

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.

Multilevel HIV prevention cohort study

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.

This study seek to arrest the HIV epidemic and its negative impact on young people in SSA communities by improving the uptake, retention and adherence of multi-level combination HIV prevention by AGYW and male partners. Specifically, the interventions supported by PEPFAR's Determined, Resilient, Empowered, AIDS free, Mentored, and Safe (DREAMS) program. Our central hypothesis is that effective uptake and retention of interventions by young people requires activation of interpersonal, family and community-level networks to support tailored and adaptive intervention delivery.

The aims of the study are:

- To quantify how multiple levels of social dynamics interact to predict uptake of multilevel interventions
- To determine how individuals life-course transitions and exposure to DREAMS impacts on retention

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

Educational and nutritional status, General health, sexual behaviour and relationships, DREAMS awareness and uptake, MTV Shuga exposure, Mobility and other risky behaviours and HSV-2 test results

1.2.3 Unit of Analysis

Basic unit(s) of analysis or observation that the study describes: For PANGEA is each record a sequence from a specimen, or are there multiple records for a single specimen or study participant

Unit of analysis is participants

1.3 Scope

1.3.1 Topics Classification

MeSH subject headings

[HIV-1](#)

[Incidence](#)

[Phylogeny](#)

[Epidemics](#)

[Population Surveillance](#)

[Rural Population](#)

[HIV Infections](#)

[Africa](#)

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.

DREAMS, multilevel, young people

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.

Demographic surveillance area of the Africa Health Research Institute.

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.

Effectiveness of this multi-level approach to HIV prevention will depend on the reach, uptake and retention of each component by AGYW and their male partners. Our analytic sample will be all male partners (men aged 15-29), residing in the AHRI surveillance area. There are 14,600 age-eligible men; we expect 58% to be resident and participate in the annual surveillance.

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
Dr Maryam, Shahmanesh	Africa Health Research Institute

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
South African Medical Research Council	SAMRC	MRC-RFA-UFSP-01/2013/UKZN HIVEPI	Genotyping funding source

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
Wilkinson, Eduan	KwaZulu-Natal Research Innovation Sequencing Platform	Cleaned, aligned and help analyse the sequence data.

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

Participants were randomly selected from a southern part of AHRI surveillance area, stratified by age, sex and week-block. A target sample size was 2500 males aged 13-35 and 500 females aged 24-29 years.

1.7 Data Collection

1.7.1 Dates of Collection

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

Data collection started on 2018 and completed on 2019.

03-Sep-2018 until 11-Jun-2019

1.7.2 Notes on Data Collection/Production

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

Not applicable.

1.8 Data Processing

1.8.1 Other Processing

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

HSV-2 specimens were processed at AHRI Durban-based laboratory

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 a 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
Dr Maryam Shahmanesh	Africa Health Research institute, University College London	m.shahmanesh@ucl.ac.uk	www.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.

See section 1.2.1

3 Variable Description

A code book of the variables in the data file.

Name	Definition	Data Type and Codes
AC	Africa Centre (AHRI)	Community level location of sampling
Sequence	Polymerase sequence of viral isolate	String of As, Cs, Ts & Gs
Date	Date of sampling in decimal format	e.g. 2010.123