI services?

- **Indicator/Thematic area of exploration:** Related to PO, related to children, related to the work environment, others.
- **Means of verification:** In-depth interviews and focus group discussions with IMCI services providers
- **Frequency of data collection:** Twice, once at the middle (3 interviews already done) and once at the end of the implementation period

Research Question 11: What is the acceptability and usability of introducing PO in IMCI service as reported by the IMCI service providers?

- **Indicator/Thematic area of exploration:** Acceptability, user-friendliness, perceived importance, perceived benefit, etc.
- **Means of verification:** In-depth interviews and focus group discussions
- **Frequency of data collection:** Twice, once at the middle (3 interviews already done), and once at the end of the implementation period

How much will data be generated?

Si.	Method of data collection	Tools	Number of variables	Sample size	Approx. size of storage
	Qualitative				
1.	Health Facility Assessment	Health Facility Assessment Tool	40-50	Baseline: 10 Round-1: 10 Round-2: 10	5 MB
2.	Structured Interview	IMCI Service Provider Knowledge Assessment Tool	30-40	Baseline: 20 Round-1: 20 Round-2: 20	5 MB
3.	Structured observation	IMCI Service Provider Structured	50-60	Round-1: 20*15=300	10 MB

		Observation Checklist		Round-2: 20*15=300	
4.	Re-Examination	Re-assessment Form	50-60	Round-1: 20*15=300 Round-2: 20*15=300	10 MB
5.	Exit Interview	Exit Interview Tool	30-40	Round-1: 20*15=300 Round-2: 20*15=300	10 MB
6.	Data Extraction	Data Extraction Tool	20-25	600-700	20 MB
	Qualitative				
7.	In-depth Interview	IMCI Service Provider Interview Guideline		10-12	300 MB
8.	Key Informant Interview	IMCI Managers Interview Guideline		6-8	200 MB

2. Data Management

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

All quantitative data will be accompanied by a data dictionary and a data quality report showing if any inconsistencies or erroneous entries exist in the dataset. All qualitative transcripts should be accompanied by descriptive variables of respondent characteristics, the name of the researcher(s) and any field notes are taken to describe the situational, institutional and cultural contexts of the interaction with the researcher.

We will develop the following documents to ensure understandability of data.

1. Entry diagram presenting the availability of different data sets
2. Description of the data sets
3. Data codebook containing the variable name, description, type, formats, value label and missing value.

Where will the data be stored and backed-up?

Data Storage: Data will be primarily stored in SQL Database hosted by icddr'b server. To ensure the reliability and safety of data, we configure RAID6 in our server.

Data Backup: A data backup system will be scheduled with SQL Database that will automatically back up the databases every day in the same server, in different drive and another in a different server. We will also take manual backup weekly and store them in a different physical location.

3. Integrity

How will you quality assure your data?

We plan to take specific measures to assure the quality of data during data collection and data management. Followings are some of the key steps that will be taken to ensure data quality.

1. **Data collection:**
    a. **Data collection tools:** We developed the data collection tools based on the globally nationally accepted relevant data collection tools to ensure optimum validity and reliability (such as Demographic Health Survey Questionnaire, Service Provision Assessment Tool, WHO IMCI Service Readiness Assessment Tool, etc.). Every tool will be pretested and field-tested before finalisation.

    b. Data collectors: We recruited data collectors with relevant experience. We organised extensive training with theoretical and practical sessions before commencing data collection. All data collectors receive refreshers training once in the middle of each data collection cycle.

    c. **Monitoring and supervision team:** The data collection team will be supervised by an on-site supervisor with extensive experience. The on-site supervisors will directly be trained by the investigators. The on-site supervisors will conduct regular monitoring visits and recheck ~5% of the data collected by the data collection team. Weekly meetings will be organised between the data collection team and the respective supervisors to discuss data quality issues and specific gaps observed during routine monitoring visits. Also, the investigators will conduct periodic visits to the data collection sites to monitor data collection and adherence to the protocols.

    d. **Data quality review:** The immediate supervisors will review the data collection tools filled in by the data collection teams to assess the completeness and internal reliability. Immediate feedback will be given to the respective data collectors if and when necessary. The immediate supervisor will sign-off each of filled-in the data collection tools before sending for data entry. The investigators will recheck ~10% of the forms from each lot.

2. **Data entry:** We are developing a dedicated data entry interface for this project which will be user friendly and very easy to use; configured to match items on the study data forms. We are using various controls like textbox, date/time, list box, dropdown list, radio button etc. and this is also configured to match items like study data forms. The

data entry system will be secured by password, with different permissions for entry/editing granted to different study personnel. All types of data validation rules have in-built with the data capture software. The validation rules include range check; data type check; consistency check; abnormal values; logical check, cross-system consistency check, data Presence Check, Skips etc.

3. **Data cleaning:** Field Supervisors will be primarily assigned to check the filled-in hard copy for completeness, accuracy and internal reliability. As all types of validation rules will be implemented in the data entry software, almost cleaned data will be stored in the database. Moreover, after entering the data, the data manager takes the frequencies of the data and check for inconsistent data in the same tables or within different tables. The filled-in hard copy forms will be checked to address any inconsistency identified through the data cleaning exercise by the data manager. A separate log will be maintained to document the steps of data cleaning. All edits will be stored as new data sets to maintain the integrity of raw data sets.

4. Confidentiality

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

The University of Edinburgh will act as the data controller for this study and icddr,b will act as the data processor on behalf of the University of Edinburgh. The University of Edinburgh as the data controller and icddr,b as the data processor, will ensure that confidentiality and anonymity of all study participants are strictly maintained at every stage of data management, data analysis and dissemination.

The following steps will be taken to ensure the confidentiality and anonymity of all study participants.

1. **Ethical and Administrative Approvals:**The study is being conducted in accordance with the principles of the International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice (ICH GCP).
   Ethical and administrative approvals were obtained from the following bodies:

	Approval body	Organisation
1.	Institutional Review Board (ACCORD)	University of Edinburgh
2.	Institutional Review Board (Research Review and Ethics Review Committee)	icddr,b
3.	Field implementation permission	National Program Manager for Newborn Heath and IMCI, Directorate General of Health

		Services, Ministry of Health and Family Welfare of Bangladesh

2. **Informed consent process:** We are taking written and informed consent from all study participants. Also, special permission were be obtained from district health managers before collecting data from the selected facilities. Confidentiality and anonymity are be maintained at all stages of data collection and analysis.

3. **Linking records and anonymisation:** Individual identifiers are be assigned to allow the data entry team to link records that will be collected in this study. Once the records are matched, individual identifiers will be removed, and data will be anonymised. The linking dataset with personal identifiers will be stored separately. Qualitative transcripts have the participant and identifying place names removed, and these are replaced with careful descriptions of the role or characteristics of the removed names to enable interpretation of the data.

4. **Data security:** The data stored on icddr,b secure server will be encrypted so that those with the correct encryption key can only access it. Hence, the encryption key will only be available to members of the immediate research team who are working on the analysis of the data.
   The server on which the data is stored is protected by user-specific passwords. Below is a list of study personnel and the respective type of access:

Title	Type of Access
Data Manager/ Developers	Administrator – Modification of database structure, Add, Delete, View, Edit
Data Management Officer	View, Add, Delete, Edit, Update the values
Data Management Assistant	View, Add, Delete, Edit through Data Capture Software
Data Entry Person/ Data Collector	Only enter data through Software
Field Manager/ Research Investigator	View through the Monitoring system
Investigators and their delegates	We will provide Data files in a different format for Statistical analysis

Only qualified data entry personnel may handle the data and access the database. All data entry personnel will be instructed not to share their username and password. All data are considered confidential, and may only be shared with members of the research team and Investigator of the project.

5. Retention and Preservation

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

All physical records containing personal identifiers will be stored under lock and key at icddr,b at all times. After entering and storing the data in the database software, that means digitalise the data; we will scan the hard copy of the data forms, compare and check the identification part of scanned copy with identification part of entered data and destroy the hard copy of data forms. A copy of the scanned form will be stored in two different icddr,b-server. Also, we will keep another copy in an external hard disk in a separate physical location. We plan to retain and preserve the digitalised data coming out of this project for the long-term by utilising icddr,b infrastructure and ensure continued access to authorised personnel.

How will the data be preserved?

There will be two datasets. The master dataset will not have any personal identifiers. All personal information will be stored in a separate dataset with linking information to merge the personal identifiers dataset with the master dataset.The master dataset and personal identifiers dataset will be stored securely and separately in the icddr,b- server. Access to these datasets will be password protected and will be restricted to the investigators of the study and the data management team leads (also described in the previous section). In aligning with the icddr,b and The University of Edinburgh internal data archival policy, all physical records containing personal data will be stored for five years at icddr,b.

6. Sharing and Publication

Which data will be shared, and how? AND

Are any restrictions on data sharing required?

1. **Intellectual Property Rights Issue:** Ownership of the data arising from this study resides with the study team. Based on the agreement between the University of Edinburgh and icddr,b it has been decided that The University of Edinburgh will be the data controller and icddr,b will act as the data processor. The University of Edinburgh and icddr,b will have joint ownership of the data.

2. **Data sharing with the University of Edinburgh:**

There may be some data/items that we can share openly in Edinburgh DataShare, and there may be some data/items that we cannot share openly, and therefore we will consider depositing the data in Edinburgh DataVaults. For the latter, we will ask each potential user to read the icddr,b data access policy (https://www.icddrb.org/images/stories/AboutUs/icddrb_data-access-policy_nov-07.pdf) and the icddr,b data licensing application and agreement (https://www.icddrb.org/images/stories/AboutUs/icddrb_dlaa_20140518_1.doc. The application will be referred to the icddr,b senior manager of research administration.

After entry and necessary cleaning icddr,b will deposit the Datasets into a secure central storage repository (Data Vault) at the University of Edinburgh. The data provided will be the most recent, complete versions that have undergone all quality control and quality assurances steps.
All the Datasets will be stored on servers of icddr,b and the University of Edinburgh for future use by the investigators. icddr,b's server will be maintained by icddr,b and the University of Edinburgh's server will be maintained by the University of Edinburgh. The repositories at icddr,b and UoE will be managed and curated by respective data managers. Direct access to the repository will only be available to authorised Project members.

3. **Data sharing with other researchers:** icddr,b recognises the public health, social and intellectual value of providing access to its data to increase the number of their innovative analyses. Considering the future use of the data and prevailing of potential research findings, this data will be available after three years of protocol completion date as recorded by the IRB, publisher, or extraordinary circumstances. We will share such data according to the following policy (see also the links above within this section):
   - Data from this study (the Datasets) will be provided to interested researchers (Recipients) for purposes of secondary data analyses upon approval of a Data Licensing Application & Agreement (Application) by the icddr,b Data Repository Committee (DRC).
   - The Datasets are subject to any and all interests and limitations imposed by statutory requirement, icddr,b institutional policy, or written agreement with any third-party institution or individual (Interested Third Party). Requests for the Datasets that would negate or infringe any such interest will not be approved.
   - Only those data that icddr,b's Internal Data Policy, donor and publisher requirements, and ethical considerations permit will be made publicly available in the Datasets.
   - Requests from interested researchers to conduct secondary data analysis of the Datasets will be considered by the DRC upon receipt of a completed Application and in discussion with the PI or her/his designated co-investigator, which must include a thorough description of the research project for which the data are requested. Recipients must also include a copy of their own institution's Institutional Review Board (IRB) approval of their study. In the event the recipient is not affiliated with an institution that has an IRB satisfactory to the DRC, a full protocol must be approved by the icddr,b internal IRB process before the application is approved.
   - The Datasets will only be provided after removing all personal identifiers. However, due to the wealth of individual-level data available on study participants, the possibility of direct identification of a study subject cannot be completely eliminated. Therefore, only IRB approvals from an expedited or full review can be accepted. Waiver for exemption of IRB approval will only be accepted from icddr,b IRB.
   - As more fully set forth in the application, use of the Datasets shall be strictly limited to use in and for the purpose of the described research project; may not be shared, sold, transferred, or otherwise released to another party without the written consent of the DRC; and must strictly comply with subjects' informed consent and all applicable IRB requirements. Under no circumstances may a Recipient attempt to identify or contact study subjects.

- The rights and interests of the respective icddr,b PI and the University of Edinburgh Investigators shall take priority over the requests of Recipients, and no grant of access to the Datasets shall be construed or allowed to infringe such priority rights and interests.
- Considerations of the DRC will include but not be limited to whether the request is appropriate and consistent with the mission of icddr,b; b whether the proposed research is consistent with the subjects' informed consent; c. Whether icddr,b researchers have been or will be involved in the development and execution of the proposed secondary analysis; d. Whether the requested access would result in duplicative research and, if so, whether granting such access would be inappropriate under the circumstances; and e. The capacity of the requesting party to correctly understand and/or interpret the requested data.
- In an abstract, manuscript, publication, dissemination, presentation or other public disclosure based on the Datasets, their PIs, icddr,b and the University of Edinburgh Investigators, and the original donor must be properly acknowledged. Recipients publishing secondary analyses based on icddr,b Datasets must give each icddr,b Researcher who made an intellectually significant contribution to the dataset an option to be listed as an author, provided it would meet international standards for authorship in peer-reviewed journals and the policies of the publishers.
- If the recipient identifies an apparent error or inconsistency in the Datasets, Recipient must convey it to the DRC as soon as is reasonably possible.
- Any proposed abstract, manuscript, publication or other formal public dissemination based on the Datasets must be submitted to icddr,b for its review and comments at least 30 days prior to submission or presentation.

4. **Publication:** The usage of the Datasets will be for scientific and academic purposes; inform the routine information systems and to help governments, policymakers and programmers plan for health services better. Findings arising from the analyses of the Datasets will be disseminated through reports, meetings, workshops, conferences and in peer-reviewed journals. All publications arising from the analyses of the Datasets will follow standard criteria of authorship (Vancouver convention- www.icmje.org) will be followed, including joint publications for pooled datasets. All the investigators will have access to the data and will be offered authorship of manuscripts written from this study. At the time of publication, fully anonymised data will be provided in accordance with the open access requirements of peer-reviewed journals. Any publication, presentation, display or release of information online, to the public, to the press, or to any third party (whether in written or verbal format) in connection with the project (each a "communication") shall be made in accordance with clauses per an agreement between RESPIRE and icddr,b. The Publishing Party is expected to follow acknowledged good practise when publishing under the project as detailed in guidelines issued by, for example, the Committee on Publication Ethics (http://www.publicationethics.org/), the council of Science Editors (http://www.councilscienceeditors.org/) and the ARRIVE guidelines (http://www.nc3rs.org.uk/arrive-guidelines). The Publishing Party, before making a communication, will provide the text of the communication.