Data set metadata record

Data set ID
SABSSM 2005 Visiting point

Data set title

Citation

Data set description
This data file contains household information about the people who usually live and slept there (the previous night) as per the time of the interview. Their age, sex, marital status, race, language spoken, education status, main source of drinking water, energy for cooking, type of toilet facility is included.

The data set contains 25 variables and 10548 cases.

Data set abstract
South Africa continues to have the largest number of people living with HIV/AIDS in the world. This study intends to understand the determinants that lead South Africans to be vulnerable and susceptible to HIV. This is the second in a series of household surveys conducted by the Human Sciences Research Council (HSRC), that allow for tracking of HIV and associated determinants over time using the same methodology used in the 2002 survey, thus making it the first national-level repeat survey. The interval of three years allows for an exploration of shifts over time against a complex of demographic and other variables, as well as allowing for investigation of the new areas. The survey provides the first nationally representative HIV incidence estimates.

The study key objectives were to: Determine HIV prevalence and incidence as well as viral load in the population; Gather data to inform modelling of the epidemic; Identify risky behaviours that predispose the South African population to HIV infection; examine social, behavioural and cultural determinants of HIV; explore the reach of HIV/AIDS communication and the relationship of communication to response; assess the relationship between mental health and HIV/AIDS and establish a baseline; assess public perceptions of South Africans with respect to the provision of anti-retroviral (ARV) therapy for prevention of mother-to-child transmission and for treating people living with HIV/AIDS; understand public perceptions regarding aspects of HIV vaccines; and investigate the extent of the use of hormonal contraception and its relationship to HIV infection.

In the 10 584 valid visiting points that agreed to participate in the survey, 24 236 individuals were eligible for interviews and 23 275 completed the interview. Of the 24 236 individuals, 15 851 agreed to HIV testing and were anonymously linked to the behavioural interviews. The household response rate was 84.1 % and the overall response rate for HIV testing was 55 %.
Time method
Cross-section

Time period comment
Person life span until 2005.

Origin
Primary data

Granularity
Micro level data

Type of data
Quantitative

Kind of data
Single tabular (Component of related data sets)

Production date
2005

Version
1.0

Countries
South Africa

Geographic coverage
Nine provinces in South Africa: Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng, Mpumalanga, Limpopo.

Geographic unit
Province and Geography type (Geotype): Urban formal, Urban informal, Tribal area (Rural informal), Rural formal (farm)

Unit of analysis
All individuals living in the selected households were selected and were eligible to participate including those living in hostels, but individuals staying in educational institutions, old age homes, hospitals, homeless people, and uniformed-service barracks were excluded from the survey. The study included individuals aged 2+ years, Africans, whites, coloured, Indians and others

Universe - Included
South African population, 2 years and older from urban formal, urban informal, rural formal (farms), rural informal (tribal area) settlements.

Universe - Excluded
Individuals staying in educational institutions, old age homes, hospitals, homeless people, uniformed-service barracks and persons younger than 2 years were excluded.

Date of collection
01 OCT 2004 - 30 JUN 2005
Some fieldwork took place over six weeks between October and December 2004, but the bulk of the fieldwork was undertaken from mid-January 2005 to June 2005.

**Mode of data collection**
- Clinical measurements
- Face-to-face interview
- Focus group
- Observation

**Sampling procedure**

This project used the HSRC's master sample (HSRC 2002). A master sample is defined as a selection, for the purpose of repeated community or household surveys, of a probability sample of census enumeration areas throughout South Africa that are representative of the country's provincial, settlement and racial diversity. The sampling frame that was used in the design of the Master Sample was the 2001 census Enumerator Areas (EAs) from Statistics South Africa (Stats SA). The target population for this study were all people in South Africa, excluding persons in so called 'special institutions' (e.g. hospitals, military camps, old age homes, schools and university hostels). The EAs were used as the Primary Sampling Units (PSUs) and the Secondary Sampling Units (SSUs) were the visiting points (VPs) or households (HHs). The Ultimate Sampling Units (USUs) were the individuals eligible to be selected for the survey. Any member of the household 'who slept here last night', including visitors was an eligible household member for the interview. This sampling approach was used in the 2001 census and is a standard demographic household survey procedure.

The sample was designed with two main explicit strata, the provinces and the geography types (geotype) of the EA. In the 2001 census, the four geotypes were urban formal, urban informal, rural formal (including commercial farms) and tribal areas (rural informal) (i.e. the deep rural areas). In the formal urban areas, race was used as a third stratification variable. What this means is that the Master Sample was designed to allow reporting of results (i.e. reporting domain) at a provincial, geotype and race level. A reporting domain is defined as that domain at which estimates of a population characteristic or variable should be of an acceptable precision for the presentation of survey results. A visiting point is defined as a separate (non-vacant) residential stand, address, structure, and flat in a block of flats or homestead. The 2001 estimate of visiting points was used as the Measure of Size (MOS) in the drawing of the sample. A maximum of four visits were made to each VP to optimise response. Fieldworkers enumerated household members, using a random number generator to select the respondent and then proceeded with the interview.

All people in the households, resident at the visiting point aged 2 years and older were initially listed, after which the eligible individual was randomly selected in each of the following three age groups 2-11, 12-14 and 15 years and older. These individuals constituted the USUs of this study. Having completed the sample design, the sample was drawn with 1 000 PSUs or EAs being selected throughout South Africa. These PSUs were allocated to each of the explicit strata. With a view to obtaining an approximately self-weighting sample of visiting points (i.e. SSUs), (a) the EAs were drawn with probability proportional to the size of the EA using the 2001 estimate of the number of visiting points in the EA database as a measure of size (MOS) and (b) to draw an equal number of visiting points (i.e. SSUs) from each drawn EA. An acceptable precision of estimates per reporting domain requires that a sample of sufficient size be drawn from each of the reporting domains. Consequently, a cluster of 15 VP was systematically selected on the aerial photography produced for each of the EAs in the master sample. Since it is not possible to determine on an aerial photograph whether a `dwelling unit' is indeed a residential structure or whether it was occupied (i.e. people sleeping there), it was decided to form clusters of 15 dwelling units per PSU, allowing on average for one invalid dwelling unit in the cluster of 15 dwelling units. Previous experience at Statistics SA indicated a sample size of 10 households per PSU to be very efficient, balancing cost and efficiency. The VP questionnaire was administered by the fieldworker, and in
follow-up, participant selection was made by the supervisor. Participants aged 12 years and older who consented were all interviewed and also asked to provide dried blood spots (DBS) specimens for HIV testing. In case of 2-11 years, parents/guardians were interviewed but DBS specimens were obtained from the children.

The sample size estimate for the 2005 survey was guided by (1) the requirement for measuring change over time and to be able to detect a change in HIV prevalence of 5 % points in each of the main reporting domains, and (2) the requirement of an acceptable precision of estimates per reporting domain, say a precision less than 74% with a design effect of 2 units. Overall, a total of 23 275 participants composed of 6 866 children (2-14 years), 5 708 youths (15-24 years) and 10 687 adults (25+ years) were interviewed. The sample was designed with the view to enable reporting of the results on province level, on geography type area and on race of the respondent. The total sample size was limited by financial constraints, but based on other HSRC experience in sample surveys it was decided to aim at obtaining a minimum of 1 200 households per race group. The number of respondents per household for the study was expected to vary between one and three (one respondent in each of the three age groups). More females (68.3%) than males (62.2%) were tested for HIV. The 25+ years age group was the most compliant (71.3%), and 2-14 years the least (54.6%). The highest response rates were found in rural formal locality types (74.5%) and the lowest in urban formal locality types (61.7%).

**Weighting**

Weighting procedures were undertaken before analysis of the data as follows:

**Step 1 - Calculating the sampling weight:**

The SAS (Survey Analysis Software) procedure Surveyselect was used to draw the sample of EAs. The EAs were drawn using probability proportional to size (pps) sampling and the estimated number of visiting points was used as the measure of size (MOS). Therefore, the data file of drawn EAs contained the selection probabilities as well as the sampling weights of these EAs. In the case of small EAs (i.e. EAs with an estimated number of visiting points less than 100), one or more neighbouring EAs were also identified and combined with the small EA to form a new primary sampling unit (i.e. PSU). Thus, a PSU is defined for the study as a cluster of at least 100 visiting points (estimated). The first step was to calculate the sampling weight of those PSUs consisting of more than one EA. This was done simultaneously with step 3.

**Step 2 - Compiling a list of substituted EAs/PSUs:**

A list of all EAs/PSUs that were substituted during the study was compiled. This list included the following details relating to the original and substituted EAs/PSUs:

- EA number, EA type description, geotype, region name, place name and reason for substitution;
- Details of EAs/PSUs entirely deleted from the sample (if any) due to whatever reason;
- Finally, detailed information was also required on the PSUs where the geotype and/or EA type were incorrect in the original pre -census 2001 EA database (e.g. an EA with an informal urban code was drawn when in fact it was a formal urban area).

**Step 3 - Integration of original and realised EA lists:**

This step involved the integration of the information on substituted and/or deleted EAs/PSUs (step 2) with the original drawn sample (step 1). This step also included the `correction' of the sampling weights of the substituted EAs/PSUs, as well as `correcting' the consequences of any deleted EAs/PSUs.

**Step 4 - Provision of information on realised sample:**
In respect of all realized EAs/PSUs or the final sample used in this study, the following information was required:

- The number of visiting points counted on the aerial photographs (or obtained by a physical count in the field);
- The number of invalid visiting points (such as an empty house, not being a residential unit, etc.) among the 15 systematically drawn visiting points in the PSU;
- The actual number of visiting points among the 15 that participated in the study. Note that, in the case of a visiting point with more than one household, a visiting point would have been considered as a participant even if only one household at the visiting point responded (i.e. if one or more records are available for that VP);
- In the case of farms, the total number of farms in the EA/PSU as well as the number of farms actually drawn.

**Step 5 - Calculating the visiting point sampling weight:**

The visiting point sampling weight was the counted number of visiting points in the EA divided by the number of visiting points participating in the survey. The final visiting point sampling weight was then taken as the product of the EA sampling weight and the visiting point sampling weight.

**Step 6 - Providing visiting point information:**

The next step involved collecting the following information for all households at every participating visiting point in all responding EAs/PSUs:

- The number of persons by gender and age for all age groups (specifically including children 2-11 and 12-14 years, youth and adults 15 years and older). Only one person was randomly drawn in each of the three age groups;
- The (dominant) race, whether or not a drawn person refused to respond;
- The number of households at visiting points that were not prepared to participate in the study;
- In the case of a farm EA, the total number of visiting points counted on the aerial photographs covering all farms or, alternatively, the total number of farms, the number of farms selected, the number of owner houses and the number of labourer houses on each selected farm and whether an owner house and labourer houses or only labourer houses were selected on the farm;
- The number of labourer houses selected on each selected farm; The total number of owner and labourer households as well as the number of selected households on each selected farm. In respect of the selected households, the same personal information regarding gender, age and race as indicated above was required.
- Refusal and other non-response information were also required.

**Step 7 - Calculation of final record weight:**

In this step the integration of all the above information occurred and the final sampling weight for each data record was calculated. This weight is equal to the final visiting point sampling weights (as given in step 5) multiplied by the selected person's sampling weight per visiting point per age group. This process produced a final sample representative of the population in South Africa for gender, age, race, locality type and province.

**Author / principal investigator(s)**

Shisana, Olive (Prof O.)

**Acknowledgement**
**Funder(s)**
Centers for Disease Control and Prevention
Human Sciences Research Council
Nelson Mandela Foundation
Swiss Agency for Development and Cooperation

**Producer(s)**
Human Sciences Research Council

**Distributor(s)**
Human Sciences Research Council

**Date of distribution**
2011-12-13

**Access conditions**

By accessing the data, you give assurance that

- The data and documentation will not be duplicated, redistributed or sold without prior approval from the rights holder.
- The data will be used for scientific research or educational purposes only. The data will only be used for the specified purpose. If it is used for another purpose the additional purpose will be registered. Redundant data files will be destroyed.
- The confidentiality of individuals/organisations in the data will be preserved at all times. No attempt will be made to obtain or derive information from the data to identify individuals/organisations.
- The HSRC will be acknowledged in all published and unpublished works based on the data according to the provided citation.
- The HSRC will be informed of any books, articles, conference papers, theses, dissertations, reports or other publications resulting from work based in whole or in part on the data and documentation.
- For archiving and bibliographic purposes an electronic copy of all reports and publications based on the requested data will be sent to the HSRC.
- To offer for deposit into the HSRC Data Collection any new data sets which have been derived from or which have been created by the combination of the data supplied with other data.
- The data team bears no responsibility for use of the data or for interpretations or inferences based upon such uses.

Failure to comply with the End User License may result in sanctions being imposed.

**Keyword(s)**
BIIOGRAPHIC DATA / COMMUNITY PARTICIPATION / DISABILITIES / ELECTRICITY / GROUP MEMBERSHIP / HOUSEHOLD INCOME / RADIO / TELEPHONES / TELEVISION / TOILETS / WATER RESOURCES

**Topics(s)**
HIV AIDS