








## DATA NOTE

# The Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA): Data on mortality, by HIV status and stage on the HIV care continuum, among the general population in seven longitudinal studies between 1989 and 2014 [version 1; referees: 2 approved, 1 approved with reservations]

Emma Slaymaker <sup>1</sup>, Estelle McLean<sup>2,3</sup>, Alison Wringe<sup>1</sup>, Clara Calvert<sup>1</sup>, Milly Marston<sup>1</sup>, Georges Reniers<sup>1,4</sup>, Chodziwadziwa Whiteson Kabudula<sup>5</sup>, Amelia Crampin<sup>2,3</sup>, Alison Price<sup>2,3</sup>, Denna Michael<sup>6</sup>, Mark Urassa<sup>6</sup>, Daniel Kwaro<sup>7</sup>, Maquins Sewe<sup>7</sup>, Jeffrey W. Eaton <sup>8</sup>, Rebecca Rhead<sup>8</sup>, Jessica Nakiyingi-Miir<sup>9</sup>, Tom Lutalo<sup>10</sup>, Dorean Nabukalu<sup>10</sup>, Kobus Herbst <sup>11</sup>, Victoria Hosegood <sup>11,12</sup>, Basia Zaba <sup>1</sup>

<sup>1</sup>Department of Population Health, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK

<sup>2</sup>Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK

<sup>3</sup>Malawi Epidemiology and Intervention Research Unit, Lilongwe, Malawi

<sup>4</sup>School of Public Health, University of the Witwatersrand, Johannesburg, 2000, South Africa

<sup>5</sup>MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

<sup>6</sup>National Institute for Medical Research, Mwanza, Tanzania

<sup>7</sup>Kenya Medical Research Institute, Kisumu, Kenya

<sup>8</sup>Department of Infectious Disease Epidemiology, Imperial College London, London, W2 1PG, UK

<sup>9</sup>MRC/Uganda Research Unit on AIDS, Entebbe, Uganda

<sup>10</sup>Rakai Health Sciences Program, Entebbe, Uganda

<sup>11</sup>Africa Health Research Institute, Durban, 4001, South Africa

<sup>12</sup>Department of Social Statistics & Demography, University of Southampton, Southampton, SO17 1BJ, UK

**v1** First published: 06 Nov 2017, 1:4 (doi: [10.12688/gatesopenres.12753.1](https://doi.org/10.12688/gatesopenres.12753.1))  
Latest published: 06 Nov 2017, 1:4 (doi: [10.12688/gatesopenres.12753.1](https://doi.org/10.12688/gatesopenres.12753.1))

## Abstract

Timely progression of people living with HIV (PLHIV) from the point of infection through the pathway from diagnosis to treatment is important in ensuring effective care and treatment of HIV and preventing HIV-related deaths and onwards transmission of infection. Reliable, population-based estimates of new infections are difficult to obtain for the generalised epidemics in sub-Saharan Africa. Mortality data indicate disease burden and, if disaggregated along the continuum from diagnosis to treatment, can also reflect the coverage and quality of different HIV services. Neither routine statistics nor observational clinical studies can estimate mortality prior to linkage to care nor following disengagement from care. For this, population-based data are required.

## Open Peer Review

Referee Status: 

	Invited Referees		
	1	2	3
<b>version 1</b>			
published 06 Nov 2017	<a href="#">report</a>	<a href="#">report</a>	<a href="#">report</a>

<sup>1</sup> **Mary-Ann Davies**, University of Cape Town, South Africa

The Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa brings together studies in Kenya, Malawi, South Africa, Tanzania, Uganda, and Zimbabwe. Eight studies have the necessary data to estimate mortality by HIV status, and seven can estimate mortality at different stages of the HIV care continuum. This data note describes a harmonised dataset containing anonymised individual-level information on survival by HIV status for adults aged 15 and above. Among PLHIV, the dataset provides information on survival during different periods: prior to diagnosis of infection; following diagnosis but before linkage to care; in pre-antiretroviral treatment (ART) care; in the first six months after ART initiation; among people continuously on ART for 6+ months; and among people who have ever interrupted ART.

2 **Matthias Egger**, University of Bern, Switzerland

3 **Tracy R. Glass**, Swiss Tropical and Public Health Institute, Switzerland

---

### Discuss this article

Comments (0)

**Corresponding author:** Emma Slaymaker ([emma.slaymaker@lshtm.ac.uk](mailto:emma.slaymaker@lshtm.ac.uk))

**Author roles:** **Slaymaker E:** Conceptualization, Data Curation, Formal Analysis, Visualization, Writing – Original Draft Preparation; **McLean E:** Conceptualization, Data Curation, Formal Analysis, Visualization, Writing – Review & Editing; **Wringe A:** Conceptualization, Writing – Original Draft Preparation; **Calvert C:** Conceptualization, Data Curation, Writing – Review & Editing; **Marston M:** Conceptualization, Data Curation, Formal Analysis, Validation, Writing – Review & Editing; **Reniers G:** Conceptualization, Funding Acquisition, Validation, Writing – Review & Editing; **Kabudula CW:** Data Curation, Validation, Writing – Review & Editing; **Crampin A:** Data Curation, Validation, Writing – Review & Editing; **Price A:** Data Curation, Validation, Writing – Review & Editing; **Michael D:** Data Curation, Validation, Writing – Review & Editing; **Urassa M:** Data Curation, Validation, Writing – Review & Editing; **Kwaro D:** Data Curation, Validation, Writing – Review & Editing; **Sewe M:** Data Curation, Validation, Writing – Review & Editing; **Eaton JW:** Conceptualization, Data Curation, Validation, Writing – Review & Editing; **Rhead R:** Data Curation, Validation, Writing – Review & Editing; **Nakiyingi-Miiro J:** Data Curation, Validation, Writing – Review & Editing; **Lutalo T:** Data Curation, Validation, Writing – Review & Editing; **Nabukalu D:** Data Curation, Validation, Writing – Review & Editing; **Herbst K:** Data Curation, Validation, Writing – Review & Editing; **Hosegood V:** Validation, Writing – Review & Editing; **Zaba B:** Conceptualization, Data Curation, Funding Acquisition, Investigation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**How to cite this article:** Slaymaker E, McLean E, Wringe A *et al.* **The Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA): Data on mortality, by HIV status and stage on the HIV care continuum, among the general population in seven longitudinal studies between 1989 and 2014 [version 1; referees: 2 approved, 1 approved with reservations]** Gates Open Research 2017, 1:4 (doi: [10.12688/gatesopenres.12753.1](https://doi.org/10.12688/gatesopenres.12753.1))

**Copyright:** © 2017 Slaymaker E *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** Bill and Melinda Gates Foundation [OPP1082114]; Wellcome Trust [085477].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**First published:** 06 Nov 2017, 1:4 (doi: [10.12688/gatesopenres.12753.1](https://doi.org/10.12688/gatesopenres.12753.1))

## Introduction

In 2014, an estimated 25.8 million people in sub-Saharan Africa were living with HIV (PLHIV), representing 70% of the global total, and 790,000 HIV-related deaths occurred<sup>1</sup>. A third of these deaths occurred in six countries: Kenya, Malawi, Tanzania, South Africa, Uganda and Zimbabwe<sup>2</sup>. The network for Analysing Longitudinal Population-based data on HIV/AIDS in Africa (ALPHA)<sup>3</sup> brings together health and demographic surveillance systems (HDSS) in these six sub-Saharan African countries, which have longitudinal data on participants and their HIV status.

These studies combine demographic, survey, clinical and verbal autopsy data to provide comprehensive, population-based estimates of the incidence of new HIV infections, the number of PLHIV, their experience of diagnosis and treatment, and their survival. For many PLHIV, these data permit estimation of seroconversion date and dating of diagnosis of infection, initiation of antiretroviral therapy (ART) and interruption of treatment.

Since the rollout of ART, overall adult mortality<sup>2,4–10</sup>, and that among PLHIV both on and off treatment, has fallen substantially<sup>11</sup>. However, the continuing high HIV mortality burden indicates that improvements in HIV testing, care and treatment are still needed.

For PLHIV, the continuum of care describes the steps necessary to achieve viral suppression<sup>2,12</sup>. Key transition points include diagnosis, assessment of ART eligibility, initiation of ART and viral suppression. Failure to progress in a timely manner from infection to diagnosis and from diagnosis to treatment will have a deleterious impact on mortality rates and onward HIV transmission.

Tracking PLHIV along each step of the HIV care continuum is challenging for several reasons<sup>2</sup>: Data on ART initiation and adherence are commonly available and readily collated from ART clinics with computerised records, but data on time of diagnosis and confirmation of eligibility for ART are less frequently recorded in formats that permit aggregation of those data for monitoring purposes. Furthermore, dates of seroconversion are not routinely estimated in the course of HIV testing or clinical care.

This data note describes a dataset containing harmonised survival data for DSS participants aged 15 years and above from eight studies in six countries over the period 1989–2014. The person-time of PLHIV is categorised by stage on the HIV care continuum: a) had undiagnosed infection, b) had learned their HIV status, but had yet to enrol in care, c) had received some pre-ART care, but had not (yet) started treatment, d) were receiving ART continuously, and e) had experienced an interruption in ART. Using these data, it is possible to estimate mortality at different stages of the HIV care continuum, and progression along the continuum.

## Methods

### Participating studies

Eight ALPHA Network studies could estimate mortality by HIV status over periods ranging from 4 to 26 years, and seven had data on HIV diagnosis and treatment: Karonga (Malawi)<sup>13</sup>; Kisesa

(Tanzania)<sup>14</sup>; Kisumu (Kenya)<sup>15</sup>; Manicaland (Zimbabwe)<sup>16</sup>; Masaka (Uganda)<sup>17</sup>; Rakai (Uganda)<sup>18</sup>; uMkhanyakude (South Africa)<sup>19</sup>. Agincourt in South Africa<sup>20</sup> had mortality data and partial HIV status data. The period with the most comparable data is 2005–14.

These studies are mainly ongoing rural HDSS<sup>21</sup> and collect demographic data on the entire population resident in specific geographical areas via frequent, focussed enquiries, and conduct regular home- or community-based HIV testing of the resident adult population. Additional socio-economic, HIV-related and sexual behaviour data are collected via questionnaire-based surveys. In uMkhanyakude there were no self-reported data on uptake of HIV care and treatment.

All resident adults are eligible for HIV testing, with the exception of Kisumu and Manicaland where HIV testing is restricted to a smaller, representative, cohort within the study. For the period in question, only Karonga and Kisumu returned HIV results automatically to participants, but all participants had the opportunity to find out their status by opting to receive the results of the research test or by undergoing HIV testing and counselling (HTC) in parallel with the research testing and in accordance with contemporary local standards for HTC, described elsewhere<sup>3</sup>.

The ALPHA network partners are independent research institutions, and began fieldwork at different times and with differing research priorities. Consequently, there is variation in study protocols and the data available. Recently, these studies have begun to link their population-based data with clinical data from local health facilities, but the information available from health facility records and the methods used to link these data to the HDSS vary. Four studies (Kisesa, Masaka, Rakai and uMkhanyakude) linked visit-level data from local ART clinics to the HDSS data. Administrative changes in clinics interrupted data linkage with the HDSS after 2012 in Rakai and 2013 in uMkhanyakude. Two studies (Karonka and Kisumu) had summary information from ART clinics linked to HDSS records. No clinical data were available for Manicaland, but some information on treatment was available from self-reported survey data. Detailed information is given in [Box 1](#).

It is worth noting that the period covered by these data includes the early phase of ART provision in these communities, and the decentralisation of treatment. Provision of treatment went ahead of strengthening clinic systems for monitoring and reporting on ART initiation, adherence and other relevant clinical measures. Many of the clinics that provide treatment to HDSS participants do not have electronic medical records (EMR) not least due to a lack of reliable electricity. During this period, most clinics could not easily undertake CD4 counts, and there was no viral load testing or resistance screening. Thus the standard of care for a typical clinic visit may compare unfavourably to that of high-income settings at the time. In addition, the lack of EMR means that some observations were not systematically recorded, for example, in many cases consultations were summarised in freehand notes. Such information is cumbersome and expensive to extract from handwritten logbooks and registers, and has thus not been included.

**Box 1. Details of clinic data available from each site and the methods used to link these data with the population-based data.****Karonga**

Links between the health and demographic surveillance system (HDSS) and clinical records were made by study fieldworkers in all 16 clinics that provide ART in the area. Those attending were interviewed and identified in the HDSS database. Initially, links were made only in ART clinics, so there is little information on pre-ART care. Clinical records were summarised every 6 months, and data to 2016 were shared with ALPHA.

**Kisesa**

Links between the HDSS and clinical records were made in two ways. Probabilistic links based on names, date of birth and place of residence were made for records from two HIV clinics from mid-2015 onwards. Individuals identified as HIV positive in the serosurvey or at routine HIV testing and counselling (HTC) in the HDSS study area were referred to local care providers, and their referral was tracked to enable ongoing linkage between clinical records and the HDSS.

**Kisumu**

A summary of patients' records was made annually since 2009 from the 19 clinics providing HIV care. Probabilistic and deterministic links were made using the Kenyan national identity number and other identifying information.

**Masaka**

Since 2012, all HIV positive participants have been routinely referred to the clinic run by the study which uses the same identifier as the research project. Pre-2012, those in selected clinic studies and those attending HTC were referred to the clinic and linked.

**Rakai**

Between 2004 and 2013, Rakai had good coverage of 19 HIV care clinics in the area covered by the HDSS and used probabilistic methods to link data with the HDSS. However, in 2013 the system for sharing clinic data with the HDSS changed and data are not available from that point onwards.

**uMkhanyakude**

Deterministic links between the HDSS and 21 HIV care and treatment clinics serving the HDSS area were made using South African national identity numbers and referrals tracking from the HDSS. Probabilistic record linkage was done whenever national ID numbers were not available.

The information needed to identify time spent in pre-ART care (still an important stage on the continuum during the period covered by these data) was not available for PLHIV in Manicaland (via self-report) or for most PLHIV in Karonga. Additionally, nobody in Manicaland is classed as having interrupted ART because the survey question on interruption of treatment did not permit dating of this event.

Local researchers supplied the dates at which ART first became available in at least one government health facility in each HDSS area, and the date when the ART rollout to all local health facilities designated as ART providers under national ART policy guidelines was complete.

**Transitions and stages on the continuum of care for HIV**

The primary purpose for constructing this dataset was to describe and compare mortality at different stages of the care continuum over time and between studies. Given the data available across the

different studies, we defined stages of the HIV care continuum based on: The estimated date of seroconversion from HIV negative to positive and the date of diagnosis of HIV infection, enrolment in HIV care, initiation of ART, stabilisation on ART and first interruption of ART. High quality data to describe temporary or permanent disengagement from pre-ART care or date of first ART eligibility were not available. The stages of the HIV care continuum are the intervals between these transitions. The care continuum, as described above, is a one-way system since people cannot return to a category after leaving it. Those who had a gap in treatment of longer than one month and those who had discontinued ART, at any time after initiation, were allocated to a group of people who had experienced an interruption in treatment, which included those who subsequently resumed ART.

Viral load was not routinely measured in the clinics serving these populations and so we could not identify people who achieved viral suppression. In lieu, we used a period of time on ART (more than six months without interruption), and used that classification as a proxy for stabilisation on treatment.

**Data sources and availability**

Data from each study were harmonised to standard specifications for analysis (Table 1, also see <http://alpha.lshtm.ac.uk/metadata/> for the detailed specifications) although not all datasets were available from every study.

**Analytical methods for preparation of the dataset**

Data from all HDSS participants, regardless of HIV status, were prepared for survival analysis based on residency data using Stata 15<sup>22</sup>. Entry into the study was on the 15<sup>th</sup> birthday or the date of first in-migration to the HDSS area (if already aged over 15), or the date of enumeration in the baseline study round. Exit was the date of out-migration or death. Gaps in residency due to temporary migration were recorded as such in the analysis. Surviving residents were censored at the time of their last HDSS observation. The dataset is organised to have multiple records for each individual, with each record corresponding to a short period of exposure which permits estimation of mortality by age, calendar time and HIV, diagnosis and treatment status. In Stata<sup>22</sup> terminology, this means the data have been stset and stsplit (see page 377 of the Survival Analysis Reference Manual<sup>23</sup>).

**Assignment of HIV status**

Person-time was allocated to one of three HIV status categories (unknown, negative, positive). Individuals could move through all three categories during study follow-up. HIV status information was available from a) periodic research HIV tests, b) linked clinic data, and c) self-report. Prior to first HIV test, individual HIV status was classed as unknown. The time between two negative tests was classed as negative and for the purpose of calculating mortality among HIV-negative adults, the period after the last HIV negative test was classed as negative for a short period, after which person-time was classed as unknown. For each study, the period of time classed as negative after the last negative test was based on the observed local incidence rates and the cut-off point was defined as the time by which 5% of people of the same age and sex had seroconverted. This ranged from 2 to 50 years in the East

**Table 1. ALPHA standard datasets available from each site.**

ALPHA standard datasets:		Karonga	Kisesa	Kisumu	Manicaland	Masaka	Rakai	uMkhanyakude
1	Information on residence in the study area, including dates of birth, migration and death.	•	•	•	•	•	•	•
2	History of HIV testing, including dates of tests, circumstance in which test was carried out, final test result, and whether or not the test result was returned to the participant.	•	•	•	•	•	•	•
3	Verbal autopsy data.	•	•	•	•	•	•	•
4	Self-reported information, from periodic surveys, on use of HIV testing services, disclosure of HIV status, use of HIV care and treatment services, ART use and interruption of ART.	•	•	•	•	•	•	•
5	HIV clinic records on enrolment in care and ART history.	•	•	•		•	•	•

African studies (mean 10.8) and from 1 to 50 years in the Southern African studies (mean 2.8). Individuals were assigned as HIV positive immediately after the first positive research test. People of hitherto unknown HIV status who were identified in HIV or ART clinics were classified as positive from the date of the first visit to the clinic. In Kisumu, participants could report their HIV status in lieu of a test and those results were used for people without a test result (less than 3% of participants).

People who were observed to seroconvert were treated exactly the same way as all other people who had tested negative and were categorised as positive only after the first positive test. Those for whom the interval between the last negative and first positive tests exceeded the cut-off were classified as unknown for a period prior to the first positive test.

#### Assignment of person time and deaths to stages on the HIV care continuum

Information on HIV diagnosis status, use of HIV-related care and ART history came from:

- receipt of HIV test result following periodic HDSS research tests
- self-report of HIV test history, diagnosis, pre-ART care and ART use, including dates
- linked clinic data, primarily on pre-ART care and ART

The information used from each source is detailed in [Table 2](#). A greater amount of information on experience of HTC, and thus diagnosis date, was available from self-reported and study data than from clinic data, whereas ART histories were derived almost entirely from clinic data.

Some individuals had information on a transition from more than one source, but no source of data is clearly superior. Clinic data did not contain information on diagnosis and did not include people who used clinics outside the study area. Self-reported data may quickly become outdated, but can include interactions with a wide range of service providers if PLHIV are willing to disclose their status to the interviewer. We considered using algorithms to give more weight to data from either clinic data or self-reported data for the transitions that could be identified in both data sources, but concluded that it was not possible to systematically deem one source to be more reliable than the other in all circumstances. Therefore, we took the earliest reported instance of any of the transitions as the date for that change regardless of the source of that data. We discarded self-reported dates that were implausible (such as, ART initiation before birth), but imposed no other constraints. Thus if, for example, a self-reported start date for ART was before the first positive test in the study, we backdated the HIV positive person time to the reported date of ART initiation.

Some participants retrospectively reported the dates when they made a transition (first positive diagnosis, linked to care, started ART, interrupted ART). For this analysis of mortality we did not use the retrospectively reported date but instead used the date of first report. This was to avoid biasing the data towards survivors, who were able to make a retrospective report of their progression along the continuum which those who had died before the survey could not do.

HIV positive person-time was allocated to one of the continuum stages listed in [Table 2](#) based on a synthesis of all the information available for each person and each transition. If at least one source indicated that an individual had made a transition, then

**Table 2. Categories on the HIV care and treatment continuum.**

Person years assigned as:	Description	Information used		
		Study research tests	Self-report	Clinic data
Undiagnosed	Never been told has HIV	Never given test results by study since first testing positive	Never received a test result* Not received a test result since first recorded positive test	Not used, cannot confirm undiagnosed status
Diagnosed	Has received $\geq 1$ positive test result	Has been given a positive test result by study	Reports having received a test result after date of first recorded positive test	Not used
In care	Has attended ( $\geq 1$ occasion) an HIV service or care clinic but has never had ART	-	Reports having attended HIV care services at least once. No restriction on type of service or frequency of visit. (ART naïve)	Attended an HIV care clinic on at least one occasion (ART naïve)
Early ART	Started ART within the past 6 months, no breaks	-	First 6 months after reported start date for ART and no gaps reported/ before first gap	First six months of continuous attendance at ART clinic
Stable ART	Been on ART longer than 6 months without a break	-	More than 6 months after reported start date for ART and no gaps reported (includes time before first gap)	After six months of continuous attendance at ART clinic <i>or, if initiation not captured, continuous attendance since first visit in someone who initiated ART elsewhere or before the clinic's data were linked with the HDSS</i>
Interrupted ART	Started ART but has experienced one or more gaps of 1 month or more in treatment	-	After first reported interruption in ART	
Unknown	Insufficient information to allocate to any of the above groups	-		

\* Distinction between taking an HIV test and receiving the results is not possible in some rounds in Masaka.

their person-time following that transition was allocated to the subsequent stage on the continuum. This process was repeated if information was available on subsequent transitions. After the last observed transition, all the remaining person-time was allocated to the last recorded stage. Transitions that occurred when the individual was not under observation in the study (due to temporary out migration) were recorded as occurring at the time that the individual returned to observation to avoid biasing the data towards temporary, rather than permanent, outmigrants and to survivors.

In addition to information on the dates when individuals made the various transitions, we also had dates at which a particular transition had not already been made from null reports in self-reported survey data (i.e. does not know HIV status, has never been to HIV care, has never taken ART, has never stopped taking ART) and from clinic history (has never started ART, has never stopped taking ART). Therefore person-time allocated to each individual's final stage can be disaggregated into time when the individual is known not to have made the transition to the next stage and the person-time contributed by people for whom there is no evidence regarding the next transition. For mortality analyses,

we did not separate the person-time in the final stage into these two groups, or use this information to censor the follow-up time, because there is a zero probability of mortality in the time before the final null report.

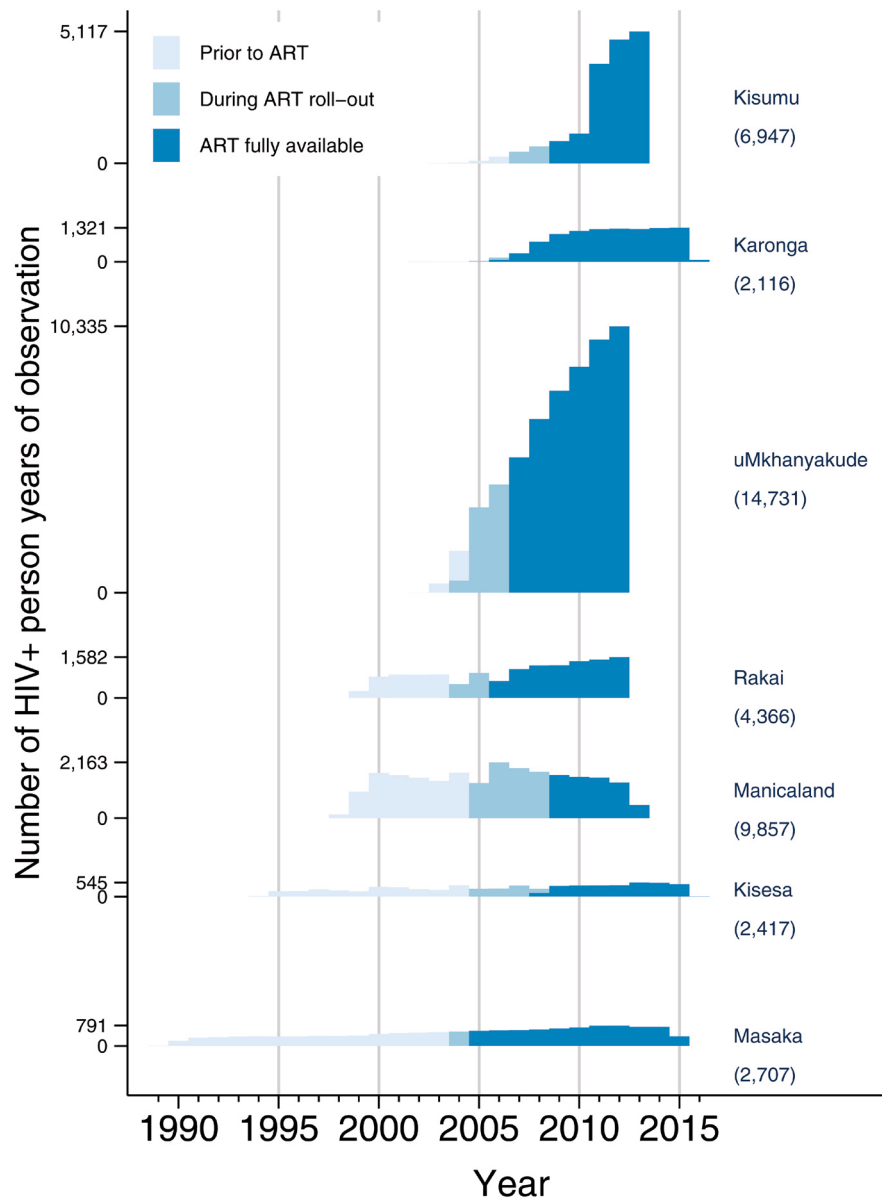
#### Continuity of care and adherence to ART

We identified anyone who ever self-reported or had been observed to have a break in treatment and distinguished these individuals from those who never interrupted ART. We were not able to make this same distinction for pre-ART care due to a lack of data for PLHIV in this stage. Based on clinic data, continuous attendance was defined as PLHIV attending appointments not more than one month after: a) next recorded scheduled visit, b) time when their pills ran out (based on records of pills dispensed at previous visit), or c) the interval between visits specified in national guidelines. This interval ranged from 1 to 3 months. PLHIV for whom we had only self-reported data regarding ART were assumed to be continuously on ART until they reported stopping or interrupting ART. When this was reported, or when a gap was identified in clinic data, PLHIV were moved to the last stage: interrupted ART.

**Relative size and annual coverage of the different studies**  
 For each study site, the total number of individuals included in this dataset is determined by the size of the population participating in demographic surveillance and the length of time the study has been in existence. The number of PLHIV included is further determined by the HIV prevalence and participation in survey rounds that included HIV testing. The proportion of participants who were HIV positive at the most recent survey ranged from 6% in Karonga to 33% in uMkhanyakude.

The average number of PLHIV under observation in each HDSS, for the periods i) before the introduction of ART; ii) during ART roll-out; and iii) once ART was widely available, are shown in Figure 1.

In terms of the number of individuals followed, the largest is uMkhanyakude in South Africa (approximately 63,000 adults in early 2012), followed by Kisumu in Kenya (circa 47,000 adults in early 2012), which began population-wide HIV testing in



**Figure 1. HIV positive person time by site and calendar year.** Numbers in brackets give total number of PLHIV ever identified.

2005 and 2008, respectively. We used only part of the Kisumu data, that from Gem district, which has more HIV data than the other two districts. The studies with the longest follow up are Masaka (beginning 1989) in Uganda and Kisesa in Tanzania (beginning 1994), which had, respectively, approximately 10,000 and 16,000 adults under observation at the start of 2012.

The introduction of ART occurred at different times in the different sites and the interval between introduction and widespread availability varied markedly ranging from one year in Masaka to 3-5 years in Kisesa and Manicaland in Zimbabwe.

### Description of variables

The variables and coding, where relevant, are given in [Supplementary File 1](#).

### Dataset validation

#### Availability of linked clinic data

The ability to link clinical records to HDSS data varies between sites and has changed over time. [Table 3](#) shows, by study, the number of PLHIV observed during 2005, 2010 and 2014 and the percentage for whom clinic data were available. The proportion of people with linked clinic data available represents the extent to which PLHIV are diagnosed and enrolled in care, and also the ability of the study site to link DSS and clinical records. In Kisesa, the total number of PLHIV fluctuates markedly because new HIV positives are identified once every 2–3 years, but exits due to death and out-migration are recorded approximately every 6 months in the HDSS.

#### Data availability over time, by HDSS and transition along the continuum

There were important differences between the studies in the availability of data. The time covered by each data source is shown in [Figure 2](#) and is disaggregated for each transition. The absolute numbers of transitions observed within each study varies substantially and so, to facilitate cross-study comparison, the height of each coloured band indicates, within a study, the relative number of transitions in each year, for each source, expressed as a proportion of the largest group in the study (at any time point).

### Impact of intermittent reporting on data quality

The validity of our results is determined by the timeliness of our data, and the degree to which we base our classification of person-years on outdated information. ALPHA studies collect survey data every one to three years, depending on study site, and the time elapsed between individual reports may be longer because a) individuals do not always participate in every round, and b) questions may have been omitted from some rounds.

Person-time after an observed transition is, by default, allocated to the next stage on the continuum. For example, after a positive test where there is no record of results being reported to the individual, person-time is allocated to the undiagnosed stage until the date of diagnosis/linkage to care or ART, if this occurs and is recorded. We may have one (or more) reports that the person has not yet learned their HIV status and that allows us greater certainty about the allocation of time to the undiagnosed stage. Following such a report, the chance of misclassification (the person learning their status but we fail to capture that information) increases with time since the last report.

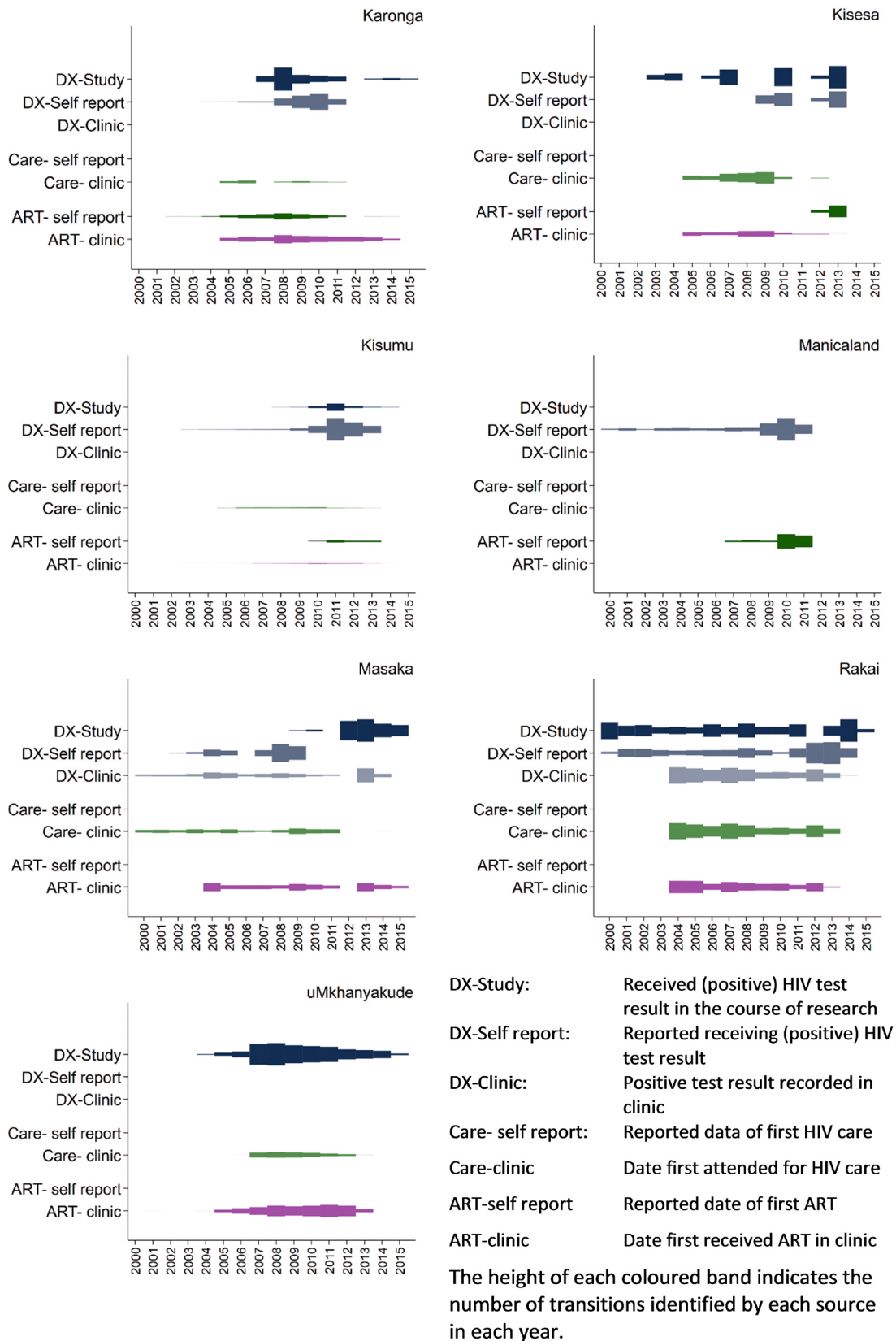
Examining the distribution of these null reports (a report stating that a transition has not been made) allows us to see how certain we are regarding allocation of person time to each category. [Figure 3](#) shows the distribution of person-time amongst PLHIV that was allocated to stages on the HIV care continuum based on not having made certain transitions: not diagnosed, not linked to care and not on ART (ART naïve). The total person-time in these states is summed and the figure shows the proportions of that time that were a) before a null report regarding the subsequent transition, b) less than two years after the last null report regarding the subsequent transition, and c) more than two years after the last null report. The fourth category shows person-time contributed by individuals for whom we have never had any information regarding the subsequent transition.

Pre-diagnosis has the greatest share of person-time prior to the last report, the lowest proportion where there is no evidence either way, and a high proportion of person time within two years of the last report on diagnosis status. This is because in most studies this

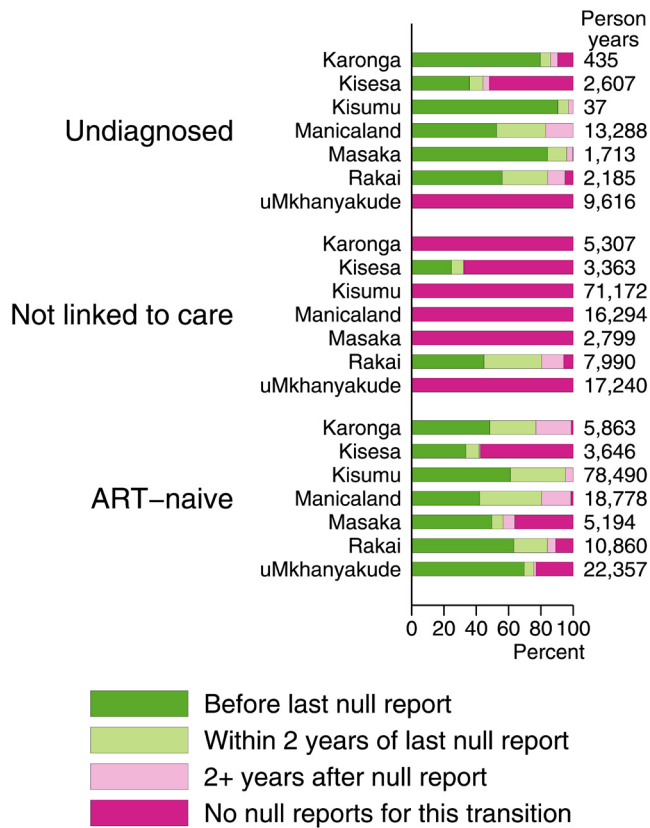
**Table 3. Number of people living with HIV (PLHIV) and the percentage linked to clinical records by study and year.**

Study	2005		2010		2014	
	Number PLHIV	% linked to clinic	Number PLHIV	% linked to clinic	Number PLHIV	% linked to clinic
Karonga	124	67.7	1296	54.6	1339	70.1
Kisesa	395	2.5	584	13.9	570	8.4
Manicaland	2571	0.0	2222	0.0		
Masaka	622	54.5	772	60.1	769	61.0
Rakai	1149	33.6	1609	47.9	1135	44.7
uMkhanyakude	4034	5.5	9705	46.5	11584	48.4
Kisumu	183	100.0	1938	65.3		





**Figure 2.** Data available from each study by data source and continuum stage, over time.



**Figure 3.** HIV positive person years prior to each transition along the continuum by the existence and recency of the information used to place them.

information is collected systematically at the same time as the HIV testing is conducted for research purposes and, in all studies except uMkhanyakude, participants are asked about testing history.

The transition to pre-ART care is the most poorly described, in part because pre-ART care was not as widely or systematically available as treatment. Five studies had no null information because this transition is only recorded by the arrival of the individual in an HIV care clinic. In Kisesa and Rakai this information is collected in self-report, which is why there is some person time known to be pre-transition.

The transition to ART is well described, with the exception of Kisesa, but despite that in most studies a substantial minority of person-years are classified as ART-naive by default because there is no evidence either way. Note that many of those individuals do have information on diagnosis.

### Ethical policies

Best practice regarding HIV testing for research purposes has changed during the lifetime of these studies. The earliest data collection made use of informed consent without disclosure. After treatment became available, research testing was accompanied by

opportunities for participants to learn their status using a variety of locally approved protocols.

Local ethical approval for these analyses was obtained by each study. The harmonisation and analysis of the pooled mortality data was approved by the ethics committee of the London School of Hygiene and Tropical Medicine (ref 6467) on 23<sup>rd</sup> August 2013.

### Consent

Written informed consent for analysis of the participants' anonymised details and publication of the results was obtained from the participants or, for verbal autopsy data, from a relative of the deceased participant.

### Data availability

The ALPHA Network harmonises the data produced by a diverse group of longitudinal HIV studies. Ownership of the data rests with the studies, or in some cases with the national government. The member studies have shared their data with the Network, but have not transferred ownership. The dataset is maintained by Emma Slaymaker, but access to the data is conditional on obtaining permission to use each study's data. Requests for permission to use the data should be directed to the ALPHA Network contact person in each study:

Study	Contact	Point of contact
Karonga	Amelia (Mia) Crampin	<a href="http://meiru.lshtm.ac.uk/">http://meiru.lshtm.ac.uk/</a>
Kisesa	Mark Urassa	via ALPHA
Kisumu	Daniel Kwaro	<a href="mailto:dkwaro@kemricdc.org">dkwaro@kemricdc.org</a>
Manicaland	Simon Gregson	<a href="http://www.manicalandhivproject.org/data-access.html">http://www.manicalandhivproject.org/data-access.html</a>
Masaka		via ALPHA
Rakai	Tom Lutalo	<a href="mailto:tlutalo@rhsp.org">tlutalo@rhsp.org</a>
uMkhayakude	via the Africa Health Research Institute data repository	<a href="mailto:help@afRICACENTRE.ac.za">help@afRICACENTRE.ac.za</a>

The ALPHA Network can facilitate such enquiries (<http://alpha.lshtm.ac.uk/>). Requests should state the name of the dataset: ALPHA\_Gates\_ready\_2015\_pooled\_2\_Aug\_2017.dta.

### Competing interests

No competing interests were disclosed.

### Grant information

Bill and Melinda Gates Foundation [OPP1082114]; Wellcome Trust [085477].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Supplementary material

Supplementary File 1: Description of variables.

[Click here to access the data dictionary.](#)

## References

- UNAIDS: **How AIDS changed everything — MDG6: 15 years, 15 lessons of hope from the AIDS response.** 2015; (accessed 1st March 2016). [Reference Source](#)
- UNAIDS: **The Gap Report.** 2014; (accessed 1st March 2016). [Reference Source](#)
- Reniers G, Wamukoya M, Urassa M, *et al.*: **Data Resource Profile: Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA Network).** *Int J Epidemiol.* 2016; 45(1): 83–93. [PubMed Abstract](#) | [Publisher Full Text](#)
- Floyd S, Marston M, Baisley K, *et al.*: **The effect of antiretroviral therapy provision on all-cause, AIDS and non-AIDS mortality at the population level – a comparative analysis of data from four settings in Southern and East Africa.** *Trop Med Int Health.* 2012; 17(8): e84–e93. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gargano JW, Laserson K, Muttai H, *et al.*: **The adult population impact of HIV care and antiretroviral therapy in a resource poor setting, 2003–2008.** *AIDS.* 2012; 26(12): 1545–54. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kasamba I, Baisley K, Mayanja BN, *et al.*: **The impact of antiretroviral treatment on mortality trends of HIV-positive adults in rural Uganda: a longitudinal population-based study, 1999–2009.** *Trop Med Int Health.* 2012; 17(8): e66–e73. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marston M, Michael D, Wringe A, *et al.*: **The impact of antiretroviral therapy on adult mortality in rural Tanzania.** *Trop Med Int Health.* 2012; 17(8): e58–e65. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Reniers G, Blom S, Calvert C, *et al.*: **Trends in the burden of HIV mortality after roll-out of antiretroviral therapy in KwaZulu-Natal, South Africa: an observational community cohort study.** *Lancet HIV.* 2017; 4(3): e113–e21. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Price AJ, Glynn J, Chihana M, *et al.*: **Sustained 10-year gain in adult life expectancy following antiretroviral therapy roll-out in rural Malawi: July 2005 to June 2014.** *Int J Epidemiol.* 2017; 46(2): 479–91. [PubMed Abstract](#) | [Publisher Full Text](#)
- Asiki G, Reniers G, Newton R, *et al.*: **Adult life expectancy trends in the era of antiretroviral treatment in rural Uganda (1991–2012).** *AIDS.* 2016; 30(3): 487–93. [PubMed Abstract](#)
- Reniers G, Slaymaker E, Nakiyingi-Miiró J, *et al.*: **Mortality trends in the era of antiretroviral therapy: evidence from the Network for Analysing Longitudinal Population based HIV/AIDS data on Africa (ALPHA).** *AIDS.* 2014; 28 Suppl4: S533–42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kranzer K, Govindasamy D, Ford N, *et al.*: **Quantifying and addressing losses along the continuum of care for people living with HIV infection in sub-Saharan Africa: a systematic review.** *J Int AIDS Soc.* 2012; 15(2): 17383. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Crampin AC, Dube A, Mboma S, *et al.*: **Profile: the Karonga Health and Demographic Surveillance System.** *Int J Epidemiol.* 2012; 41(3): 676–85. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kishamawe C, Isingo R, Mtenga B, *et al.*: **Health & Demographic Surveillance System Profile: The Magu Health and Demographic Surveillance System (Magu HDSS).** *Int J Epidemiol.* 2015; 44(6): 1851–61. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Odhiambo FO, Laserson KF, Sewe M, *et al.*: **Profile: the KEMRI/CDC Health and Demographic Surveillance System—Western Kenya.** *Int J Epidemiol.* 2012; 41(4): 977–87. [PubMed Abstract](#) | [Publisher Full Text](#)
- Sherr L, Lopman B, Kakowa M, *et al.*: **Voluntary counselling and testing: uptake, impact on sexual behaviour, and HIV incidence in a rural Zimbabwean cohort.** *AIDS.* 2007; 21(7): 851–60. [PubMed Abstract](#) | [Publisher Full Text](#)
- Mulder DW, Nunn AJ, Wagner HU, *et al.*: **HIV-1 incidence and HIV-1-associated mortality in a rural Ugandan population cohort.** *AIDS.* 1994; 8(1): 87–92. [PubMed Abstract](#) | [Publisher Full Text](#)
- Sewankambo NK, Wawer MJ, Gray RH, *et al.*: **Demographic impact of HIV infection in rural Rakai District, Uganda: results of a population-based cohort study.** *AIDS.* 1994; 8(12): 1707–13. [PubMed Abstract](#) | [Publisher Full Text](#)
- Tanser F, Hosegood V, Bärnighausen T, *et al.*: **Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey.** *Int J Epidemiol.* 2008; 37(5): 956–62. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kahn K, Collinson MA, Gómez-Olivé FX, *et al.*: **Profile: Agincourt health and socio-demographic surveillance system.** *Int J Epidemiol.* 2012; 41(4): 988–1001. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ye Y, Wamukoya M, Ezeh A, *et al.*: **Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Saharan Africa?** *BMC Public Health.* 2012; 12(1): 741. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Stata Statistical Software: **Release 15.** [program]. College Station, TX: StataCorp LLC, 2017. [Reference Source](#)
- StataCorp: **Stata 15 Survival Analysis Reference Manual.** College Station, TX: Stata Press, 2017. [Reference Source](#)

## Open Peer Review

Current Referee Status:



### Version 1

Referee Report 07 December 2017

doi:[10.21956/gatesopenres.13813.r26102](https://doi.org/10.21956/gatesopenres.13813.r26102)



#### Tracy R. Glass

Department of Clinical Research Unit, Swiss Tropical and Public Health Institute, Basel, Switzerland

This paper describes the valiant efforts of many people to collect quality data for research purposes. From experience in just one of these countries' HDSS data collection, I know the challenges they have faced. I think they did a good job attempting to give an accurate picture of the data they collected and its limitations both for use and for drawing valid and generalizable conclusions.

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 04 December 2017

doi:[10.21956/gatesopenres.13813.r26075](https://doi.org/10.21956/gatesopenres.13813.r26075)



#### Matthias Egger

Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

This is the first time I review a Data Note for Gates Open research. According to the instructions for authors, "Data Notes are brief descriptions of research datasets that include details of why and how the data were created; they do not include any analyses or conclusions". In general I felt the data note on the network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA) is well written and

I have approved it. The report could be a bit more detailed and clear in some places, and I have made suggestions for improvements below. These can all be dealt with within a minor revision.

1. The longitudinal studies participating in ALPHA were set up for different reasons and with different research priorities. It would be helpful to detail these reasons and priorities in greater detail. What were the primary research objectives of the different studies? What is the primary research objectives of ALPHA?
2. Clearly, the studies were not established to study the HIV cascade of care and data were obtained from different sources, including documented test results from HDSS surveys, linked clinic data, and self-reported data. I was surprised to read that “it was not possible to systematically deem one source to be more reliable than the other ...”. Isn’t a documented HIV test more reliable than a self-reported HIV status? Could this be clarified?
3. Following from above, if data from different sources were available, could the validity of the different sources be assessed, for example by considering documented test results as the gold standard. This might allow corrections, for example for social desirability bias. Please comment.
4. It would be useful to know how many participants had a negative HIV test before their positive test? Also, how many individuals were considered HIV positive because they attended HIV clinics, without any testing data? Can this be gleaned from Figure 2? Not entirely clear to me.
5. A time window of 1 month seems rather short to define interruption of treatment, and prone to misclassification due to data errors. Any single undocumented visit or pharmacy claim is likely to lead to misclassification of patients as interrupted treatment. Please comment on the choice of this window – usually longer windows are chosen (3 months or 6 months).
6. How exactly were the HDSSs and clinic data linked? I would have thought a data note would exactly be the place where some technical details should be provided on the type of probabilistic or deterministic linkage used.
7. The section on the Assignment of HIV status is not entirely clear. First the authors say that the period after the last HIV negative test was classed as negative for a short period (how short?) but then go on to state that the period of time classed as negative after the last negative test was based on the observed local incidence rates, that the cut-off point was defined as the time by which 5% of people had seroconverted (how was this threshold chosen?) and that this ranged from 2 to 50 years in the East African studies. Please clarify.
8. It would be interesting to see a discussion of how a cascade of care can be constructed using these data, including the definition of numerators and denominators at each step of the cascade, and what particular issues arise in the different studies.
9. Important limitations of the dataset include the lack of CD4 cell and viral load data, and the time period covered, which for some studies ends in 2012. This could be discussed in greater detail, particularly in the light of the recent changes in WHO guidelines (“treat all”).
10. Figure 2 needs more explanation and a scale for the horizontal bars that vary in width over time.

11. No datasets are provided and I could therefore not assess the question: “Are the datasets clearly presented in a useable and accessible format?”

**Is the rationale for creating the dataset(s) clearly described?**

Partly

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

No

**Competing Interests:** I collaborate with some of the authors within the framework of the International epidemiological Databases to Evaluate AIDS (IeDEA) and the Measurement & Surveillance of HIV Epidemics (MESH).

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Referee Report 01 December 2017

doi:[10.21956/gatesopenres.13813.r26079](https://doi.org/10.21956/gatesopenres.13813.r26079)



**Mary-Ann Davies**

Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa

This is a well written article describing the data available in the ALPHA network that could potentially be used to assess mortality along the HIV care continuum.

My specific comments are outlined below.

*Abstract: Among PLHIV, the dataset provides information on survival during different periods: prior to diagnosis of infection; following diagnosis but before linkage to care; in pre-antiretroviral treatment (ART) care; in the first six months after ART initiation; among people continuously on ART for 6+ months; and among people who have ever interrupted ART.*

Given the sentence above – as well as the title of the article, I was expecting that the article would actually describe the survival (or its inverse – the mortality) during the different steps along the care continuum. However the article’s focus is actually on the data that is available to describe this mortality, rather than on survival or mortality itself. I suggest that the title and abstract are reworded to make this clear that the article is about the availability and quality of data from the alpha network that could be used to examine mortality along the HIV care continuum, rather than suggesting that the article will describe mortality itself. Perhaps the title should read: “The Network for Analysing Longitudinal Population-based HIV/AIDS data on Africa (ALPHA): Assessing the availability and quality of data on mortality...”

*Methods: Some participants retrospectively reported the dates when they made a transition (first positive diagnosis, linked to care, started ART, interrupted ART). For this analysis of mortality we did not use the retrospectively reported date but instead used the date of first report. This was to avoid biasing the data towards survivors, who were able to make a retrospective report of their progression along the continuum which those who had died before the survey could not do.*

This is not completely clear. I fully understand the need to make some adjustment to avoid biasing the data towards survivors, but it is not clear to me how using the data of first report rather than the retrospectively reported date would achieve this. Surely this would mean that for many patients who retrospectively report they have been diagnosed, linked to care and started ART and interrupted ART it would appear that all of these events happened on the same day. Since this is an article about data validation it may also be helpful to actually show what the difference would be using the date of first report vs the retrospectively reported date.

*Dataset validation: Availability of linked clinical data.*

*The proportion of people with linked clinic data available represents the extent to which PLHIV are diagnosed and enrolled in care, and also the ability of the study site to link DSS and clinical records.* Since proportion of people linked to clinic is a combined outcome both of actually linking to care (a key clinical and programmatic outcome that we really want to measure accurately) and ability of the study site to link DSS and clinical records –which is a research function and has nothing to do with program performance or clinical outcomes, it seems odd to present this as a combined outcome which is really not meaningful, particularly when the percent linked to clinic ranges from 0-100%. I am not sure of the value of this table or of presenting the information unless it is possible to separate the two outcomes or to validate in some way for each site among the diagnosed people actually linked, what proportion the study is able to identify as linked.

I found Figure 2 very difficult to interpret and am not sure what the key message is. I suggest this is made clearer. I think a table with absolute numbers may be clearer as well as some summary text highlighting the key messages.

The article doesn't seem to have a "Conclusion" or "Take home message". This would be very helpful and I would suggest that the authors add this.

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

No

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** HIV cohort research, pediatric HIV clinical epidemiology

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

---