Tutorial for analyses of retention in care according to WHO's definition of loss to follow-up

Andreas Haas, Elizabeth Zaniewski, Matthias Egger

for

Correspondence to:
Andreas Haas
Tel: + 41 31 631 38 67
Email: andreas.haas@ispm.unibe.ch

8 September 2017

Table of Contents

Summary

The World Health Organization (WHO) classifies patients >90 days late for their last scheduled appointment loss to follow-up (LTF). Scheduled appointment dates are often not recorded in HIV program data. In this tutorial, we illustrate how WHO's definition of LTF can be implemented in HIV program data and describe statistical analyses of retention in care and other treatment outcomes including mortality, LTF and transfer between facilities. We provide a dataset with 15,000 simulated patients on antiretroviral therapy (ART) and STATA code to reproduce data management files and statistical analyses.

Dataset

The dataset contains simulated data from three HIV care programs each with 5,000 patients on ART. Patients initiated ART between 1 Jan 2004 and 21 Jun 2015. The dataset is comprised of five tables containing the key variables used in the IeDEA-WHO global analyses of retention in care (Table 1). The simulated data tables are in STATA Version 14 format and can be found under …/RETENTION_TUTORIAL/ SIMULATED_DATA/INPUT_TABLES/.

Table 1: Description of variables in dataset

Variable name	***Type***	**Definition**
PAT table		
ID	*Character*	Code to identify patient (Cohort Patient ID)
PROGRAMME	Character	HIV programme
HAART_D	*YYYY-MM-DD*	Date of ART start (Date of starting antiretroviral combination therapy with any highly active regimen)
CLOSE_D		Date of database closure (provided by ART programme)
DEATH_Y	*Numeric:* *0=No* *1=Yes* *9=unknown*	Has the patient died?
DEATH_D	*YYYY-MM-DD*	Date the patient died.
TRANSFER_D	*YYYY-MM-DD*	Date the patient transferred care to another facility.
TRANSFER_Y	*Numeric:* *0=No* *1=Yes* *9=unknown*	Has the patient transferred?
ART table		
ID	*character*	Code to identify patient (Cohort Patient ID)
ART_SD	*YYYY-MM-DD*	Date of initiation of treatment
ART_ED	*YYYY-MM-DD*	Date of stopping of treatment
RNA table		
ID	*character*	Code to identify patient
RNA_D	*YYYY-MM-DD*	Date of measurement
CD4 table		
ID	*Character (or numeric if possible)*	Code to identify patient (Cohort Patient ID)
CD4_D	*YYYY-MM-DD*	Date of all CD4 measurements (cells or percent)
VIS table		
ID	*Character (or numeric if possible)*	Code to identify patient (Cohort Patient ID)
VIS_D	*YYYY-MM-DD*	Dates of all patient visits

STATA code ("Do-files")

STATA code to reproduce the data management files and statistical analyses of this tutorial are contained in eight "do-files" (Table 2). These STATA do-files have the file format filename.do and can be found under …/RETENTION_TUTORIAL/ DO/.

If you use the do-files with your own HIV program data you need to run them in the correct order. If you use the simulated dataset, you can run any do-file after you have first run 00_DO_FIRST.

Table 2: Description of do-files

File name	**Description**
00_DO_FIRST	Defines file paths and installs user-written packages. Always run first when you start a new STATA session.
01_CLOSING_DATE	Checks completeness of data from beginning of study until closing date. If datasets are incomplete, we calculate an earlier closing date.
02_CLINIC_ENCOUTERS	Generates table with all possible dates indicating that a patient was seen at the facility (clinic encounters table). Dates from all five input tables (pat, vis, art, cd4, and rna) are pooled in the encounters table.
03_VISIT_FREQUENCY	Determines typical appointment schedules of ART programs based on descriptive analyses of clinic encounter dates and visual inspection of frequency plots.
04_NEXT_APPOITNMENT_DATE	Calculates the next appointment dates of patients. Validates the calculated next appointment dates.
05_LOSS_TO_FOLLOW-UP	Classifies patients as LTF according to six different definitions.
06_DEFINE_OUTCOMES	Determines ART outcomes of patients based on the six definitions of LTF. Generates exit date for survival analyses.
07_PLOT_OUTCOMES	Generates stacked area plot of cumulative incidence functions of ART outcomes based on WHO's definition of LTF. Compares Kaplan-Meier estimates of retention in care for all six definitions of LTF.

Detailed description of do-files

00_DO_FIRST

Specify on line 8 of this do-file the location where the accompanying folder “RETENTION_TUTORIAL” has been saved on your local drive.

For example: global dd "C:/Data” (use / instead of \ )

Save the do-file and run this do-file at the beginning of each STATA session.

01_CLOSING_DATE

It is crucial for the valid analyses of retention in care that the closing date of the database is correct. Closing dates provided by ART programmes are often incorrect in IeDEA data. For example, an ART program may report a closing date of 31 Dec 2015 but pharmacy data (ART table) are only available until 30 Jun 2015.

Do-file “01-CLOSING_DATE” checks the completeness of all tables from the beginning of the study (1 Jan 2004) to the closing date. If the closing date is incorrect (i.e. data in one or more tables are incomplete), an earlier closing date is selected. Never choose a closing date that is later than the closing date provided by the ART programme, even if all tables contain data after the provided closing date.

This do-file creates frequency plots of the number of records in all five input tables over calendar time for each of the ART program. Plots are saves in the folder “…\RETENTION_TUTORIAL\FIGURES\ CLOSING_DATE”.

Program A (Figure 1a) has complete data in all tables until the provided closing date, so we do not need to correct the closing date. Program B (Figure 1b), however, does not have complete data up until the provided closing date, so we need to choose an earlier closing date to ensure completeness of data from beginning of study until database closure (Figure 2).

The corrected closing date is saved in the variable my_close_d and will be used in all further analyses.

Figure 1a: Frequency plot of number of records over time

Figure shows the frequency of records in the patient table, ART table, visits table, CD4 table, and viral load table over calendar time for program A. The red line represents the reported closing date and the dashed line the corrected closing date.

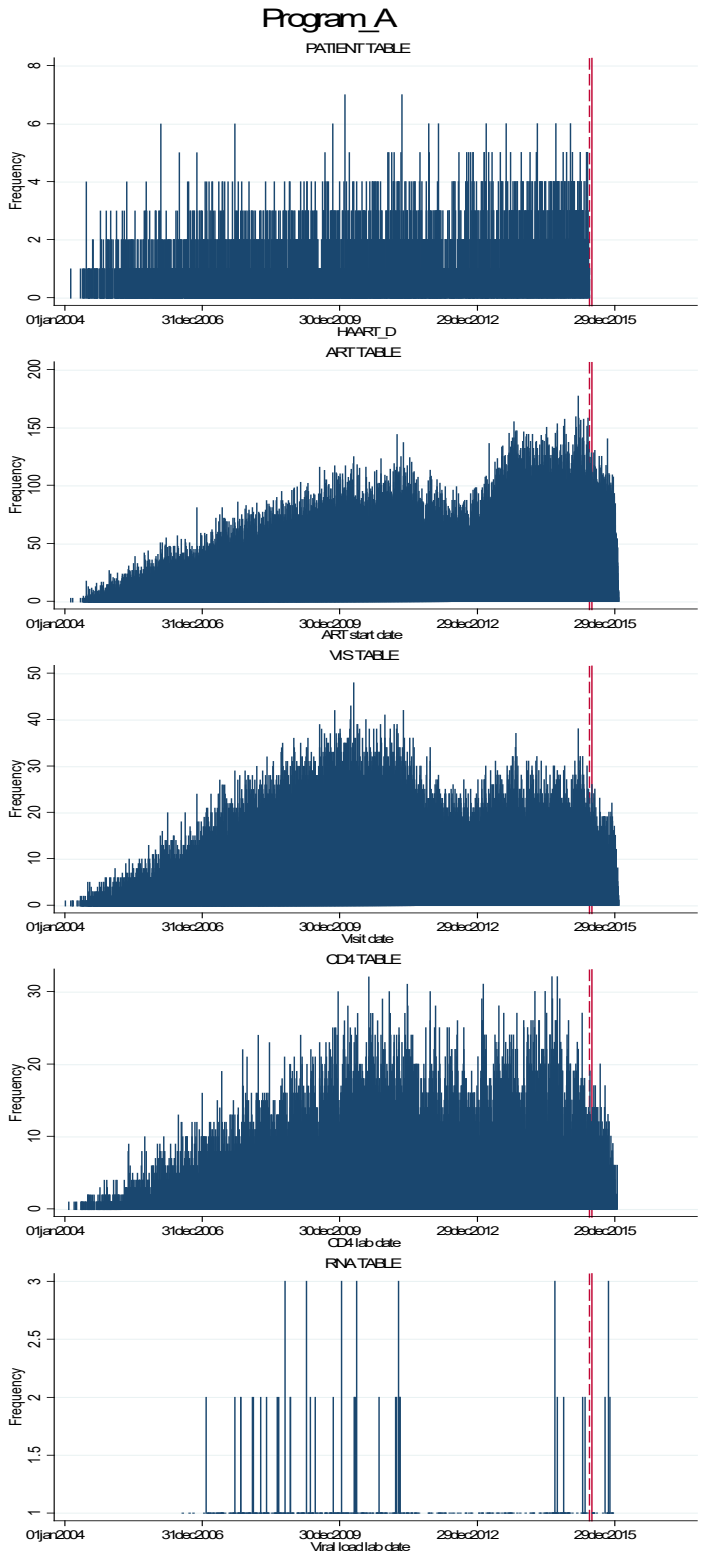
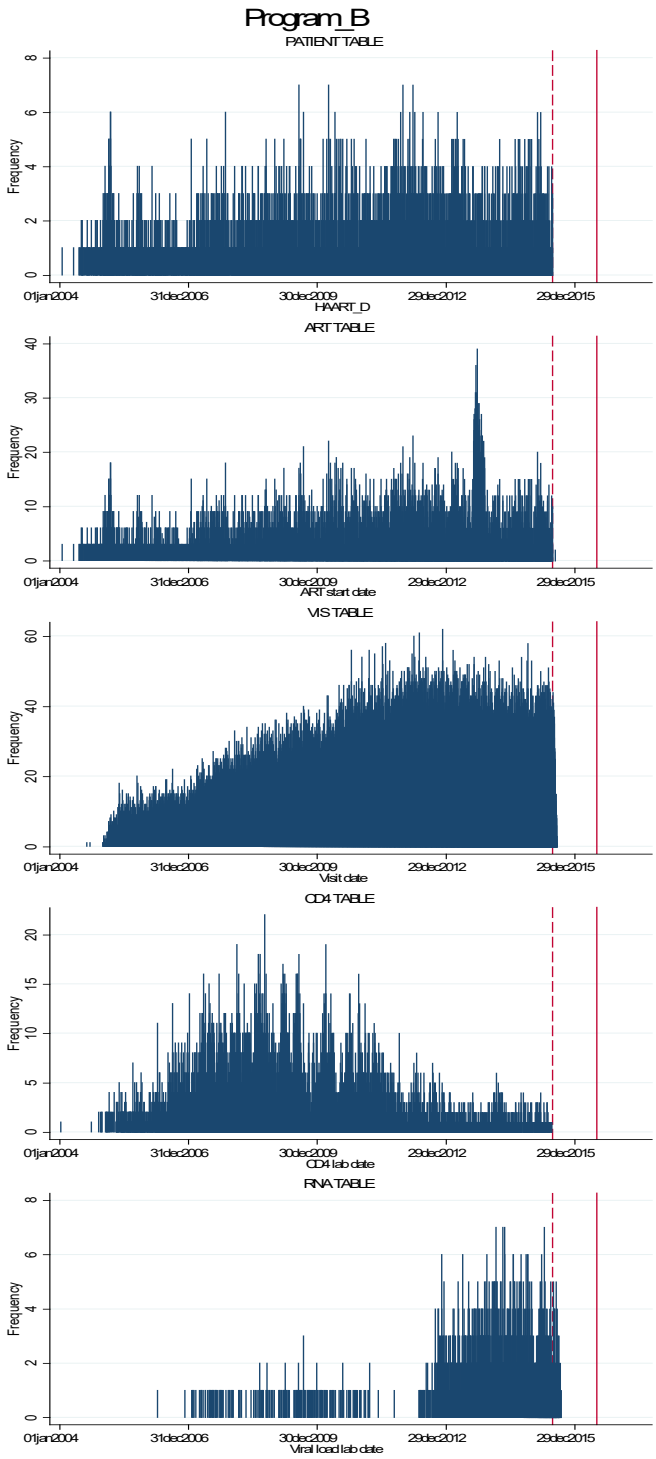


Figure 1b: Frequency plot of number of records over time

Figure shows the frequency of records in the patient table, ART table, visits table, CD4 table, and viral load table over calendar time for program B. The red line represents the reported closing date and the dashed line the corrected closing date.



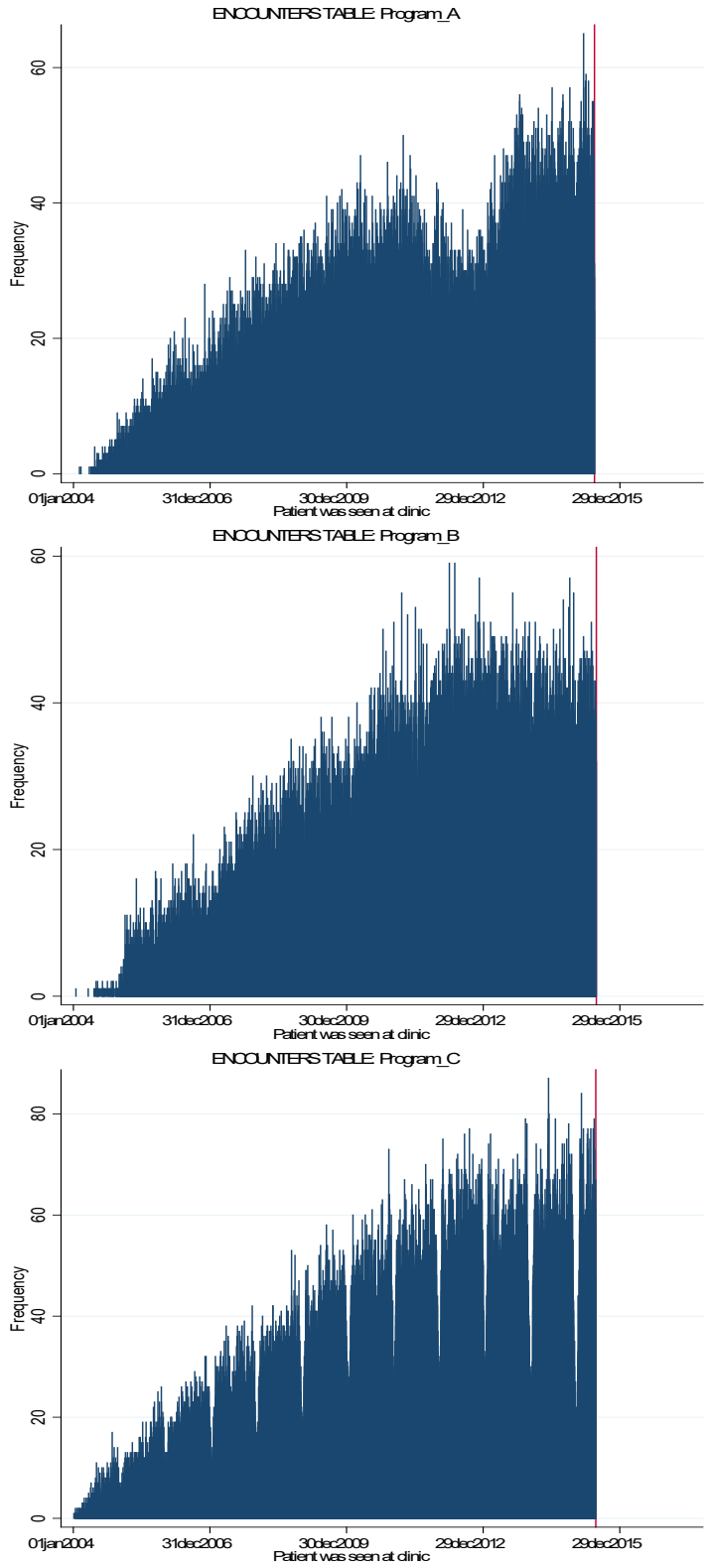
02_CLINIC_ENCOUNTERS

This do-file generates the clinic encounters table, which may contain multiple rows per patient. The clinic encounters table contains all dates (enc_sd) a patient was seen at the facility and is created by pooling all encounter dates contained in the five data tables (rna_dmy, cd4_dmy, art_sd, vis_dmy). The encounters table is unique for combinations of id and enc_sd. Figure 3 shows all the encounter dates pooled from the five data tables for patient A_0015. The encounter table only contains dates up until the corrected closing date.

Figure 2: Encounter table for patient "A_0015".

	id	enc_sd	programme	my_clos~d	haart_d
318.	A_0015	19feb2009	Program_A	30jun2015	19feb2009
319.	A_0015	03apr2009	Program_A	30jun2015	19feb2009
320.	A_0015	05jul2009	Program_A	30jun2015	19feb2009
321.	A_0015	07aug2009	Program_A	30jun2015	19feb2009
322.	A_0015	13sep2009	Program_A	30jun2015	19feb2009
323.	A_0015	07mar2010	Program_A	30jun2015	19feb2009
324.	A_0015	07may2010	Program_A	30jun2015	19feb2009
325.	A_0015	08aug2010	Program_A	30jun2015	19feb2009
326.	A_0015	22aug2010	Program_A	30jun2015	19feb2009
327.	A_0015	10sep2010	Program_A	30jun2015	19feb2009
328.	A_0015	07oct2010	Program_A	30jun2015	19feb2009
329.	A_0015	06nov2010	Program_A	30jun2015	19feb2009
330.	A_0015	06dec2010	Program_A	30jun2015	19feb2009
331.	A_0015	09jan2011	Program_A	30jun2015	19feb2009
332.	A_0015	13feb2011	Program_A	30jun2015	19feb2009
333.	A_0015	15apr2011	Program_A	30jun2015	19feb2009
334.	A_0015	24jul2011	Program_A	30jun2015	19feb2009
335.	A_0015	23sep2011	Program_A	30jun2015	19feb2009
336.	A_0015	23oct2011	Program_A	30jun2015	19feb2009
337.	A_0015	23dec2011	Program_A	30jun2015	19feb2009
338.	A_0015	29dec2011	Program_A	30jun2015	19feb2009
339.	A_0015	03jun2012	Program_A	30jun2015	19feb2009
340.	A_0015	05aug2012	Program_A	30jun2015	19feb2009
341.	A_0015	07sep2012	Program_A	30jun2015	19feb2009
342.	A_0015	07oct2012	Program_A	30jun2015	19feb2009
343.	A_0015	07dec2012	Program_A	30jun2015	19feb2009
344.	A_0015	10feb2013	Program_A	30jun2015	19feb2009
345.	A_0015	10mar2013	Program_A	30jun2015	19feb2009
346.	A_0015	12apr2013	Program_A	30jun2015	19feb2009
347.	A_0015	12jul2013	Program_A	30jun2015	19feb2009
348.	A_0015	08sep2013	Program_A	30jun2015	19feb2009
349.	A_0015	25apr2014	Program_A	30jun2015	19feb2009
350.	A_0015	14aug2014	Program_A	30jun2015	19feb2009
351.	A_0015	23oct2014	Program_A	30jun2015	19feb2009
352.	A_0015	27nov2014	Program_A	30jun2015	19feb2009
353.	A_0015	28dec2014	Program_A	30jun2015	19feb2009
354.	A_0015	29jan2015	Program_A	30jun2015	19feb2009
355.	A_0015	05mar2015	Program_A	30jun2015	19feb2009
356.	A_0015	02apr2015	Program_A	30jun2015	19feb2009
357.	A_0015	30apr2015	Program_A	30jun2015	19feb2009
358.	A_0015	01may2015	Program_A	30jun2015	19feb2009
359.	A_0015	04jun2015	Program_A	30jun2015	19feb2009

Figure 3: Encounter tables containing dates up until the corrected closing date (red line).



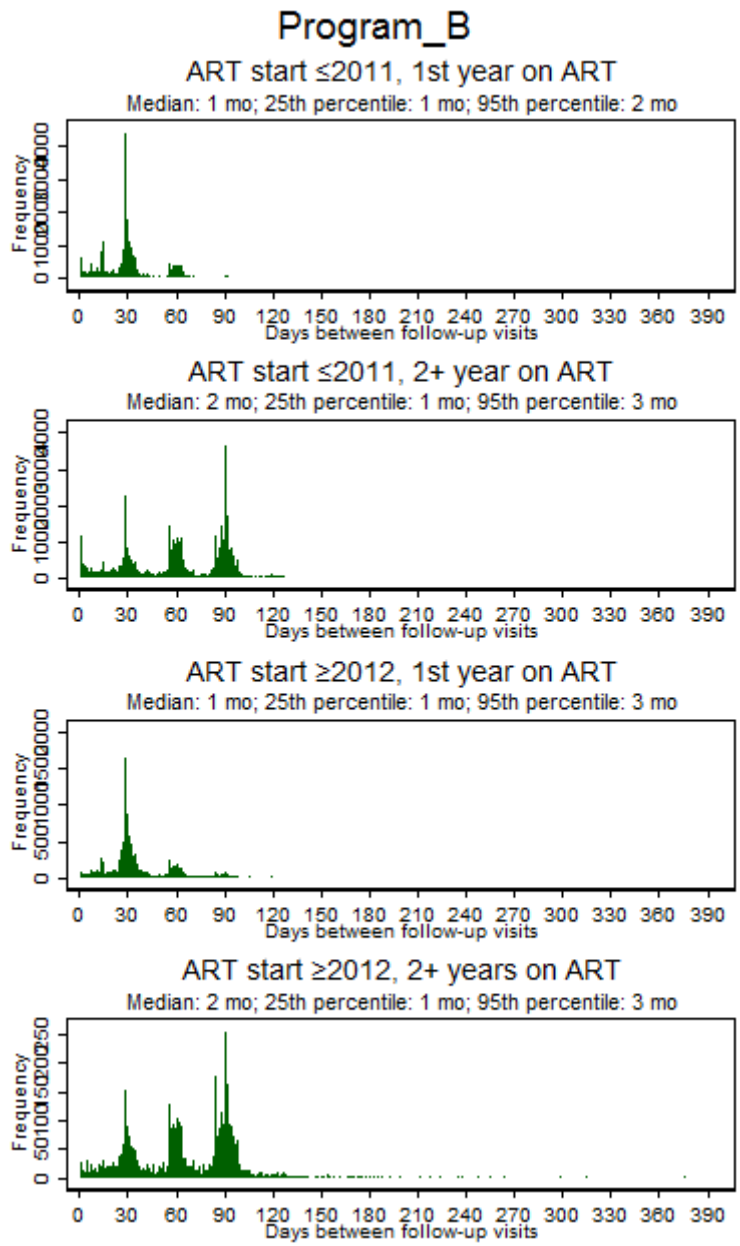
03_VISIT_FREQUENCY

In do-file "03_VISIT_FREQUENCY" we determine the typical appointment schedules of ART programs based on descriptive analyses of clinic encounter dates. We correct these estimates based on visual inspection of frequency plots and knowledge of treatment guidelines and local practices, if necessary.

The do-file calculates the median, 25th and 95th percentile of the gaps between visits (in months) for each ART programme by number of years on ART (1 year, 2+ years) and year of ART start (≤2011, ≥2012). The do-file further generates frequency plots for the gap between visits (in days) for ART program by number of years on ART and year of ART start (Figure 4) and saves them in the folder "RETENTION_TUTORIAL\FIGURES\VISIT_FREQUENCY".

Based on visual inspection of the frequency plots and knowledge of treatment guidelines and local practices, gaps between visits that are <25th percentile or >the 95th percentile may be considered "unscheduled visits". If the median, 25th or 95th percentile values are implausible, we can replace them with more plausible values. (Only plausible visit gaps are used to predict patients' next appointment date in do-file 04_NEXT_APPOINTMENT_DATE). For instance, we could set the value saved in p95 for patients on ART for 2+ years in programme B to 4 months, as it seems that some patients were followed up every 4 months. When in doubt, select the larger gap to avoid misclassifying patients as LTF.

Figure 4: Frequency plot of gap between visits for programme B.



04_NEXT_APPOINTMENT_DATE

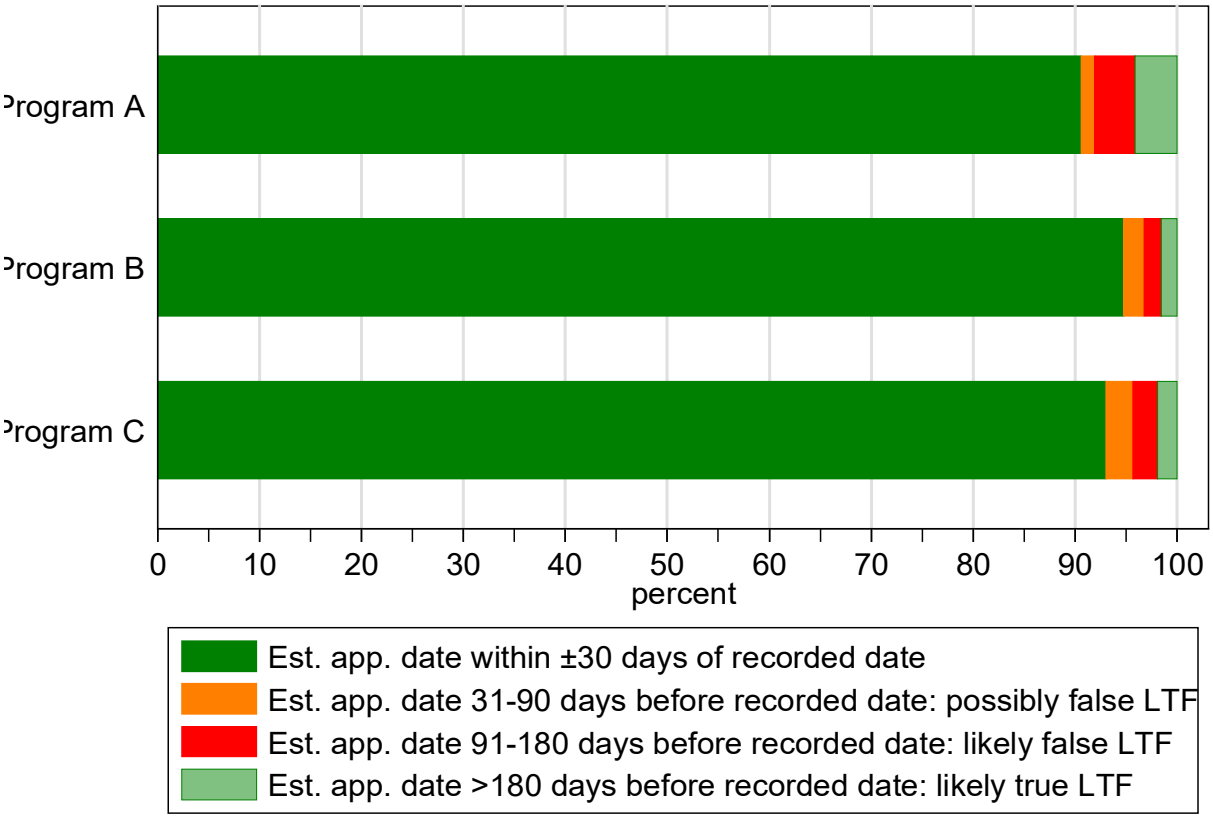
In do-file "04_NEXT_APPOINTMENT_DATE" we calculate patients' next appointment date. The next appointment date is derived by adding the patient's most recent and plausible gap between two consecutive visits (as determined in do-file 03_VISIT_FREQUENCY) to the last visit date.

To validate the algorithm, we calculate the next appointment date based on data up to the next to last visit (i.e. recorded last visit date is excluded) and compare it with the recorded last visit date (Figure 5). For 93% of the patients, the next appointment date was correctly calculated (i.e. within ±30 days of the recorded last visit date). For 2.0% of the patients, the calculated next appointment date was 30 to 90 days before the recorded last visit date. Although it is possible that this error could lead to incorrect classification of patients (false LTF), the WHO definition has a 90 day window so patients visiting the clinic on time would not be misclassified. For 2.7% of patients the calculated next appointment date was 90 to 180 days before the recorded last visit date. These patients would be misclassified as LTF

even if they visit the clinic on time, as this extends beyond the 90 day window. For 2.5% of patients the calculated next appointment date was more than 180 days before the recorded last visit date. These patients are likely correctly classified “true” LTF.

Figure 5: Agreement of the calculated and recoded last appointment date.

Figure shows the agreement of calculated next appointment date (based on data until the next to last visit) and the recorded last visit date by ART program.



05_LOSS_TO_FOLLOW-UP

In do-file “05_LOSS_TO_FOLLOW-UP” we classify patients LTF according to two different definitions across three time windows (90, 180, 365 days late) to create six possible scenarios.

First, we use the WHO definition across the three different time windows. Patients more than 90, 180, or 365 days late for their next scheduled clinic appointment who did not return to care before database closure are classified as LTF on the day of their last visit, respectively.

Second, we use another definition that classifies patients LTF the first time they have a gap in care of more than 90, 180, or 365 days, respectively. This definition can be used to assess time trends in retention (Johnson et al Am J Epi 2014). We recommend using this definition with a time window of at least 180 days and only if transient treatment interruptions are reliably recorded. This may not be the case at all IeDEA clinics.

06_DEFINE_OUTCOMES

In do-file "06_DEFINE_OUTCOMES" we define ART outcomes for patients according to the six LTF scenarios defined in the previous do-file. ART outcomes are retained in care, transferred out, LTF, and dead. Outcomes are mutually exclusive. Dead and transfer out always supersede retention in care and LTF. Patients not LTF, transferred out, or dead are classified as retained in care.

We further calculate the date patients exit the survival analyses. Patients exit the analyses after a maximum of 10 years of follow-up or when they stop being at risk for LTF (i.e. depending on time window used in the LTF definition 90, 180, or 365 days before database closure, respectively).

07_PLOT_OUTCOMES

Do-file "07_PLOT_OUTCOMES" generates a stacked area plot with cumulative incidence functions of ART outcomes for each of the six LTF scenarios (Figure 6).

We follow patients from ART initiation for a maximum of 10 years. Individuals who transferred are censored on the date of transfer. Individuals retained on ART are censored 90 days before database closure, when they were no longer at risk of loss to follow-up. In sensitivity analyses, we compare Kaplan-Meier estimates for retention in care for all six scenarios of LTF (Figure 7).

Figure 6: Cumulative incidence of ART outcomes after ART initiation.

We use the Aalen-Johansen estimator to calculate the cumulative incidences of documented death and loss to follow-up and death. Death and LTF are competing risks. We use the Kaplan-Meier estimator to calculate the cumulative probability for retention in care. Death and LTF are failure events in Kaplan-Meier analyses.

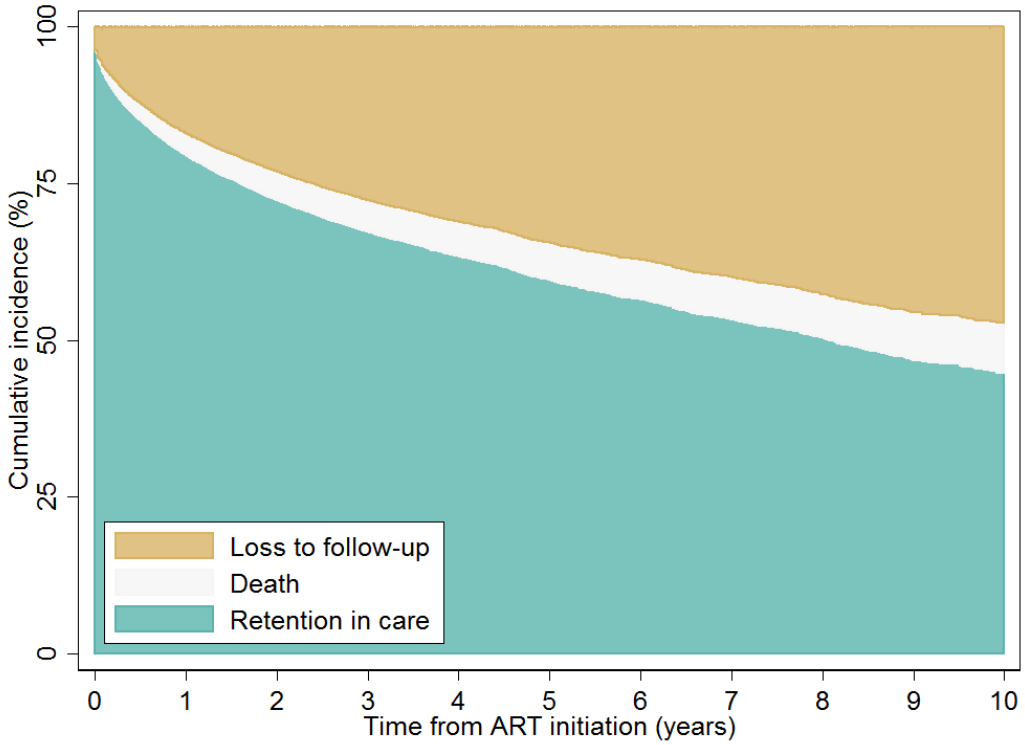


Figure 7: Kaplan-Meier estimates for retention in care based on different definitions of loss to follow-up (LTF).

Solid lines: Patients late for more than more than 90, 180, or 365 days for their clinic appointment who did not return to care before database closure are classified LTF

Dashed lines: Patients ever late for a clinic appointment for than 90, 180, or 365 days are classified LTF.