Contextual documentation for the Shared ALPHA-MEIRU data

2021

MEIRU data FAQ

1. What is the MEIRU web address?
   - https://www.meiru.info/
2. What is the MEIRU data catalogue link?
   - MEIRU data catalogue is not available online
3. What is the reference/ citation to the Karonga HDSS cohort profile?
   - (Amelia C Crampin et al. 2012)

4. Where do the data for the residency file and HIV test file come from? What study/ studies? How frequent?

ALPHA Specification 6.1: Residency data

- Mainly comes from the HDSS/ Continuous Registration System (Amelia C Crampin et al. 2012)
- Annual census rounds are conducted in combination with monthly Key Informant (KI) meetings to report births and deaths. A KI is a community leader trained in reporting core events of migration, births and deaths happening within their community within the HDSS.
- Migration is recorded during the annual census rounds and whole house migrations are reported by the KI.

ALPHA Specification 6.2b: HIV testing data

- From the serosurveys – rapid test results from sero survey (Floyd et al. 2013)and from other studies – the MEIRU NCD survey (Amelia Catharine Crampin et al. 2016; Price et al. 2018), and Cooking and Pneumonia Study (Mortimer et al. 2017) data collected in the Karonga HDSS.
- Before the serosurvey – adult health survey and from any rapid test done in the various studies.

5. Preparation of the Residency file from source data:
   - How is a resident or residency defined?
     - The Karonga HDSS cohort profile (Amelia C Crampin et al. 2012) provides the definition of a household and a resident used in the HDSS.
     - A household is defined as a group of individuals who live together and recognise the same household head (Amelia C Crampin et al. 2012)
     - A household member is defined as a person accepted as belonging to the household.
     - Every household member is a resident
     - A visitor is a person expected to return to their home somewhere after the duration of their visit.
     - Unlike other HDSS, no time duration is attached to defining a resident predicated on being a household member or visitor, but rather what is reported by the interviewee. It is based on cultural perceptions of household membership.

- The distinction between residents and visitors is made in the field during data collection based on the above definitions also provided in the HDSS field protocols

- Is internal migration acknowledged? If it is, how is it handled?
  - Both internal and external migration are reported annually. This starts with field team meeting with KI who report about whole household migrations. This is then followed by individual household re-census to follow up on individual migrations

- Undocumented absences, Loss to follow up or hanging cases
  - when there are gaps in an individual's residency information for some observations – how is this handled?
  - Undocumented absences are not explicitly experienced in Karonga, they are not recorded in the database. The field team records reported movements, deaths, and births. The respondent has to report all departures and arrivals for the household. Household members with no reported event during a census round are assumed to be present.
  - Checks are done in the databases to assess the quality of the data. These include verifying that for internal migration each departure from one household has a corresponding arrival at another household. If for any departure a corresponding arrival cannot be established or vice versa, then the concerned migrants are referred ot as unresolved migrants. These are resolved between the data management team and the field team prior to the development of the ALPHA data specifications.

- Are the data separately stored for each data collection round?
  - No
- 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 and lost to follow up?
  - End of observation – those present in the study area, the date is given as 31 Dec in the last year we have complete data on
  - Exit type for those still in the study area at the end of the observation is ***present in the study*** in the data.
  - Losses to follow up – not identified in the Karonga data

- Missing dates? DoB how is it handled? Midpoint/ randomly assign, Masaka example
  - If month is missing June is used
  - If day is missing the 15th is used
  - If year of birth or year of other events of interest is not known, either the middle of the following bespoke bands defined according to various cultural/ historical landmark events among them national independence, coming back of Dr Kamuzu Banda a prominent nationalist who led Malawi after independence from overseas in 1958, the sinking of the Vipya ferry in lake Malawi in 1946, among others, or the middle of the decade where the year of interest lies are used:
    - <1900
    - 1900-13
    - 1914-33
    - 1934-45
    - 1946-57
    - 1958-63

>1964
1890-99
1900-09
1910-19
1920-29
1930-39
1940-49
1950-59
1960-69
1970-79
1980-89
1990-99

- Interview dates – same hhold, next-door neighbour etc
    - Not required in the specification. In Karonga, this is entered in during fieldwork before data get into databases used for preparing ALPHA specifications
- Inconsistencies in the static variables?
    - There is one database table which works as the authoritative source of information for variables such as sex, date of birth and others.
    - An algorithm is used to resolve these discrepancies this is resolved as part of routine data processing, not at the time of creating a specification
- Inconsistencies in the status variables? Hhold id
    - Not checked during specification preparation, assumption is that all inconsistencies in household id numbers are resolved during routine data management.

- Inconsistencies in the events order? What reference point is used to resolve these? What must happen before what?
    - Handled during data processing through standard data validation checks. Issues that cannot be resolved in the data processing office are brought to the attention of the field team and households are revisited to double check as needed
- Inconsistencies in the events dates?
    - Overlapping episodes
        - Dates changed in a consistent manner to remove overlaps and ensuring that each person gets the full number of person-years that belongs to them.
        - Polygynous heads – left in only one household per time – cover them for the entire period they are resident without overlaps
- Recording of deaths recorded outside census rounds, how are these brought in? those who die outside the study area with a last observation date in the census
    - KI monthly update of events – trigger Verbal Autopsy for deaths
    - During the annual census deaths maybe picked up and trigger a VA
    - If an individual dies outside of the HDSS area before a departure is recorded, the field team checks with the reporting household whether the deceased was still considered a household member at the time of death, if so, a VA is done, if no longer considered a member at the time of death, they are recorded to have departed at a point prior to their death as reported by household proxy
    - VA – no death form, full VA required before a death appears in the database

- Other data recorded outside census rounds
    - None

6. Preparation of the HIV testing data from the source data
    - The preparation of the specification includes steps to identify and exclude everyone not in the residency data
    - For everyone included in the residency data, all HIV tests ever recorded for those individuals are included in the specification

    - HIV testing protocol(s)
        - HIV testing protocols are described in publications from study (White et al. 2007; Glynn et al. 2001; Amelia C Crampin et al. 2003; McGrath et al. 2007; Molesworth et al. 2010; Tafatatha et al. 2015). Initially before rapid testing, the Edgeware and Organon was used. After the introduction of rapid testing, the study switched to UniGold HIV rapid test. Apart from this, quality control of rapid tests is done by testing every 1 in 10 negative sample and all positive samples using Edgeware and Organon.

    - Do you include reported testing and results?
        - No
    - What are the sources of your HIV test results?
        - These are provided in the study metadata
    - Survey round name – description to help understand these better
        - These are provided in the study metadata
    - Handling inconsistencies in the test results
        - Resolved at data management level, not handled during specification preparation. Further checks are done centrally at ALPHA secretariat
    - Handling gaps – one negative test then a huge gap then a positive or another negative test? Unusual testing pattern, how is this investigated and verified if it's the same person
        - ID system is robust so not checked during specification preparation
    - How are rounds of data collection linked and links between studies? Are inconsistencies test results
        - This is done during data processing in the data management system through a robust system of longitudinally linked IDs
    - Handling missing dates
        - Very rare in the data
    - Handling inconsistent dates
        - Not checked during specification preparation