Data Documentation Template

1 Study Description

1.1 Identification

1.1.1 Title: Impact evaluation of PEPFAR's determined resilient empowered AIDS-free mentored and safe (DREAMS) initiative in UMkhanyakude District

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.

Data collected from a nested cohort of adolescent girls and young women (AGYW) aged 13-22 years collected between 2017 and 2018 within the Africa Health Research Institute's (AHRI) Southern Population Intervention Platform Study Area (SPIPSA), KwaZulu-Natal, South Africa.

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.

The main aim of this study is to evaluate the impact and effectiveness of the scale up of the DREAMS HIV prevention package of biological, behavioural and social interventions in reducing HIV incidence in adolescent girls and young women residing in the uMkhanyakude district of KwaZulu-Natal. To achieve this aim, the changes in different outcomes will be assessed over time. The primary outcome will be HIV incidence and other key secondary outcomes will include knowledge of own HIV status, sexual debut, HSV-2, number of sexual partners, age-disparity with sexual partners, ever been pregnant, condom use, unmet need for contraception, transactional sex, education (remaining in school) and experiences of violence.

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

Longitudinal survey data

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

Individual

1.3 Scope

1.3.1 Topics Classification

MeSH subject headings

HIV prevention

Interventions

Adolescents

Young people

Sexual health

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 nested cohort; sexual history; Adolescent Girls and Young Women (AGYW); GEMS scale; HIV testing; Partnership loop; Interventions; Coverage

1.3.3   Country

Indicates the country or countries covered in the file

South Africa

1.3.4   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; KwaZulu-Natal, uMkhanyakude district.

1.3.5   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.

<<Add here>>

1.4   Producers and Sponsors

1.4.1   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

1.4.2   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
Bill & Melinda Gates Foundation (BMGF)	BMGF	grant agreement number: OPP1136774	Impact Evaluation of PEPFAR funded DREAMS interventions

1.4.3   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
Jaco Dreyer and Nondumiso Mthiyane	AHRI	Data management, cleaning and analysis

1.5   Sampling

1.5.1   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

Adolescent girls and young women aged 13-22 years who were residents in the demographic surveillance area of the Africa Health Research Institute. A total of 3013 participants were randomly selected to obtain a target sample size of 800 after 2 years of follow-up, allowing for 40% non-contact/loss-to-follow-up. Sampling was stratified by age group and area (week-blocks).

1.6   Data Collection

1.6.1   Dates of Collection

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

1.6.2   Data collected started on 03 May 2017   Notes on Data Collection/Production

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

1.7   a team of trackers (10) collect data electronically using tablets and RedCap software. They use a sampling list generated by the statistician from the eligibility list to recruit eligible participants. Recruitment and enrolment of participants is done at the household. All sensitive information on sexual history is self-filled on the tablet by the participant. Interviews are conducted in isiZulu. Data Processing

1.7.1   Other Processing

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

1.8   Data Access

We will add these bits

1.9   Contacts

1.9.1   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	AHRI	m.shahmanesh@ucl.ac.uk	

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.

3   Variable Description

A code book of the variables in the data file.

Name	Definition	Data Type and Codes