

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.

Incidence of tuberculosis infection among adolescents in rural KwaZulu-Natal, South Africa, 2017-2018

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 datasets provided here will be used for analyses in estimating prevalence of *M. tuberculosis* infection among adolescents aged 10-19 years

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

Survey data

1.2.3 Unit of Analysis

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

Adolescents (aged 10-19) selected to participate in the study

1.3 Scope

1.3.1 Topics Classification

MeSH subject headings

Tuberculosis; *Mycobacterium tuberculosis* infection; adolescent; IGRA; risk factors

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.

Tuberculosis;

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.

uMkhanyakude district in northern 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.

Individuals aged 10-19 years who are resident members of households in the AHRI DSS area

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
Professor Alison Grant	LSHTM; AHRI
Kathy Baisley	LSHTM; AHRI
Themba Mzembe	LSHTM; AHRI
Dr. Anita Edwards	AHRI

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
Viiv Healthcare Positive Action Programme	Viiv	NA	HAYA Project

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
Dr. Kobus Herbst	AHRI	Chief Information Officer
Dickman Gareta	AHRI	Head of Research Data Management
Njabulo Dayi	AHRI	Data management, data cleaning

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

A randomly selected sample of 1100 adolescent resident in the AHRI DSS

1.7 Data Collection

1.7.1 Dates of Collection

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

November 2017-December 2018

1.7.2 Notes on Data Collection/Production

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

Participants were enrolled from their homes by a team of research nurses. The following clinical and demographic information was collected using a standard questionnaire: age, school attendance, hospitalization, BCG vaccination, previous or current household TB contact, history of HIV testing and attendance to indoor gathering locations. Participants were offered HIV a rapid HIV test and a blood sample was obtained for testing TB infection using QFT-plus assay at the project laboratory in Durban.

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.

NA

1.9 Data Access

NA

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
Themba Mzembe	LSHTM; AHRI	Themba.Mzembe@lshtm.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.

The dataset contains participants' survey responses and laboratory results for *M. tuberculosis* infection. The purpose of the data is to estimate the prevalence of and risk factors for *M. tuberculosis* infection among adolescents.

3 Variable Description

A code book of the variables in the data file.

Name	Definition	Data Type and Codes