Data set metadata record

Data set ID
SABSSM 2008 Guardian Infant

Data set title
South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey (SABSSM) 2008: Guardian Infant - All provinces

Citation

Data set description

The guardian data of the SABSSM 2008 study covers information from the parents or care givers of children younger than 2 years on matters ranging from biographical information of the child and parent/guardian, the child's home environment, care and protection, sources of information on HIV and AIDS, media impact, history of health services, male circumcision and child health services.

The data set contains 161 variables and 1715 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 third in a series of household surveys conducted by Human Sciences Research Council (HSRC), that allow for tracking of HIV and associated determinants over time using a slightly same methodology used in 2002 and 2005 survey, making it the third national-level repeat survey. The 2002 and 2005 surveys included individuals aged 2+ years living in South Africa while 2008 survey included individuals of all ages living in South Africa, including infants younger than 2 years of age. The interval of three years since 2002 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 the prevalence of HIV infection in South Africa; examine the incidence of HIV infection in South Africa; assess the relationship between behavioural factors and HIV infection in South Africa; describe trends in HIV prevalence, HIV incidence, and risk behaviour in South Africa over the period 2002-2008; investigate the link between social, values, and cultural determinants and HIV infection in South Africa; assess the type and frequency of exposure to major national behavioural change communication programmes and assess their relationship to HIV prevention, AIDS treatment, care, and support; describe male circumcision practices in South Africa and assess its acceptability as a method of HIV prevention; collect data on the health conditions of South Africans; and contribute to the analysis of the impact of HIV/AIDS on society.

In the 13440 valid households or visiting points, 10856 agreed to participate in the survey, 23369 individuals (no more than 4 per household, including infants under 2 years) were eligible to be interviewed, and 20826 individuals completed the interview. Of the 23369 eligible individuals, 15031 agreed to provide a blood specimen for HIV testing and were anonymously linked to the behavioural questionnaires. the household
response rate was 80.8%, the individual response rate was 89.1% and the overall response rate for HIV testing was 64.3%.

**Time method**
Longitudinal: Trend/Repeated cross section

**Origin**
Primary data

**Granularity**
Micro level data

**Type of data**
Quantitative

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

**Production date**
2009

**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. Settlement types (Geotype): Urban formal, Urban informal, Tribal area (Rural informal), Rural formal (farm).

**Geographic unit**
Enumerator area (EA).

**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 of all ages in years, Africans, whites, coloured, Indians and others

**Universe - Included**
South African population of all individuals 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 and uniformed-service barracks were excluded.

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

Sampling procedure

As in previous surveys, a multi-stage disproportionate, stratified sampling approach was used. A total of 1 000 census enumeration areas (EAs) from the 2001 population census were selected from a database of 86 000 EAs and mapped in 2007 using aerial photography to create a new updated Master Sample as a basis for sampling visiting points/households. The selection of EAs was stratified by province and locality type. Locality types were identified as urban formal, urban informal, rural formal (including commercial farms), and rural informal. In the formal urban areas, race was also used as a third stratification variable (based on the predominant race group in the selected EA at the time of the 2001 census). The allocation of EAs to different stratification categories was disproportionate; that means, over-sampling or over-allocation of EAs was done, for example, in areas that were dominated by Indian, coloured or white race groups to ensure that the minimum required sample size in those smaller race groups was obtained.

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 preceded with the interview.

All people in the households, resident at the visiting point were initially listed, after which the eligible individual was randomly selected in each of the following three age groups: under 2 years, 2-14 years, 15-24 years and 25+ years. 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 0-11 years, parents/guardians were interviewed but DBS specimens were obtained from the children.

The sample size estimate for the 2008 survey was guided by the (1) requirement for measuring change over time in order to detect a change in HIV prevalence of 5 percentage points in each of the main reporting domains,
namely gender, age-group, race, locality type, and province (5% level of significance, 80% power, two-sided test), and (2) the requirement of an acceptable precision of estimates per reporting domain; that is, to be able to estimate HIV prevalence in each of the main reporting domains with a precision level of less than 4%, which is equivalent to the expected width of the 95% confidence interval (z-score at the 95% level for two-sided test). A design effect of 2 was assumed.

Overall, a total of 20826 interviewed participants composed of 4981 children (0-14 years), 5344 youths (15-24 years) and 10501 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 (68.9%) than males (62.02%) were tested for HIV. The 25+ years age group was the most compliant (68.8%), and 2-14 years the least (58.9%). The highest testing response rate was found in urban informal settlements (72.5%) and the lowest in urban formal areas (62.8%).

Weighting

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 0-2, 2-14, 15-24, and 25+ years). Only one person was randomly drawn in each of the 4 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.); Rehle, Thomas (Prof. TMG); Simbayi, Leickness (Prof. LC)

**Acknowledgement**
Gita Ramjee (Dr G); Parker (Dr W); Puren (Dr A)
Medical Research Council (MRC); CADRE; National Institute of Communicable Diseases; Reproductive Health Research Unit of the University of the Witwatersrand