Overview of SHIMS Datasets

Analysis Datasets

The analysis datasets correspond to data files developed in conjunction with manuscripts reporting results from the SHIMS study. For many analysis questions, these analysis datasets will define a relevant cohort for additional questions of interest. Analysis data files for new questions can be derived by merging new variables from the raw data files with these analysis datasets.

Prevalence Manuscript

This data file contains the records for reporting the HIV Prevalence findings from the Baseline SHIMS Cohort: the 18,172 Swaziland participants enrolled at baseline. The majority of the analysis variables are derived (i.e. computed) from the raw data collected at baseline, as documented.

HIV status at baseline of each participant, as supplied in this data set, was adjudicated by the SHIMS Principal Investigators and combined information from the HIV testing at the Households together with confirmatory testing at the Swaziland National Laboratory.

Participants are included in this data set if they were at least 18 and not yet 50 at the time of enrollment, had known gender and HIV status of HIV-infected or HIV-uninfected (i.e. HIV indeterminate status are omitted).

This dataset contains survey weights to allow computation of population based estimates, using survey methods. Three weights are included
wt_design: 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.
wt_poststratification: 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.
wt_population: Combines the design and poststratifiation weights, the sum of the weights estimates the distributions of the total population of Swaziland.
wt_cohort: Final survey weights for use in analysis, where the sum of the weights 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.

Reference:

Bicego GT, Nkambule R, Peterson I, Reed J, Donnell D, Ginindza H, Duong YT, Patel H, Bock N, Philip N, Mao C, Justman J. Recent patterns in population-based HIV prevalence in Swaziland. PLoS One. 2013 Oct 15;8(10):e77101. doi: 10.1371/journal.pone.0077101. PMID: 24143205; PMCID: PMC3797108.

PVL Manuscript

This data file contains the records for reporting the HIV Plasma Viral Load findings from the HIV-infected Baseline SHIMS Cohort: the 5,803 Swaziland participants who were HIV-infected at baseline. The majority of the analysis variables are derived (i.e. computed) from the raw data collected at baseline, as documented.

Plasma viral load, measured by the Swaziland National Laboratory, is included in the file.

This dataset contains survey weights to allow computation of population based estimates, using survey methods. Weights were used directly from the entire cohort for the subset of HIV-infected (i.e. the weights are identical to the HIV prevalecne cohort).
Three weights are included.
wt_design: 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.
wt_poststratification: 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.
wt_population: Combines the design and poststratifiation weights, the sum of the weights estimates the distributions of the total population of Swaziland.
wt_cohort: Final survey weights for use in analysis, where the sum of the weights 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.

Incidence Manuscript

This data file contains the records for reporting the HIV Incidence findings from the HIV-uninfected SHIMS Incidence Cohort: the 11,897 Swaziland participants who were HIV-uninfected at baseline and enrolled in the incidence cohort (i.e. agreed to be retested for HIV six months later). The majority of the analysis variables are derived (i.e. computed) from the raw data collected at baseline, as documented.

HIV status at follow-up (i.e. identified HIV incidence cases) was adjudicated by the study Principal Investigators and includes HIV testing in the household, confirmatory testing conducted in the Swaziland National Laboratory and confirmatory testing of HIV status at baseline. There were 148 incident cases identified.

Participants are included in this data set if they were at least 18 and not yet 50 at the time of enrollment, had known gender and confirmed HIV status of HIV-uninfected at enrollment.

This dataset contains survey weights to allow computation of population based estimates, using survey methods. Weights were recomputed for this cohort, to adjust for the non-response resulting from loss to follow-up at the second HIV testing visit.
Three weights are included.
wt_design: 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.
wt_poststratification: After completion of the second HIV test at 6 months post baseline, poststratification weights were computed correcting for 1) non-response by region, gender and age, based on the complete sample frame from the household survey and 2) non-response by region, gender and age, based on the enrolled cohort.
wt_population: Combines the design and poststratification weights, the sum of the weights estimates the distributions of the total population of Swaziland.
wt_cohort: Final survey weights for use in analysis, where the sum of the weights 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.

Baseline Household Questionnaire

This dataset correspond to data files for all sampled households. Data from consenting households was collected on all individuals in the household.

Household Level Dataset

This data file contains household level variables for 14,891 selected households.

Household composition

This data file contains data on listed individuals for 14,951 households approached by SHIMS. 35 household IDs were omitted from the Household Level Dataset, as no data items were entered on the CRF.

Household Additional Contact Attempts

This data file details additional contact attempts for households not reached with the initial 6 attempts.

Refusal: Pre-Cohort Survey

This data file details refusal reasons for eligible individuals in the household who did not consent for participation in the Pre-Cohort (Baseline) Survey

Baseline Cohort

This data file contains the raw data collected from eligible participants who consented to enroll in the Pre-Cohort (Baseline) survey: the 18,172 Swaziland participants who were tested for HIV at baseline. Two versions of the questionnaire were administered: routine (RQ) and extended (EQ), where the EQ was administered to approximately 10% of the cohort.

Combined RQ and EQ

This data file contains the raw data variables common to both the RQ and EQ, merged to form a complete dataset for all participants.

Routine Questionnaire

This data file contains the raw data variables common to both the RQ and EQ, merged to form a complete dataset for all participants.

Extended Questionnaire

This data file contains the raw data variables common to both the RQ and EQ, merged to form a complete dataset for all participants.

Refusal: Short-Term Cohort

This data file details refusal reasons for eligible HIV-uninfected individuals in the household who did not consent for participation in the Short-term Cohort (Incidence Cohort)

HIV Incidence Cohort

These data files contain the raw data collected from eligible participants who consented to enroll in the Short-term Cohort (Incidence Cohort): the 11,977 Swaziland participants who were tested in the household as HIV-uninfected at baseline and consented to participate in the cohort.

Short-term Cohort Enrollment Status

This data file contains variables collected on all consented Pre-cohort Survey participants about their eligibility and enrollment in the Short-Term Cohort.

Pilot Short-Term Cohort Follow-up

This data file contains data from the first 49 participants in the 6-month follow-up visit collected using pilot forms

Short-Term Cohort Follow-up

This data file contains the 6-month follow-up visit data collected from the 11,977 persons who consented to short-term follow-up.

Termination

This data file contains the reasons for early termination for participants lost to follow-up in the short-term follow-up cohort.

CD4 count cohort

An amendment to the SHIMS study was implemented, which sought to evaluate CD4 counts on approximately 1000 SHIMS participants who were HIV-infected at baseline. At the time of implementation, participants were selected at random amongst HIV-infected participants whose informed consent period had not elapsed (i.e. those enrolled later in the cohort had a higher chance of selection into Cohort C)

Enrollment Status: Cohort C

This data file contains the enrollment status information for those approached for participation in Cohort C.

Questionnaire: Cohort C

This data file contains the information collected from those consented and enrolled in Cohort C.

CD4 Data: Cohort C

This data file contains the CD4 counts measured on participants enrolled in Cohort C.

Refusal: Cross Sectional Cohort C

This data file contains the reasons for refusal amongst those approached for participation in Cohort C.

Reference:

Justman J, Reed JB, Bicego G, Donnell D, Li K, Bock N, Koler A, Philip NM, Mlambo CK, Parekh BS, Duong YT, Ellenberger DL, El-Sadr WM, Nkambule R. Swaziland HIV Incidence Measurement Survey (SHIMS): a prospective national cohort study. Lancet HIV. 2017 Feb;4(2):e83-e92. doi: 10.1016/S2352-3018(16)30190-4. Epub 2016 Nov 16. PMID: 27863998; PMCID: PMC5291824.

Acknowledgements:

Manuscripts based on analysis of the SHIMS data should include the following acknowledgement: SHIMS was funded by the President's Emergency Plan for AIDS Relief (PEPFAR) by the Centers for Disease Control and Prevention. In addition, the following organisations or entities were involved in data collection and analysis: the Government of the Kingdom of Swaziland's Ministry of Health, including the National Reference Laboratory and the

Swaziland National AIDS Programme, and the Swaziland Central Statistical Office; the U.S. Centers for Disease Control and Prevention; ICAP at Columbia University; Epicentre South Africa; Maromi South Africa; the Fred Hutchinson Cancer Research Center, Statistical Center for HIV/AIDS Research and Prevention (SCHARP); and, the U S Agency for International Development (USAID).

Contact

For further information about these datasets, please contact Deborah Donnell of Fred Hutchinson Cancer Research Center (deborah@fredhutch.org).
For further information about SHIMS please contact Jessica Justman, ICAP, Columbia University jj2158@cumc.columbia.edu.