Data Documentation Template

1 Study Description

1.1 Identification

1.1.1 Title

Contains the full authoritative title of the data collection. A full title should indicate the geographic scope of the data collection as well as the time period covered.

Lending a hand: developing a support structure for young migrants in South Africa and Uganda

1.2 Overview

1.2.1 Abstract

An unformatted summary describing the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they conducted the study. A listing of major variables in the study is important here.

We use mixed methods to adapt, implement and evaluate the feasibility of an intervention. We will use participatory methods working with youth advisory groups in both countries already engaged in the design of this project, to develop and implement the intervention. Then we will conduct a mixed methods process evaluation to assess which facets of the package are valued and test our assumptions about the challenges young migrants face and the appropriateness of the intervention offered.

The work is divided into four phases across 30 months: Phase 1 (months 1-3): recruitment of staff and other preparations; Phase 2 (months 4-9): Baseline assessment and co-designing the intervention with young people and other stakeholders; Phase 3 (months 10-21): Implementation of the intervention; Phase 4 (months 22-30): Analysis, dissemination and discussions on refinement of the intervention and, if the intervention is feasible, the design of a full-scale trial.

1.2.2 Kind of Data

The type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, etc. No description just a single phrase, e.g. Genetic sequences

Types of data
Quantitative survey data will be captured at the beginning and end of the intervention period, and from follow-up of participants from the first survey who go on to use our services. Qualitative data, using rapid qualitative assessment (generating observational, interview and group discussion data), will be collected prior to intervention implementation to update existing data on the sites (collected during preparatory work documenting migration experiences of young people). During the intervention, diary data documenting numbers of contacts with support structures and records of call hub usage will be collected. In-depth interviews with selected participants and other stakeholders will be undertaken.
Format and scale of the data
Quantitative data will be collected offline on encrypted Android tablets using Open

Data Kit (ODK) or REDCap. ODK/REDCap are well-established software for electronically collecting and managing data. The advantage of using these electronic data capture systems is that they allow internal validation and consistency checks to be incorporated during the data collection process. The data we collect will be encrypted and subsequently securely synced to secure servers at each site. We will generate one record per participant per survey (i.e. if a participant is sampled twice a record will be made in both databases).

Some in-depth interviews and discussion groups will be recorded at the time of data collection. These recordings will be transcribed and translated in English and stored as password protected word documents on MRC/UVRI and LSHTM/AHRI secure servers. Access to the server folders are restricted only to specific research staff working on the study. The recordings will be deleted after transcription.

Observational data and interview/discussions where it is not feasible to use recordings (where the participant does not agree to recording, for example) will be in note form at the time of data collection and will be written up into full scripts and stored in the same way as the transcription files.

1.2.3 Unit of Analysis
Basic unit(s) of analysis or observation that the study describes:

n/a

1.3 Scope

1.3.1 Topics Classification
MeSH subject headings

n/a

1.3.2 Keywords
Keywords summarize the content or subject matter of the survey. As topic classifications, these are used to facilitate referencing and searches in electronic survey catalogues.

Adolescents; mobility; migration; South Africa; health; wellbeing

1.4 Coverage

1.4.1 Country
Indicates the country or countries covered in the file

South Africa

1.4.2 Geographic Coverage
Information on the geographic coverage of the data. Include the total geographic scope of the data, and any additional levels of geographic coding provided in the variables.

PIP area – and Mtubatuba, KwaZulu-Natal

1.4.3 Universe
A description of the population covered by the data in the file; the group of persons or other elements that are the object of the study and to which the study results refer. Age, nationality, and residence commonly help to delineate a given universe, but any of a number of factors may be involved, such as age limits, sex, marital status, race, ethnic group, etc. The universe may consist of elements other than persons, such as specimen, sample or isolate. In general, it should be possible to tell from the description of the universe whether a given individual or element (hypothetical or real) is a member of the population under study. Also known as universe of interest, population of interest, and target population.

All young people who have migrated, with particular attention to young people who have moved into or out of UMkhanyakude (who are within twelve months of their move to the new location) while still providing advice to migrants of longer standing, who may be planning a further move or trying to make a change in their employment, for example

1.5 Producers and Sponsors

1.5.1 Investigators

The person, corporate body, or agency responsible for the data collection's substantive and intellectual content. Repeat the element for each author and use the affiliation attribute if available. Invert first and last name and use commas. Remarks: The author in this element should be the individual(s) or organization(s) directly responsible for the intellectual content of the data collection.

Name	Affiliation
Janet Seeley	AHRI/LSHTM

1.5.2 Funding

The source(s) of funds for production of the data collection. If different funding agencies sponsored different stages of the production process, use the role attribute to distinguish them. Also includes a field for the grant/contract number of the project that sponsored the data collection effort.

Agency	Abbreviation	Grant number	Role
MRC/NIHR	MRC/NIHR	MR/S023607/1	

1.5.3 Acknowledgements

Statements of responsibility not recorded in the title and statement of responsibility areas. Indicate here the persons or bodies connected with the work, or significant persons or bodies connected with previous editions and not already named in the description. For example, the name of the person who cleaned the data collection might be cited here, using the role and affiliation attributes. Does not include funders.

Name	Affiliation	Role

1.6 Sampling

1.6.1 Sampling Procedure

The type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction

For the baseline survey : The sample size calculation for the baseline and end-line surveys is based on a simple one sample test of proportions to demonstrate acceptability defined as the proportion of our participants who use our services is at least 50%. To do this, we will simply count the proportion of baseline participants offered the intervention who make contact with our peer-supporter intervention by three months. In the end-line survey, following 9-10 months of intervention delivery, we will measure the proportion of the population who have engaged with at least one of the health-related, skills-related or social service components of the intervention. We have been conservative with our sample size calculation, allowing for the fact that sample size estimates such as these are very sensitive to the assumption of the expected proportion. We expect uptake to be 60%; as long as it is at least 50% we will achieve power of greater than 94%.

<<Add here>>

1.7 Data Collection

1.7.1 Dates of Collection

Contains the date(s) when the data were collected/produced.

Data collected across 2020-2022

1.7.2   Notes on Data Collection/Production

Used to describe noteworthy aspects of the data collection/production.

<<Add here>>

1.8   Data Processing & Appraisal

1.8.1   Other Processing

Used to indicate additional information about the methodology and processing involved in the production of the dataset.

<<Add here>>

1.9   Data Access

We will add these bits

1.10   Contacts

1.10.1   Contact persons

Individuals listed as contact persons will be used as resource persons regarding problems or questions raised by the user community. The URI attribute should be used to indicate an URN or URL for the homepage of the contact individual. The email attribute is used to indicate an email address for the contact individual.

Name	Affiliation	Email	URI
Janet Seeley	AHRI/LSHTM	Janet.seeley@ahri.org	

2   File Description

2.1   Data Files

2.1.1   Contents

Abstract or description of the file. A summary describing the purpose, nature, and scope of the data file, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer when they created the file. No need to repeat information already contained in the abstract in paragraph 1.2.1.

<<Add here>>

3   Variable Description

A code book of the variables in the data file.

Name	Definition	Data Type and Codes