Data specification for ALPHA HIV incidence data

The incidence dataset includes data on resident study participants aged 15-49 who have had one negative HIV test and at least one subsequent HIV test done within the study for research purposes. Data may be split into multiple observations for each person to describe time spent in different residencies, different age groups and different calendar year periods. All timings are given as age at event. To enable multiple imputation of seroconversion dates we have provided the times (ages) of last negative test and first positive test.

Table 1: Specification for shared ALPHA incidence dataset

Variable name	Description	Coding	
study_name	Study name	1	Karonga
		2	Kisesa
		3	Manicaland
		4	Masaka
		5	Rakai
		6	uMkhanyakude
		7	Agincourt
		8	Kisumu
		9	Ifakara
		10	Nairobi
idno	Participant ID number for sharing		
sex	Sex	1	Men
		2	Women
fouryear	Calendar year, grouped in 4 years post 2005	0	1995-99
		1	2000-04
		2	2005-08
		3	2009-12
		4	2013-16
agegrp	Five year age group	3	15-19
		4	20-24
		5	25-29
		6	30-34
		7	35-39
		8	40-44
		9	45-49
timein	Start of episode (age)		
timeout	End of episode (age)		
timelastneg	Age at last negative test		
timefirstpos	Age at first positive test		
fail	Seroconversion occurred at the end of this episode	0	No
		1	Yes
_t0	Stata variable: age at start of episode		
_t	Stata variable: age at end of episode		
_st	Stata variable: in survival analysis		
_d	Stata variable: failure (seroconversion)		