

## Contextual documentation for the shared ALPHA data:

### Manicaland

28<sup>th</sup> March 2022

#### Quick links

1. What is the Manicaland study web address?  
<http://www.manicalandhivproject.org/>
2. What is the Manicaland study data catalogue link?  
Manicaland data catalogue is not available online
3. What is the reference/ citation to the Manicaland study cohort profile?  
(Gregson et al. 2017) <https://bmjopen.bmj.com/content/7/10/e015898>
4. Where are questionnaires available?  
Questionnaires: <http://www.manicalandhivproject.org/questionnaires.html>

## Study Background

### Study description

The Manicaland general population open cohort HIV sero-survey data are from the Manicaland province in Eastern Zimbabwe. The study is described in detail here <https://bmjopen.bmj.com/content/7/10/e015898>

### Study area

The first five rounds of the study covered people resident in 12 areas: two small towns, four agricultural estates, two roadside settlements and four subsistence farming areas. From round 6, eight of the study areas were followed-up.

### Study eligibility

Households, and their adult residents, within the study areas. See <https://bmjopen.bmj.com/content/7/10/e015898> for detailed eligibility criteria for each round.

From round 3 onwards, all eligible adults were interviewed. In rounds 1 and 2 only one person from each marital group was interviewed.

### Study process

The data collection staff begin with a list of households (enumerated at the last round) and work with village guides and village community health workers to ascertain whether any have been dissolved and whether there are any new households to be enumerated in this round.

Each household is visited for completion of a household census. For households which have previously been surveyed the data collection instruments contain information on the individuals included at the last round to enable follow up of individuals who may have died or moved away. Census data is obtained for each household. One person answers for entire household, either the household head, or senior member of the household, or their proxy depending on who is present during the researchers' visit.

After the census has been completed the information is used to identify eligible household members for a detailed survey questionnaire. The eligibility for individual interview has changed over the rounds. Full details are listed here - <http://www.manicalandhivproject.org/eligibility-criteria.html>.

The survey includes demographic information, sexual behaviour, relationships, fertility, pregnancy history, knowledge, awareness and attitudes towards HIV, wider health condition. This is conducted shortly after the household census. Following the individual interview, respondents are invited to take an HIV test for research purposes.

## Timings and scope of study rounds

There were six rounds of cohort study between 1998 and 2013. See Gregson 2017 Supplementary Table 1 for timings and response rates for study rounds

<https://bmjopen.bmj.com/content/bmjopen/7/10/e015898/DC1/embed/inline-supplementary-material-1.pdf?download=true>

HIV status was ascertained in each round.

The seventh and eighth rounds were a modified follow-up of the cohort and the data are not included here.

## Data collection and management

Until round 5, all data collection was using paper instruments with some questions in the sexual relationships section answered using a confidential self-completion method

([https://sti.bmj.com/content/80/suppl\\_2/ii36](https://sti.bmj.com/content/80/suppl_2/ii36)).

In round 6 the EpiCollect software was used.

Rounds 7 and 8 (not included here) used ODK.

## Residency and household definitions

Households were initially defined as a group of people who regularly eat together from the same cooking pot.

Residency is not used as a concept, people who are present in the household at the time of the household census are recorded and those who meet the eligibility criteria are invited for the individual interviews. In all rounds, the household census captured data on people who were considered to be usual residents or to stay in the household on a regular basis and who stayed at least one night in the last month (regular visitors). Regular visitors would include people who might work and stay away from home but come back for a few days at the end of the month and/or during holidays. From site 5 (out of 12) in round 2 as part of a collaboration with other partners on a mobility study, and in all sites in subsequent rounds, non-regular visitors were also included in the household census listings/questionnaires (i.e., by asking for any other people who stayed in the household on the night before the interview). So, in practice, a combination of both de jure and de facto.

## Migration

The study lacks a mechanism for internal migration reconciliation. Movements within study areas are captured as much as possible but it is reliant on the person being recorded with the same name in different rounds and locations. This means that internal migrants can be recorded as two different individuals. Consequently, internal migrations are underestimated.

More than a month visit – these captured when reported in the individual questionnaire

Work done within Manicaland study on relationships between migration and incidence of HIV is available in: [Mundandi 2006](#) and [Coffee 2005](#)

## Date imputation, including right censoring

**Right censoring** - how is this handled for those who continue to reside in the area at the time of last observation, those who migrated out, those who died, lost to follow up?

- In the household questionnaire, respondents were asked the date the person started staying in the household (if new since the last round); the date of death if they died; whether the person still stays in the household (or did at the date they died); and how many nights is it since they last slept in the household (from which one can estimate the date they left if they no longer stay in the household. All dates just month and year
- Assume that everyone who is still resident is censored at the date of the last household report or individual interview, whichever is later. For those who have out migrated or died, follow up stops at date of outmigration or death (which may be imputed)
- Dates of death should be captured in the household questionnaire (e.g. Q032 in R4; month/year). If a verbal autopsy interview was done, a more reliable date of death will be available (still month/year).

## Imputation of missing dates of migration to/from households

The date of arrival in, or departure from, the household was collected. If this was missing the midpoint between survey rounds was used. If dates of arrival/departure cannot be ascertained, follow up starts/ends on the dates of the first/last household interview in which the individual was present.

## Imputation of missing dates of birth (DoB)

If the day of birth is not known, 1 is used. If the month is not known, January is used. If the date of birth is missing entirely, the date reported in another round is used, if available. If no DoB is available it is back calculated from age at interview. If DoB or reported age differs between rounds, the most commonly reported one is used. N.B. this is more of an issue in early rounds as in the later rounds the data collection forms were pre-populated with DoB. In the case of a discrepancy in reported age or DoB between the individual questionnaire and the household census, then the individual questionnaire data is used.

## Data quality checks

### Household census: what is done with previously enumerated individuals for whom there is no data in current round

- When a household member who was seen in the previous round is not present in the current round, the household respondent is asked about their destination and for an estimated date when they left
- For household members who reappear in the household after missing round(s), the household respondent is asked for the date when they returned
- Individual questionnaire collects information on how long the person has been in the village but date of entry to household is used to define residency episodes.
- Any queries to do with establishing whether the individual has out-migrated or died are resolved in the field or during data management. If no further information is obtained the person's follow-up finishes at the date of the last enumeration.

### Inconsistencies in the static variables such as sex

- Round to round checking is done during data cleaning prior to data analysis
- Where inconsistencies are found, the most internally consistent and/or most frequently recorded sex etc. is used and captured in the core data set
- Self-reported information preferred over information from household questionnaire

### Inconsistencies in the reported order of key events

- ? What the reference point is – what has to happen before what?
- Data quality supervisor reviews questionnaires before they are entered to identify and resolve inconsistencies – data cleaning checks to be made. Where necessary, the field team can revisit a respondent to resolve identified errors.
- End of round data cleaning procedure – data office has a checklist for data validation and cleaning. This check list is stored on the Manicaland study servers and can be made available on request.

### Recording of deaths- are there any other sources of data on deaths?

- All death data is based on information reported during census/ survey round
- Checklist of households done in the previous round and individuals previously interviewed. Household proxy is asked what has happened to any individuals who are listed but no longer present.
- Field work supervisor meets with village guide and team and the village guide updates the fieldworkers about who has passed away, moved out etc. This is done before household interviews start in the villages in each site. They go through the household and individual checklists found in that village in the previous round systematically with the village guides and update the checklists accordingly.
- Verbal autopsy interview visit follows report of a death in the household. The date of death information updated based on VA interview. This is usually after the death has been confirmed by the remaining household members (if the household still exists in the area)

## Other data recorded outside census rounds

- There are no key informants for demographic data
- There are no clinic data linked to the cohort

## Cross-round linking and harmonisation

### Linking of individuals across rounds

- Pre-population of the household list
- Multiple identifiers to link across rounds
- within round identifiers and
- across round identifiers
- These inconsistencies are rare because of the robust identification across rounds, whenever they are encountered, they are handled at data management level in consultation with lab and field team

### Linking of HIV test results to individuals

- Fieldworkers send samples for testing results to the lab; lab using same identifiers as survey so no linking is required.
- Lab team checks dates of sample when they receive the samples from the field team. Any cleaning required happens at this point.
- In earlier rounds, tests were sometimes delayed so date of test was assigned as the date of sample collection.

### Data harmonisation

- Data for each round are stored separately and collated for analysis or when developing ALPHA data specifications. All data are stored in same database on the project's servers. Data are stored in data tables according to the survey activity and questionnaire section for each round. They follow a fairly similar structure across the rounds but there are differences due to changes in the questionnaire over the rounds and changing data management practices over the years.

## HIV testing protocol

A detailed description of the HIV testing protocols is given in <https://bmjopen.bmj.com/content/7/10/e015898>

### Are self-reported test results included

Not included in ALPHA data but were asked to individual survey respondents from round 3 onwards

- Though not included in the data submitted to ALPHA, work done in Manicaland on self-reported testing is available in [Cremin et al 2010](#)

## Reconciliation of discrepant HIV test results across rounds (retroconvertors/seroreversion)

Nothing done, data include results as obtained. Due to funding constraints, DBS samples from participants who have tested positive in previous rounds are not re-tested in later rounds – so this form of reconciliation is not possible

## Glossary

**Residency episode:** A period of time when an individual is continuously living in a household defined by the start and end dates of that period. Studies can vary in how they define households, and what constitutes a period of continuous residence. Variations usually reflect local differences in residency patterns, and allow capture of data on, for example, capture seasonal labour migrations where relevant.

**Internal migration:** A change of residence within the study area. These can be challenging to capture correctly in the data. The exit from the original household and entry into the new household is straightforward because they are reported as exiting/entering at the next data collection round. However it is challenging to identify the person in the new household as the same person who has left the old household unless enough information on the person has been collected in both households to unambiguously identify the individual. Dates of birth are not always well known or accurately reported by proxy respondents and names may change (for example upon marriage or the birth of a child) or be spelled in a variety of ways. National identity documents are not universal and where held, are not necessarily available to proxy respondents.

**Migration reconciliation:** The process of identifying the same individual across periods of residence in different households. This may be done by fieldworkers, at the data processing stage or using a combination of both approaches.

**Date imputation:** Dates may not be completely reported or accurately recorded (particularly when using paper forms). Construction of the datasets requires a date and so a missing or incomplete date must be replaced with a complete data. The rules used to impute the date depend on the event for which the date is being imputed and the study procedures.

**Right censoring:** Data are right-censored if the subject under study has not yet experienced the event which ends their period of follow up. In these open-cohorts, follow-up ends when the individual dies. Most of the cohort is alive throughout the study period and is resident at the most recent survey round. Their data are incomplete, and are described as right-censored.

**Residency gaps:** When an individual leaves the study area their follow up is paused (potentially indefinitely) and their residency episode ends at the date of out-migration. If the person rejoins the study they begin a new residency episode upon their return and there is a gap in their follow up time which corresponds to the period when they lived outside the study area.

**Seroconversion:** For HIV, seroconversion refers to infection with HIV and the subsequent production of detectable antibodies to HIV. Although this is a process which takes a number of weeks, we assign a single date for this which we use as a proxy for the point of acquisition of HIV.

**retroconverter:** In the uncommon setting of a longitudinal population-based study on HIV, participants are regularly and repeatedly tested for HIV following the prevailing diagnostic testing algorithms. In many survey rounds people living with HIV are retested, alongside HIV-negative participants. In some instances, tests on people who have previously tested positive, perhaps on multiple occasions, return a negative result. Testing of samples is blind to the previous results of the individuals. When this happens we refer to these individuals as retroconvertors.