

# Manicaland Project

Supporting information for the core dataset v2

**Survey Interviews: 1998 – 2013**  
(Round 1 – Round 6)



---

**Imperial College  
London**

Supported by  
**wellcome** trust



Prof. Simon Gregson  
School of Public Health  
Imperial College London  
St Mary's Campus

Tel: +44 (0)20 7594 3279  
Email: [s.gregson@imperial.ac.uk](mailto:s.gregson@imperial.ac.uk)

---

# Contents

Data Assess Statement.....	3
Confidentiality.....	3
License Agreement Terms .....	4
Foreword.....	5
Introduction .....	6
Project Overview.....	6
Data Collection .....	7
Eligibility Criteria for General Population Survey .....	8
Data Processing .....	9
Description of Coredata.....	10
Table 1: Overall variables .....	11
Table 2: Round Specific Variables .....	12

## **Data Assess Statement**

Data generated by the Manicaland Project can be obtained from the website: <http://www.manicalandhivproject.org/data.html>. Here we provide a core dataset which contains a sample of socio-demographic, sexual behaviour and HIV testing variables from all 6 rounds of the main survey. Additional data used in the production of recent academic publications is also available. If further data is required, a data request form must be completed (available to download from our website) and submitted to [s.gregson@imperial.ac.uk](mailto:s.gregson@imperial.ac.uk). If the proposal is approved, we will send a data sharing agreement which must be agreed upon before we release the requested data.

## **Confidentiality**

The Manicaland Project has made every effort to preserve the privacy of cases and persons whose characteristics are within this database. We have removed all individual identifying information which was on the original data files, such as names. The only geographic identification is the district (i.e. Mutasa, Makoni, Nyanga) and study site name in which the household is located. Household numbers (which are unique to each site) were used instead of any postal addresses.

## **License Agreement Terms**

Your use of this dataset implies the acceptance of the following conditions. If you do not agree with the following conditions, please do not use this dataset.

Any results obtained from using (part) of the data should give full credential information to the Manicaland Project (<http://www.manicalandhivproject.org/>) and cite the following reference (this reference also includes some supplementary material on the Manicaland Project for further reading):

**HIV decline associated with behaviour change in eastern Zimbabwe.**

Gregson S, Garnett GP, Nyamukapa CA, Hallett TB, Lewis JJ, Mason PR, Chandiwana SK, Anderson RM (2006). *Science* 311(5761): 664-6.

This supporting document should be cited as follows:

**Manicaland Project – Supporting information for the core dataset v2**

prepared by the Data Management team of the Manicaland Project (2016).

*School of Public Health, Imperial College London, London, UK.*

## **Foreword**

The release of the core dataset of the first four rounds of adult individual data as part of the Manicaland Project will allow researchers interested in HIV prevalence and incidence to conduct their own analysis of HIV status in the Manicaland region, Zimbabwe. The variables in this release of the survey were selected by the Data Management staff and the Project Director, with input from the Data Analysis team.

*Manicaland Project  
Data Management*

# Introduction

## Project Overview

The Manicaland HIV/STD Prevention Project is a major collaborative scientific research initiative that has been underway in rural areas of eastern Zimbabwe since the early 1990s. The project has one of the largest prospective population-based surveys in Africa. It was set up to investigate the prevalence of HIV in Manicaland province and the demographic and social effects of the epidemic in the region. Initially, a comparative study was undertaken comparing a site in Honde Valley, an area severely affected by the HIV epidemic, with a site in Rusitu Valley, a socio-demographically similar area in Manicaland thought to be less affected by the epidemic.

Following this, plans were made to expand this comparative study to conduct a community randomised controlled trial to evaluate the effects of community-based programmes. In order to ensure a large enough sample of study sites for the community-randomised trial, the project was expanded to include 12 study sites from across three districts in Manicaland. Due to the remote positioning of the Rusitu Valley site, it was not included in the trial.

The study was piloted in February 1998 and the full study was launched in July 1998 and was completed in February 2000. Around 10,000 adults were interviewed in the baseline survey. Respondents answered questions relating to demographics, sexual behaviour and HIV prevention. A sample of blood was collected onto filter papers from each respondent and tested for HIV. The second round of the survey collected similar data and was used to evaluate the success of the community-based intervention.

Since then, the cohort study has continued and grown in Manicaland. Our questionnaire now covers diverse topics ranging from marital and sexual relationships, health and access to treatment and knowledge and awareness of HIV/AIDS. There have also been a number of spin-off studies conducted as part of the project (see the Manicaland Project website for further details).



## Data Collection

The Manicaland HIV/STD Prevention Project is an ongoing population-based open cohort study. Twelve communities from three different districts representing four socio-economic population strata were selected for this study. From the least to the most developed, the communities represent subsistence-farming areas (St. Theresa's, St. Killian's, Honde Mission, and Bonda Mission), roadside trading settlements (Watsomba and Nyabadza), large-scale agricultural estates that produce tea, coffee, and timber (Selborne, Sheba, Katiyo, and Eastern Highlands estates), and commercial centres within the rural areas (Nyanga and Nyazura). All survey rounds are conducted in a phased manner (one site at a time). Sites 1, 5, 11 and 12 dropped from the survey after round 5. Round 6 of the survey therefore consisted of only 8 sites.

No.	Site Name	R1	R2	R3	R4	R5	R6
1	<i>Katiyo</i>	X	X	X	X	X	
2	<i>Eastern Highlands Estates</i>	X	X	X	X	X	X
3	<i>Bonda Mission</i>	X	X	X	X	X	X
4	<i>Honde Mission</i>	X	X	X	X	X	X
5	<i>Selborne</i>	X	X	X	X	X	X
6	<i>Sheba</i>	X	X	X	X	X	
7	<i>Nyazura</i>	X	X	X	X	X	X
8	<i>Nyanga</i>	X	X	X	X	X	X
9	<i>Watsomba</i>	X	X	X	X	X	X
10	<i>Nyabadza</i>	X	X	X	X	X	X
11	<i>St. Killian's</i>	X	X	X	X	X	
12	<i>St. Theresa's</i>	X	X	X	X	X	

All local residents in the study areas are enumerated in household censuses which are performed at each round and take place at the same time as the individual surveys. Individuals eligible for the study are identified using our inclusion criteria. The eligibility criteria have changed over time and more details by survey round can be found below.

Where eligible individuals were not available for interview at the first visit, appointments are arranged and up to two further visits are made. Written informed consent is obtained from each study participant. Ethical approval for the studies was granted by the Medical Research Council of Zimbabwe, the Applied and Qualitative Research Ethics Committee, Oxford University (round 1-2), Imperial College London (round 3 onward), Biomedical Research and Training Institute, and Institutional Review Board.

## **Eligibility Criteria for General Population Survey**

### **Round 1**

Men aged 17-54 years and women aged 15-44 years who had slept in an enumerated household in a study site for at least 4 nights in the last month and who had also been staying in the same area at the same time one year prior to the interview date. Within any given marital grouping (i.e. a husband and one or more wives), one member of the grouping was selected at random (without replacement) as eligible for interview in the general population cohort.

### **Round 2**

All men and women enumerated in a household in a study site and interviewed as a participant in the general population cohort in round 1 (regardless of whether their ages now exceeded the initial age-limit) and still resident in the same study site. Also, all men and women who were below the relevant age-limit (i.e. 17 and 15, respectively) at round 1 but who were aged above this age-limit ('previously under-age') and met the residence criteria at the time of the round 2 census were treated as eligible. In sites 1-4, new in-migrants and visitors were treated as ineligible for the study. In sites 5-12, new in-were treated as eligible and the residence criteria were dropped. Round 1 participants reported to have out-migrated to neighbouring areas or to the cities of Harare or Mutare were also treated as eligible and followed-up; these individuals were treated as living in 'site 13' in round 2.

### **Round 3**

The eligible age-ranges for both men and women were each extended to ages 15-54 years; except for previous cohort participants who continued to be eligible if they now exceeded the upper age-limits. All previously under-age household members and new in-migrants and visitors (in all sites) were considered eligible and the restriction to one interview per marital grouping (applied in rounds 1 and 2) was dropped. All out-migrants were excluded from eligibility for interview.

### **Rounds 4 and 5**

Individual eligibility criteria as in rounds 4 and 5 (but restricted to 2/3rds of households).

### **Round 6**

Individual eligibility criteria as in rounds 4 and 5 (but restricted to 2/3rds of households in 8/12 study sites).

## **Data Processing**

HIV serological testing was done on dried blood spot specimens using a highly sensitive and specific antibody dipstick assay for all individuals who signed the written informed consent. All positives and a 10% sample of negatives were confirmed at baseline. For follow-up studies, only the samples from those participants where sero-conversion was indicated were re-tested. New participants at each cohort who were found to be seropositive were also re-tested for confirmation.

Data from all rounds of each survey have been cleaned and combined into a longitudinal SQL relational database. The project's data analysis coordinator and database manager ensure that data collected in each survey round are entered, validated, checked for consistency and prepared for analysis within a maximum of one year of the completion of data collection. Data are not released until they meet acceptable standards for scientific analysis.

## Description of Coredata

The core dataset combines individual-level data over the first five survey rounds. Data from rounds six and any new rounds will be added to the publicly downloadable core dataset within 18 months of becoming available for analysis by project investigators, staff, and collaborators.

Each line of the core dataset corresponds to an individual who completed at least one individual survey at any round and for whom at least one lab-confirmed HIV test result exists. This core dataset consists of a total of 40,872 individuals out of whom 9819 participated in round 1, 8246 in round 2, 16566 in round 3, 12,707 in round 4, 14426 in round 5 and 8902 in round 6.

The core dataset further consists of selected demographic measures (e.g. age, sex, educational level), behavioural indicators (e.g. age at sexual debut, number of sexual partners in life and in the last year, condom use), HIV testing experience (e.g. years since last test, self-reported and lab-confirmed HIV test result) amongst others; see Table 1 for full list. The core dataset also contains summary measures across all rounds which capture date of participation, birth and death, as well as the date of first and last negative and positive result respectively, to assess for sero-conversion; see Table 2 for full list.

Table 1: Overall variables

<b>Variable Name</b>	<b>Description</b>	<b>Format</b>
<b>IDNR</b>	<i>Unique identifying number over all study rounds and sites</i>	
<b>sex</b>	<i>Gender of respondent</i>	1 Male 2 Female
<b>birthmonth</b>	<i>Month of birth</i>	M
<b>birthyear</b>	<i>Year of birth</i>	YYYY
<b>m_entry</b>	<i>Month at which first entered the survey</i>	M
<b>y_entry</b>	<i>Year at which first entered the survey</i>	YYYY
<b>round_entry</b>	<i>Round at which first entered the survey</i>	
<b>m_exit</b>	<i>Month of last survey</i>	M
<b>y_exit</b>	<i>Year of last survey</i>	YYYY
<b>round_exit</b>	<i>Round of last survey</i>	
<b>m_death</b>	<i>Month of death</i>	M
<b>y_death</b>	<i>Year of death</i>	YYYY
<b>m_fstneg</b>	<i>Month first tested negative in survey</i>	M
<b>y_fstneg</b>	<i>Year first tested negative in survey</i>	YYYY
<b>m_lstneg</b>	<i>Month last tested negative in survey</i>	M
<b>y_lstneg</b>	<i>Year last tested negative in survey</i>	YYYY
<b>m_fstpos</b>	<i>Month first tested positive in survey</i>	M
<b>y_fstpos</b>	<i>Year first tested positive in survey</i>	YYYY
<b>m_lstpos</b>	<i>Month last tested positive in survey</i>	M
<b>y_lstpos</b>	<i>Year last tested positive in survey</i>	YYYY

Table 2: Round Specific Variables

<b>Variable Name</b>	<b>Description</b>	<b>Format</b>
<b>rXdistrict</b>	<i>Study district</i>	1 Mutasa 2 Makoni 3 Nyanga 1 Katiyo 2 EHTE 3 Bonda Mission 4 Honde Mission 5 Selborne 6 Sheba 7 Nyazura 8 Nyanga 9 Watsomba 10 Nyabadza 11 St. Killian's 12 St. Theresa's
<b>rXsite</b>	<i>Study site ID</i>	
<b>rXhid</b>	<i>Household ID (Unique by rXsite)</i>	
<b>rXline</b>	<i>Line number (Unique by rXhid)</i>	
<b>rXm_round</b>	<i>Month of Interview</i>	M
<b>rXy_round</b>	<i>Year of interview</i>	YYYY
<b>rXhivrsIt</b>	<i>Lab-confirmed HIV (Borderline HIV test results are coded as missing)</i>	1 positive 2 Negative
<b>rXage</b>	<i>Age of respondent</i>	
<b>rXyrsvill</b>	<i>Years respondent lived in village</i>	995 Since birth 996 Visitor
<b>rXoutside</b>	<i>Years respondent lived outside the village</i>	996 Visitor 0 None 1 Primary 2 Secondary 3 Higher 1 Estates 2 Manufacturing or building 3 Police or army 4 Education 5 Services or retail 6 Informal 7 Student 8 Unemployed 9 Other
<b>rXedlev</b>	<i>Level of education</i>	
<b>rXwsect</b>	<i>Sector of employment</i>	1 Professional/managerial 2 Self-employed 3 Skilled labour 4 Manual/unskilled labour
<b>rXwlevl</b>	<i>Level of employment</i>	1 Good health 2 Recurring sick
<b>rXhealth</b>	<i>Health status over last few months</i>	

		3 Serious illness
<b>rXhivtest</b>	<i>Had HIV test in lifetime?</i>	1 Yes 2 No
<b>rXhivtestyrs</b>	<i>Years since last HIV test (missing if rXhivtest !=1)</i>	1 Positive 2 Negative 98 Don't Know 99 Won't say
<b>rXhivself</b>	<i>Self-reported HIV status (Rounds 3 – 6 only)</i>	1 Positive 2 Negative 98 Don't Know 99 Won't say
<b>rXtkndrugs</b>	<i>Self-reported ART usage for seropositive participants (Rounds 3 – 6 only)</i>	1 Yes 2 No
<b>rXagefsex</b>	<i>Age at first sex (missing if never had sex)</i>	
<b>rXptnrlife</b>	<i>Number of sexual partners in life (missing if never had sex)</i>	
<b>rXptnryear</b>	<i>Number of sexual partners in last year (missing if never had sex)</i>	
<b>rXcondom2w</b>	<i>Condom use in last 2 weeks</i>	0 Never 1 Occasionally 2 Always