Variables Property of Incidence Dataset

Number of observations: 11897

Number of variables: 30

Purpose: Analysis dataset for Incidence manuscript

Variable	Label	Description	Definition	Type	Format	Format Definition
uid	uid	Patient ID number		numeric		
hhid	Household ID	Household ID	First seven digits of ptid + 00	numeric		
SHIMS_EA	EA	EA number	Swaziland EA Used for sampling	numeric	BEST12.	
Region	Region	Region number	First digit of ptid	numeric	regionf.	1=”Hhohho”; 2=”Manzini”; 3=”Shiselweni”; 4=”Lubombo”;
urban	resid	residence	Residence of household	numeric	resid.	1=”urban” 2=”Rural”
REQ_erdt	RQ/EQ date	T1 Enrollment Date	Date recorded on REQ form	char		
fup_date	Follow-up Date	Follow-up date at T2	Date recorded on SCF form	numeric	DDMMYY10.	DDMMYY10.
gender	gender	Gender at T1	A composite variable derived from REQsex and sex reported on HC/AHA	numeric	1ender.	1=”Male” 2=”Female”
age	age	Age at T1	A composite variable of REQyears, REQmy, and age recorded on HC/AHA.	numeric	BEST12.	age in years
age_grp	age group	Age group at T1 (18-49)	Age group at T1 ranging from 18-49	numeric	agegrp18.	1=’18-19’;2=’20-24’;3=’25-29’;4=’30-34’; 5=’35-39’;6=’40-44’;7=’45-49’;
education	education	Education at T1	Derived based on REQhlsa	numeric	eduf	1=’Primary’; 2=’Secondary’; 3=’Tertiary’;

						4='Did not attend'; 5='Missing';
marital	marital status	marital status at T1	if Rqmarrie=3 and sex_age>0 then marital=1; if Rqmarrie=3 and sex_age=-1 then marital=2;if Rqmarrie=3 and sex_age=. then marital=3; if Rqmarrie in (1,2) and Rqplwy=1 then marital=4; if Rqmarrie in (1,2) and Rqplwy=2 then marital=5; if Rqmarrie in (1,2) and Rqplwy=. then marital=6; 7="REF"=7, 8="missing item" 9="missing plate"	numeric	marif.	1="Not married, ever had sex"; 2="Not married, never had sex"; 3="Not married, unknown sex history"; 4="Married, living with partners"; 5="Married, partners living elsewhere"; 6="Married, unknown living situation"; 7-9="Missing";
sex_age_t1	sex debut age	sex debut age at T1	if Rqhnref=1 or Rqftage=0 then, sex_age_t1=-1; else sex_age_t1=Rqftage;	numeric	sexage.	-1="Never had sex"
sex_age_t2	sex debut age	sex debut age reported at T2	if SCFhnref=1 or SCFftage=0 then, sex_age_t2=-1; else sex_age_t2=SCFftage;	numeric	sexage.	-1="Never had sex"
sexprt_num_t1	sex partners in the last 6 months (num)	number of sex partners in the last 6 months at T1	Derived based on RQtnp6m	numeric	sexnumf.	-1, -2,-3 ="Missing";
sexprt_num_t2	sex partners in the last 6 months (num)	number of sex partners in the last 6 months at T2	Derived based on SCFtn6pm	numeric	sexnumf.	-1,-2,-3="Missing";
sexprt_cat_t1	sex partners in the last 6	number of sex partners in the last 6	Categorical variable of sexprt_num_t1	numeric	sexcatf.	1="0"; 2="1"

	months (cat)	months				3=”2 or more”;
sexprt_cat_t2	sex partners in the last 6 months (cat)	number of sex partners in the last 6 months	Categorical variable of sexprt_num_t2	numeric	sexcatf.	1=”0”; 2=”1”; 3=”2 or more”;
preg_t1	pregnancy status	Female participant pregnancy status at T1	Derived based on Rqpreg	numeric	pregf.	1=”Yes”; 2=”No”; 3-5=”Missing”;
preg_t2	pregnancy status	Female participant pregnancy status at T2	Derived based on SCFpreg	numeric	pregf.	1=”Yes”; 2=”No”; 3-5=”Missing”;
circum_t1	circumcision status	Male participant circumcision status at T1	Derived based on Rqcircum	numeric	circumf.	1=”Circumcised”; 2=”Uncircumcised”; 3-5=”Missing”;
circum_t2	circumcision status	Male participant circumcision status at T2	Derived based on SCFcircum	numeric	circumf.	1=”Circumcised”; 2=”Uncircumcised”; 3-5=”Missing”;
hiv_testhistory	hiv test history	hiv test history	Derived based on RQtest, RQresult, RQrltest.	numeric	selfhivf.	0=”Never had test before” 1=”Positive” 2=’Negative” 3=”Indeterminate” 4=”DK” 5=”REF” 6=”Ever had test – did not receive positive or negative result” 7=”missing”
hivstatus_t1	hiv status	hiv status at T1	Derived based on HTR form and lab datasets	numeric	hivstatusf.	1=”Neg” 2=”Pos”
hivstatus_t2	hiv status	hivstatus at T2	Derived based on FHT form and lab datasets	numeric	hivstatusf.	1=”Neg”; 2=”Pos”;
seroconvert	seroconvert status	Participants’ seroconversion	if hivstatus_t1=1 and hivstatus_t2=2 then seroconvert=1; if hivstatus_t1=1 and	numeric	serof.	1=”Yes”; 2=”No”;

			hivstatus_t2=1 then seroconvert=0			
wt_poststratification	ResponseWgt	Weight to adjust nonresponse	After survey completion, poststratification weights were computed correcting for non-response by region, gender and age, based on the complete sample frame from the household survey.	numeric		continuous
wt_design	probability of selection	Probability of selection of EAs	The design weights of the survey design: Eight strata defined by region and rural/urban were used. Evaluation Areas (EAs) were selected with weight proportional to population size and 26 households were selected at random from all households in the EA.	numeric		continuous
wt_population	UnrakedWgt	Weight, combined design and response, before raking applied	Combines the design and poststratifiation weights, the sum of the weights estimates the distributions of the total population of Swaziland.	numeric		continuous
wt_cohort	RakedWgt	Raked Weight according to population	Final survey weights for use in analysis, where the sum of the weights	numeric		continuous

			approximately equals the total population of the enrolled cohort. These use the post-stratification weights further corrected using raking to match the region, gender and age categories of the 2007 DHS population census. These correspond to the weights used in the manuscript.			