Plan Overview

A Data Management Plan created using DMPonline

Title: Social-Action Messages to Reduce Transmission of COVID-19 in North India

Creator: Wendy Olsen

Principal Investigator: Wendy Olsen

Data Manager: Wendy Olsen

Affiliation: University of Manchester

Template: UKRI Template Customised By: University of Manchester

Project abstract:

In India, rapid transmission of the SARS-CoV2 virus could mean a sudden health shock involving great expense to some families during 2020-2022. This project in 2020 aims to improve understandings around health messaging and the transmission of viral disease through a series of three activities – secondary data combining, issuing leaflets, and grassroots narratives. The collaboration we already started across north India includes participants in a series of six rural civic workshops in 2019, participants in a national workshop on women's Labour Supply in 2020, and participants in several workshops in Delhi over the years 2018-2020. Building up this network, we include women's studies specialists, development economists, population demographers, and activists for the groups that face discrimination in the society. For all these, the use of quantitative evidence in data-combining is potentially empowering. The mixing of local voices with such evidence is very powerful. To have local leaflets with a strong validation mechanism through our partnership will strengthen not just the COVID19 response but social development generally over the period 2020-2022.

ID: 56545

Last modified: 20-07-2020

Grant number / URL: Global Challenges Research Fund pump priming grant (Research England)

Social-Action Messages to Reduce Transmission of COVID-19 in North India

Manchester Data Management Outline

1. Will this project be reviewed by any of the following bodies (please select all that apply)?

- None of the above

2. Is The University of Manchester collaborating with other institutions on this project?

- Yes - Part of a collaboration and owning or handling data

3. What data will you use in this project (please select all that apply)?

- Re-use existing data (please list below)
- Acquire new data

Acquire new data

20 long phone calls take place: one each with 10 rural Panchayat leaders (one person from each of 10 villages known to the AMS staff, in the area surrounding Lucknow and Kanpur). The leaders are half female. Each is initially asked to comment on the leaflets and on the issues of transmission, awareness, and sanitation. Based on local discussions in their village about what is feasible and what is being said, the leader will give us suggestions for revision of each leaflet during the SARS-CoV2 period. We record part of each conversation under the permission of participants.

Secondary data

Dataset 1: NFHS 2015/2016

Dataset 2: Delhi NCR Coronavirus Telephone Survey- Round 1 (subject to approval and depersonalisation of the microdata)

We focus on common variables (age, sex, social group, state, house type, health indicators), and identify additional variables from Dataset 2 if available. These cover: how aware of the disease the person was by 3 April 2020. B. how aware they are about sanitation reducing the transmission of this virus. C. How many social-physical contacts the person had in the past one day in April 2020

4. Where will the data be stored and backed-up during the project lifetime?

- University of Manchester Research Data Storage Service (Isilon)

1. Dropbox for Business (only for collaboration and sharing)
2. University of Manchester Research Data Storage Service (Master copy of data)

The research data storage space created earlier for Prof. Olsen can be used to hold these new data. This data storage appears as R: drive on her desktop pc. When necessary for sharing data across Indian sites, we will use encrypted file transfer and Dropbox for Business. Prof. Olsen already has a University of Manchester Dropbox for Business which can be used to hold data securely. The address list held in the data storage area will not contain sensitive personal data. The other research data types do not contain personal data.

5. If you will be using Research Data Storage, how much storage will you require?

- < 1 TB

Wendy Olsen was allocated some storage space on a previous project so she understands how to use the replicated and snapped access. This is used as a safe backup area. The access to this data storage is via Univ of Manchester VPN. Please note that we will instruct all research team members to delete all unused data from their drives.

6. Are you going to be working with a 3rd party data provider?

- No

7. How long do you intend to keep your data for after the end of your project (in years)?

- 0-4 years

Questions about personal information

Personal information, also known as personal data, relates to identifiable living individuals. Special category personal data is more sensitive information such as medical records, ethnic background, religious beliefs, political opinions, sexual orientation and criminal convictions or offences information. If you are not using personal data then you can skip the rest of this section.

Please note that in line with data protection law (the General Data Protection Regulation and Data Protection Act 2018), personal information should only be stored in an identifiable form for as long as is necessary for the project; it should be pseudonymised (partially de-identified) and/or anonymised (completely de—identified) as soon as practically possible. You must obtain the appropriate ethical approval in order to use identifiable personal data.

8. What type of personal information will you be processing (please select all that apply)?

- Personal information, including signed consent forms

Personal information is collected on each interviewee. This information will be retained for 5 years after our Working Paper is published. The information includes their age, gender and occupation

(broad category). We mean to keep the personal contacts (name and phone numbers) only for the initial contact purpose and do not include in the report. During the interview, we ask consent for future contact. If persons allowed for further contact, we keep their contacts to provide them with feedback and information on how to reduce infection risks. As soon as the process is finished, name and phone numbers are removed from the shared folder.

9. Please briefly outline how you plan to store, protect and ensure confidentiality of the participants' information.

Personal data held with caution.

1. Keep the excel spreadsheet files under password encryption at all times, and the single paper original under lock-and-key at all times.
2. Use only Encrypted Laptops, Encrypted USB sticks and secure filestore.
3. all other data held in this project are textual data with no names, no personal data. This includes no significant place names except broad city names such as Lucknow, Delhi, etc.

10. If you are storing personal information (including contact details) will you need to keep it beyond the end of the project?

- Yes – Other (explain below)

11. Will the participants’ information (personal and/or sensitive) be shared with or accessed by anyone outside of the University of Manchester?

- Yes - Informal sharing without contractual arrangements

12. If you will be sharing personal information outside of the University of Manchester will the individual or organisation you are sharing with be outside the EEA?

- Yes

Yes, as described earlier, the other holders are resident in India.

13. Are you planning to use the personal information for future purposes such as research?

- Yes

14. Who will act as the data custodian or information asset owner for this study?

Wendy Olsen

15. Please provide the date on which this plan was last reviewed (dd/mm/yyyy).

2020-06-01

0. Proposal name

0. Enter the proposal name

Social-Action Messages to Reduce Transmission of COVID-19 in North India

1. Description of the data

1.1 Type of study

We aim to improve health and well-being in Uttar Pradesh state of India in this project. This project improves health messaging about COVID19 through secondary data combining, survey data by telephone mode, semi-structured interviews, and as an output, issuing leaflets in 2 languages. Our further analysis of grassroots narratives will be published in academic outlets.

1.2 Types of data

BROAD OVERVIEW. We acquire new data via phone.

First we use Secondary Dataset: NFHS 2015/2016

We focus on common variables (age, sex, social group, state, house type, health indicators), and build a regression model. This is to be analysed statistically without reference to the semi-structured interviews nor our original phone survey.

TYPES OF DATA HELD.

LIST 1. Non-sensitive, non-personal data.

a) secondary data with permissions;

b) survey data by telephone mode, each max 20 minutes; informed consent is verbal.

c) semi-structured interviews, also by telephone mode; each max 60 minutes; informed consent is verbal.

LIST 2. Potentially sensitive personal data.

The key sensitive personal data are:

d) The list of phone-interview names of Gram Panchayat (Village) elected heads. This list is going to be held in India only. Available in an updated format at URL https://panchayat.gov.in/sarpanch-address. We are not sharing the phone numbers with anybody.

e) Phone call transcripts which, prior to translation, are held overseas. These cross the desk of translators in India.

The 'survey' [which means a quantitative questionnaire] will be purely anonymous numeric, coded data. These are sent to the UK from India as rectangles of numbers.

The respondents are all professionals, who have been elected to a public position in a Gram Panchayat (Village area) in Uttar Pradesh. None are unemployed. Therefore, none of our respondents are vulnerable, and none are children.

The anonymised transcripts will be sent to the UK. No data held in the UK from this project will hold personal information. No data held in the UK will be sensitive personal information.

1.3 Format and scale of the data

LIST 1. Non-sensitive, non-personal data.

a) secondary data with permissions: National Family and Health Survey India.

b) survey data by telephone mode, We invite 3000 people to response. Out of these only 50% response rate is expected. These 1500 respondents then, each of whom is a Gram Panchayat leader, were each already elected to hold this role as a job. 15 questions per respondent. Excel spreadsheet 1500 rows *15 columns.

c) semi-structured interviews, also by telephone mode, with a selection of 10 of the same village leaders. See 'NVIVO data' below.

List 2. Potentially sensitive personal data.

The key sensitive personal data are:

d) The list of phone-interview name list with phone numbers and addresses - held in India only. (These are overseas-held qualitative data, personal, potentially sensitive).

3000 officially and publicly known Gram Panchayat leaders' names and their phone number, which is published in India. The India counterparts ring the number, ask for participation, and if agreed, hold the name, number and other information in the India desk of the sub-contractor, IIDS. We expect 1500 will refuse and 1500 say 'yes' to begin participation. The name and phone-number info is not sent to the UK.

e) Phone call transcripts which, prior to translation, are held overseas. These cross the desk of translators in India.

The transcripts must contain the name of the respondent (who is a professional) while being translated. Then names are removed. Then the numbered transcripts are inserted into NVIVO software. Finally the anonymised Transcripts are also sent to the UK from India. They are in 2 languages. English, Hindi.

2. Data collection / generation

2.1 Methodologies for data collection / generation

This project uses broadly three types of data: secondary data, survey data by phone more, and interview data also gathered via phone mode.

SECONDARY : Indian NFHS 2015/2016 is an open dataset, which is commonly used in many studies on health in India. This dataset contains depersonalised microdata and no personal information.

SURVEY: We also gather primary data through a short phone survey (net number = 1500 respondents) and semi-structured phone interviews (10 of the 1500 respondents, a second phone call), all of whom are rural community leaders. The *survey questions* include general questions about how much they are aware of Covid-19, whether their communities are prepared for potential risks of infections, and how they deal with neighbours and family in terms of social distancing, sanitation, and hygiene. (*document attached in Ethics RM).

INTERVIEWS: We extend the phone call to a second call for a longer, tape-recorded phone interview for 10 cases, half female and half male. (Document also attaced as Interviews, in Ethics RM)

The tapes are a data medium (MP3 for example) which is non-anonymous. No tapes are sent to the UK. The India based colleagues destroy the MP3 recordings as soon as our transcription work is done.

2.2 Data quality and standards

a) Secondary data is available through the website. NFHS provides the full dataset through the DHS website. We have submitted the aim of this research and the name of our institute, and datasets are immediately provided. Proper citation will be written on the report. Each research organisation asks for permission separately (UOM, IIDS).

Re Data Types b, c:

b) Interviews are done bu the IIDS staff. 3000 short calls are made. From among these, around 1500 refuse to participate. The remaining 1500 15-20minute phone calls involve IIDS staff member ticking off answers on the survey form in their computer.

c) 10 20-minute long interview phone calls take place: one each with 10 rural Panchayat leaders (one person from each of 10 villages, within the same 38 districts surrounding Lucknow and Kanpur, UP, India.

Based on local discussions in their village about what is feasible and what is being said, the leader will give us their view of what has happened so far in the management of COVID19 and SARS-COV2 in their area.

NOTE:

We ensure quality of translation of the survey and interview questions by having them translated to Hindi then back-translated to English. Piloting took place in Delhi with non-vulnerable professionals.

3. Data management, documentation and curation

3.1 Managing, storing and curating data

re DATA TYPE (a) The secondary NFHS data is not shared among our staff. Instead we share code. Each staff member is responsible for their own registration with the NFHS.

Then, using the 'code', the staff can re-create the same dataset on each desktop using 'derived variables'.

DATA TYPES b and c:

Personal data: Staff members will use encrypted laptops and desktops for the temporary storage of the name and contacts, and only depersonalised data will be exchanged through Dropbox and

maintained on the Univ of Manchester Research Data Storage. Personal data requires following the laws of UK, EU, and India. Requires SFTP and should not be saved on a laptop or desktop for any long period of time (not greater than 3 weeks). While held on any desktop, laptop or phone, the personal data must be encrypted.

Non-Personal Data: we do have initially a Dropbox for Business, which will be regularly backed up. We use this to share non-personal research documents. This can only be used by the project team for non-personal data. Then, we will use Univ of Manchester secure data transfer, known as SFTP via UOM, to keep our transfers safe from outsiders.

- Naming conventions: Districts are named using Census codes. Villages are named via the Gram Panchayat ID, which is an official code.
- Version control: The master copy is in UOM Research Data Storage, and is anonymous. Prof. Olsen is responsible for ensuring each file has the date in its filename, allowing version dominance.
- Clear, consistent file structure: The files are ordered via naming carried out in India by IIDS. These include subdirectory for English and subdirectory for Hindi for each document set. The document sets are:
- b) SURVEY DATA. English prevalent. A few Text items are also translated into Hindi.
- c) INTERVIEW DATA. Balance of English and Hindi, obtained by using the natural native language of each respondent, then translating to the other language. Here the directory structure is:
    - Interviews\Original\Audio and \Transcript.
    - Interviews\TranscriptsAllEnglish
    - Interviews\TranscriptsAllHindi
- When sending the files to the UK, the Audio files are omitted but the three-fold directory structure is maintained. This allows us to see what the original (native, or 'first') language of the interview was.
- When an interview is part English part Hindi that works fine, transcribe it VERBATIM. All semi-structured interviews must have VERBATIM exact transcripts.

- NOTES: Any hard copies of data will be securely stored e.g. locked filing cabinets in secure premises
- Research Data Storage is secure and replicated (i.e. automatically backed up).
-

3.2 Metadata standards and data documentation

Re DATA TYPE c) Each semi-structured interview is up to 60 minutes long. This provides 10 hours, i.e. 200 pages of transcripts, potentially. The data is analysed using NVIVO software.

How we analyse the qualitative data:

The team analyses the keyness of each single word in the transcripts using a novel statistical routine, and sets of words are counted using NVIVO. Next, discourses are then analysed using NVIVO. Some of the work is done in India, and some in UK. NVIVO is only used with anonymised transcripts.

IIDS has clear ethical guidelines about how to handle data, ensuring we conform with local laws on data protection. Translators are managed strictly within IIDS using IIDS usual methods of keeping personal data under careful management, and under Indian law, and IIDS is aware of the UK Laws on data protection. We hold various files such that, there will be a full list of all files, with the date of origin.

SHARING THE QUALITATIVE DATA: A single NVIVO file is created (with no personal data) and shared among the research team members who are co-authoring.

3.3 Data preservation strategy and standards

UOM Side

We write the ethics application (separately), with all attachments through Ethical Review Manager (ERM). Team members fully comply with the university ethics guidelines. Training occurs annually for data protection and ethics.

Regarding both the survey and the semi-structured interview, our PI and Co-PI staff ensure that no sensitive or personal questions are asked. All interviewees are adults who are leaders of the communities. It is also essential to protect researchers' privacy. We avoid revealing personal information of the researchers to participants.

INDIA SIDE

The Audio recordings will be deleted after transcription and translation into both Hindi and English.

The transcript originals, which are not anonymous at the moment of receiving them from the translator, (showing village name for example) will be held in a secure manner *only* in the India offices following the IIDS' own standards of ethics and data protection.

Then after anonymisation, the survey data and transcripts, showing only village Number and not phone numbers nor names, will be sent to the UK using safe encrypted methods.

At that point they enter the Research Data Storage service.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

The University of Manchester Information Security Policy (http://documents.manchester.ac.uk/display.aspx?DocID=6525) aims to protect information through controls and responsibilities which are in line with recognised information security standards and which support compliance with relevant legislation.

Detail:

1. Personal data: we plan to preserve the personal data initially in the India offices of IIDS, but then delete it all 5 years after project is finished, ie our working paper is published.

2. Non-personal data: using NVIVO we store this for future analysis. There is no need to destroy this after 5 years as it holds no personal identifiers and there is no risk of disclosure of personal ids.

WHAT-IF SITUATIONS (Hacking)

In any case the respondents, being professionals elected to public office, are able to deal with any risk if a village were identifiable and if a particular quote were traced back to them. They are not vulnerable people. The process of 'de-anonymisation' which a hacker could engage in is not a risk here because even if the professional Sarpanch [Mayor] were faced with their own quotations they would be able to deal with any challenge or criticism.

SENSITIVE TOPICS

The level of deaths and severe COVID19 cases is also mainly a public matter, not a private matter in India. Every case must, by law, be registered. Every death must, by law, be registered. The commentaries made in interviews by Sarpanches [Mayors] could be considered to touch upon sensitive

topics in view of the epidemic.

We will be dealing with 'distress' and having a 'de-brief' as good practice although these are non-vulnerable adult professionals.

4.2 Main risks to data security

Personal data: team members will use encrypted laptops and desktops for the temporary storage, and exchange data on encrypted USB sticks or via SFTP (secure file transfer). Then, we delete the old copies.

The IIDS adds pseudonyms to each village leader in the 10 semi-structured interview transcripts. They anonymise this information at the earliest opportunity.

We store the excel and NVIVO files and the word files of transcripts and list of Villages on the Univ of Manchester Research Data Storage system. These do not include sensitive personal information, indeed the files do not include personal information.

Non-personal data: each team members will be granted access to the secure file transfer or file store arrangement. Dropbox for Business will offer research data sharing between research team members of this project. The data will be restored in the University of Manchester Research Data Storage.

5. Data sharing and access

5.1 Suitability for sharing

Any of our published papers, outputs, and webinars will not include personal or sensitive information. Since it is not personal data, no restrictions on data sharing are required. We keep the remaining raw data, other than published data, in Dropbox with password access to restricted members of the research team only.

1. We will write a working paper, including the analysis of interview and data combining results.
2. We will later publish leaflets to acknowledge the information about Covid-19 and sanitation and hygiene. Leaflets are shared online. We also plan to use webinars to disseminate the findings.
3. Using Twitter, Facebook or other online platform, we share the research results, emphasising the importance of communications and partnership as well as sanitation and quarantine in reaction to the disease Covid-19.

Primary data sources will not be shared with anyone other than team members. For ensuring trust in research, we do not allow sharing of our NVIVO files at all beyond our research team.

5.2 Discovery by potential users of the research/innovation data

Not applicable

5.3 Governance of access

Not applicable

5.4 The study team’s exclusive use of the data

Not applicable

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

We do not intend to submit the data to the UK Data Archive. Resources in the project are not extensive and so these data cannot be deposited.

5.6 Regulation of responsibilities of users

Not applicable

6. Responsibilities

6. Responsibilities

UK Side

Prof Wendy Olsen. She delegates some reminders to team members. At all times, an awareness of ethics and data security has been developed in this team. Regular reminders both verbally and in writing will be given.

INDIA Side

Prof Pal. As Director of IIDS, Prof Pal ensures that the project gets the Ethics Clearance certificate by end of June 2020. (*Certificate is enclosed in the ERM.*)

Negotiations were underway in May-July 2020 to finalise the methods of enquiry. We are balancing safety and health/vulnerability risks. Prof Pal trains the staff at the IIDS and ensures that legal requirements are met.

STAFF support

Regular reminders both verbally and in writing will be given by Prof Olsen and Prof Pal.

7. Relevant policies

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	URL or Reference
Data Management Policy & Procedures	http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=33802
Data Security Policy	https://documents.manchester.ac.uk/DocuInfo.aspx?DocID=14914
Data Sharing Policy	Not applicable
Institutional Information Policy	https://documents.manchester.ac.uk/display.aspx?DocID=6525
Other	
Other	The University of Manchester Records Management Policy http://documents.manchester.ac.uk/display.aspx?DocID=14916 The University of Manchester Publications Policy http://documents.manchester.ac.uk/display.aspx?DocID=28526 The University of Manchester IT policies and guidelines http://www.itservices.manchester.ac.uk/aboutus/policy/ The University of Manchester Intellectual Property Policy http://documents.manchester.ac.uk/display.aspx?DocID=24420
In India	Legal system in India is covered via the IIDS ethics and governance procedures.

8. Author and contact details

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details

Wendy Olsen, phone 07891 266635 , email wendy.olsen@manchester.ac.uk. During COVID19 time I often use skype wendyolseninmanchester .

Assistance by Ms Jihye Kim, Research Assistant, July 2020. Email jihye.kim@postgrad.manchester.ac.uk.