South African national household survey of HIV/AIDS prevalence, behavioural risks and mass media impact — detailed methodology and response rate results

Olive Shisana, David Stoker, Leickness Chisamu Simbayi, Mark Orkin, Francois Bezuidenhout, Sean Edwin Jooste, Mark Colvin, Johan van Zyl

Objectives. To describe the methodology used in a recent survey of HIV/AIDS in South Africa and to present the response rates.

Methods. A cross-sectional, national household-based survey was conducted using second-generation surveillance procedures. A complex multistage sampling technique was used to create a master sample of 1 000 census enumerator areas out of a total of 86 000 nationally. Aerial photographs were taken and used to randomly select more than 10 197 households and ultimately 13 518 individuals from a sampling frame of 31 321 people. Phase 1 of the study involved notifying the household residents about the study and collecting key demographic information on respondents aged 2 years and older. This information was used to randomly select up to 3 respondents from each household: 1 adult (25 years and older), 1 youth (15 - 24 years), and 1 child (2 - 14 years). In phase 2 nurses interviewed respondents and collected oral fluid specimens for HIV testing. In the case of children aged 2 - 11 years, parents or guardians were interviewed, but HIV testing was performed on the selected children. Questionnaire data were anonymously linked with HIV test results.

Results. A total of 9 963 persons agreed to be interviewed and 8 840 were tested for HIV, yielding a response rate of 73.7% and 65.4% respectively. However, only 8 428 (62.3%) HIV test results were correctly matched with behavioural data. The results showed that those tested for HIV did not differ from those not tested in terms of key determinants.

Conclusion. It is possible to use community-based surveys to study the prevalence of HIV in the general population.

S Afr Med J 2004; **94:** 283-288.

In South Africa over the past decade, antenatal surveillance systems have constituted the primary basis for monitoring the spread of HIV. The strengths of using antenatal data are that is it relatively easy and inexpensive to obtain, it does not suffer from participation bias because consecutive patients are tested without individual consent, and the results are comparable between countries. However, the drawbacks of only using antenatal data are that assumptions must be made when extrapolating the findings from pregnant women using public sector health services to the whole population.

Human Sciences Research Council, Cape Town
Olive Shisana, ScD
Leickness Chisamu Simbayi, MSc, DPhil
Sean Edwin Jooste, MA

Private Consultant, PO Box 446, Newlands, Pretoria, 0049
D J Stoker, Math et Phys Dr

Human Sciences Research Council, Pretoria
Mark Orkin, MA, PhD
Johan van Zyl, BA (Hons)

Geospace International, Pretoria
Francois Bezuidenhout, BA (Hons)

Medical Research Council, Durban
Mark Colvin, MB ChB, MS

Several small-scale population-based HIV surveys have been conducted in South Africa.[1-3] However, there has not been a study on a large scale using a representative national sample. This study was undertaken to address this.

Method

Conceptual framework

The study was modelled on the second-generation surveillance system designed by the World Health Organisation (WHO), UNAIDS and Family Health International.[4,5] It involves conducting anonymous linked HIV testing in order to determine the relationship between HIV status and sociodemographic and behavioural practices. The purpose of conducting the study was to provide data to be used as a basis for designing interventions to prevent new infections and to forecast the likely development of the epidemic.

Sampling

The study population comprised all persons living in South Africa over 2 years of age and residing in homes, i.e. excluding individuals living in educational institutions, old age homes, hospitals and uniformed service barracks but including those living in hostels. A sample of 10 000 households was estimated to be appropriate to obtain reasonable estimates in the various

domains. The households comprised 1 200 Indian, 1 800 coloured, 2 200 white and 4 800 black households. Three persons in each household could potentially be selected, 1 from each of the following age groups: 2 - 14 years, 15 - 24 years and 25 years and older.

This is a probability sample of census enumeration areas (EAs) throughout South Africa, and representative of the country's population according to provincial distribution, settlement patterns and racial diversity.

To enable repeated household surveys, a master sample was created.[1] The sampling frame used in the design of the master sample was the enumerator areas (EAs) of the 2001 census carried out by Stats SA. The EAs constituted the primary sampling units (PSUs) in the master sample.

The sample was explicitly stratified by province and locality type of the EAs. Locality types were urban formal, urban informal, rural formal (including commercial farms) and tribal authority areas. In the formal urban areas, race was also used as a third stratification variable. The master sample therefore allowed for reporting of results at the level of province, type of locality, age and race group.

The sample was then drawn by selecting 1 000 EAs throughout South Africa. To obtain an approximately self-weighting sample of visiting points (VPs) within each of the explicit strata: (*i*) the EAs were drawn with a probability proportional to the size of the EA, using the 2001 census estimate of the number of VPs in the EA database as a measure of size (MOS); and (*ii*) an equal number of VPs ($N = 11$) were drawn from each PSU.

The population group allocation is only approximate since coloureds and whites will also be found on farms and coloureds in informal urban EAs.

To meet the criterion of having acceptable precision estimates by race group, the EA sample had to be allocated highly disproportionately to the explicit strata. This disproportionate allocation of the EA sample according to race resulted in a considerable overrepresentation of the 'urban formal' locality type, since the vast majority of Indians and whites live in formal urban areas. Furthermore, as Indians live primarily in KwaZulu-Natal and Gauteng and coloureds in the Western Cape this procedure resulted in a relative over-representation of these provinces, but this disproportionate sampling was later corrected by weighting the results.

Digital aerial photography

High-resolution, colour digital aerial photographs were obtained for 915 EAs. In the case of 44 EAs, digital aerial photographs had already been taken during the planning phase of the 2001 population census. The use of digital aerial photography for the dwelling sampling process and the field data collection was unique and greatly enhanced the validity and efficiency of the study. It enabled us to identify and select VPs in office using a desktop geographical information system (GIS) and hence field listing was effectively reduced to areas containing apartment blocks, hostels and large farms.

Field data collection methodology

Data for the study were collected in two phases. During phase 1, which lasted from March to August 2002, the selected households were notified about the survey.

During phase 1 the following information was assembled and collected for each EA.

1. A route description was provided by the first team, to assist later teams in navigating to a particular EA.

2. An orientation map of the area surrounding the EA.

3. An enumeration area map showing the selected VPs, cadastral data, prominent landmarks and road names, where available.

4. A photo sheet depicting the digital photos of the dwellings, where available.

5. A questionnaire was completed for each selected VP. This questionnaire contained details of household participation in the survey, reasons for non-participation, basic demographic information for the household(s), as well as global positioning system (GPS) co-ordinates for the specific VP. Later, the selected respondents were listed on this form.

6. An EA fieldwork form was completed containing information on the general attributes of the EA as well as locality information.

7. All the captured digital data were written onto compact disk (CD).

Besides the set of maps and photographs, the fieldworkers were also equipped with GPS units that were used to capture the geographical co-ordinates for each selected VP. The GPS also assisted fieldworkers to navigate to the selected EAs and VPs. This proved extremely useful, particularly in rural and farm areas. When fieldworkers did not find a person at a VP, two repeat visits were made before the VP was noted as a non-contact. Non-contacts were more likely to be in suburban areas. In cases where EAs were vacant, such as industrial areas and office buildings or where non-contacts were unacceptably high because of lack of access, these EAs were replaced with populated EAs of the same locality type within the same area. EAs were replaced if the cut-off point of six or more VPs agreeing to participate was not achieved.

In phase 2 of the study, which ran from April to September 2002, the actual HIV/AIDS survey was undertaken by re-visiting the VPs enumerated in phase 1. Retired, registered nurses were used as fieldworkers.

Phase 2 of the study had two main purposes. Firstly, to complete the relevant questionnaires and secondly to obtain

oral fluid specimens for HIV testing. Four different questionnaires were developed for each of the age groups interviewed: adults, youth, children aged 12 - 14 and parents/guardians of children aged 2 - 11 years. Parents/guardians answered for the children aged 2 - 11. The questionnaires for adults and youth, and to some extent for children aged 12 - 14, focused on the following issues: circumcision, marital status and marriage practices, sexual practices, perceived risk of HIV, voluntary counselling and testing, sexual violence, substance use, knowledge, attitudes and perceptions of HIV/AIDS, mass media and health status. In addition to the above and in common with the questionnaire for parents/guardians of children aged 2 - 11 years, the questionnaire for children aged 12 - 14 also looked at the home environment and care and protection. Another important issue covered by both questionnaires was educating children on life issues.

Where the suspicion arose that refusal rates were high owing to a mismatch between interviewers and respondents, revisits were made to these specific EAs, employing teams that matched the race and language composition of the area.

HIV testing

We used the OraSure HIV-1 oral specimen collection device in combination with the Vironostika HIV UNI-FORM II plus O enzyme-linked immunosorbent assay (ELISA) test kits to collect specimens for HIV testing. The OraSure/Vironostika testing methodology is the only methodology approved by the US Food and Drug Administration for use on oral fluids. According to the manufacturers the sensitivity and specificity of the OraSure device when tested with the Vironostika enzyme immunoassay are 99% and 99%.[6] The advantages of the OraSure device are that it is relatively easy to obtain specimens in young children and the specimen remains stable in ambient conditions for 21 days after specimen collection. All specimens were collected, transported, stored and tested according to detailed standard operating procedures.

Laboratory procedures

It was also important to ensure that HIV testing conducted by the three laboratories used was of the highest standard possible. In order to qualify to be part of the study, the two main criteria for selection were that the laboratories needed to have substantial experience in conducting HIV ELISA tests, and needed to do such tests as part of routine operations. Furthermore, laboratories needed to have acceptable internal and external quality control measures in place. One member of the consortium, namely the Medical Research Council (MRC) of South Africa, had extensive experience with HIV testing. The MRC was responsible for the accreditation status, quality control and audit procedures used by the laboratories.

In addition to the routine external and internal laboratory control measures, an additional quality control measure for the purposes of comparing inter-laboratory consistency was implemented specifically for the study. Thirty OraSure devices were taken by the MRC and each of 10 volunteers had 3 OraSure devices inserted, 1 after the other. The triplicate OraSure devices were sent to the 3 participating laboratories for a blind analysis. There was good agreement on the optical density (OD) ratios between the 3 laboratories.

Data capturing, data management, and data analysis

Before data entry started, a template was created for each questionnaire, which controlled the length and type of the fields, and linked each field to a validation programme to control the parameters of each field. The data were also double entered and verified.

Data analysis was performed initially using SPSS.[7] The final analysis was performed using SAS[8] or STATA 7.0.[9] Data were analysed with adjustment for clustering.

Overall quality control — re-interview of the sample

Upon completion of the fieldwork an evaluation survey was carried out to ascertain the overall compliance with procedures and the quality and accuracy of the findings.

Fifty EAs were selected from the original 1 000 EAs for inclusion in the evaluation survey, and 35 interviewers operating in teams carried out the evaluation survey.

Five distinct instruments were adapted for use in the re-interview survey. Evaluation instruments used in phase 1 included a questionnaire to validate the spatial location of the EAs and to check the accuracy of the maps and photographs provided to the field teams. For phase 2 evaluation, a shortened version of the adult, youth and caregiver questionnaires was used. The evaluation survey confirmed the validity of the information collected during both phases.

Weighting of sample

The following weighting procedures were undertaken before analysis of the data.[10-12] The data file of drawn EAs contained the selection probabilities as well as the sampling weights of these EAs. These weights reflected the disproportionate allocation of EAs to the strata. Where EAs were substituted it was necessary to calculate the sampling weight of those PSUs consisting of more than 1 EA.

Next, was to calculate the VP sampling weight. This weight was the counted number of VPs in the EA, proportionally corrected for invalid VPs, divided by the number of VPs participating in the survey. The final VP sampling weight was the product of the EA sampling weight and the VP sampling weight.

The next step in the weighting process was to assemble demographic information on all persons in all households at every participating VP in all responding EAs/PSUs, in order to calculate individual weights. In each of the 3 broad age groups the individual weight was the total number of individuals in that age group. This weight was adjusted for non-responding households.

In the final step the household information was integrated and the final sampling weight for each data record was calculated. This weight was equal to the final VP sampling weights multiplied by the selected person's sampling weight per VP per broad age group.

Ethical approval

Ethical approval for the study was obtained from the Interim Ethics Committee of the Human Sciences Research Council (HSRC). Written, informed consent was obtained directly from all adult respondents aged 15 years and older. In the case of children aged 2 - 14 years, consent was first sought from the parents and/or guardians, followed by secondary consent by the 12 - 14-year-olds, while in the case of children aged between 2 and 11 'consent' was sought to obtain an oral fluid specimen for HIV antibody testing. In order to compensate participants for their time, R50 (about US$5) and R20 (about US$2) were paid to the heads of households and single mine hostel dwellers respectively.

Pilot study

A pilot study that tested the entire study's procedures was conducted in 13 EAs in Gauteng and parts of neighbouring North West province. Among the valuable lessons learnt were the fact that the best times to visit households was during evenings and weekends, and that the questionnaire was a bit too long and therefore its length had to be reduced slightly.

Results

Response rate results are presented in Table I. At the 7 249 VPs where heads of households agreed to participate in the survey, 14 450 potential participants were selected after the completion of phase 1 (4 001 children, 3 720 youths and 6 729 adults). Of these, 13 518 (93.6%) were actually contacted during phase 2. A small proportion (6.4%) of potential respondents could not be approached because of logistical constraints such as inclement weather, impassable terrain or obstruction by 'gatekeepers'.

Of those who were contacted during phase 2, 9 963 (73.7%) agreed to be interviewed. Of the 9 963 interviewees, 8 840 agreed also to give an oral fluid specimen to be tested for HIV. However, only 8 428 specimens (62.3% of the 13 518 people in the original sample who were contacted) were anonymously linked to the behavioural interviews of the same individuals.

Table I shows the provincial breakdowns. Mpumalanga and Limpopo provinces had the lowest response rates for HIV testing, while the Northern Cape and Western Cape had the highest response rates. Regarding EA locality types, the highest response rate was in rural areas while the lowest was in urban formal areas. More females participated than males. With regard to race, the highest response rate was among coloureds and the lowest among whites. The most frequent reasons for participants refusing to take part in phase 2 of the study were,

Table I. Response rates for phases 1 and 2 of the study

Province	EAs realised in phase 1	EAs realised in phase 2		Total valid VPs in phase 1	VPs realised in phase 1	Selected respondents in phase 2		Realised respondents in phase 2		Respondents interviewed and tested		Respondents interviewed and not tested	All respondents interviewed	
	N	*N*	%	*N*	*N*	%	*N*	*N*	%	*N*	%	*N*	*N*	%
WC	125	125	100.0	1 363	890	65.3	1 806	1 809	100.1	1 267	70.0	56	1 323	73.1
EC	131	130	99.2	1 317	940	71.4	1 930	1 875	97.2	1 221	65.1	265	1 486	79.3
NC	76	73	96.1	759	545	71.8	992	913	92.0	694	76.0	35	729	79.8
FS	74	68	91.9	772	544	70.5	947	805	85.0	540	67.1	81	621	77.1
KZN	186	180	96.8	1 852	1 426	77.0	2 991	2 644	88.4	1 579	59.7	445	2 024	76.6
NW	74	73	98.7	747	581	77.8	1 086	1 042	95.9	626	60.1	110	736	70.6
GT	180	162	90.0	1 731	1 142	65.9	2 287	2 139	93.5	1 272	59.5	255	1 527	71.4
MP	74	72	97.3	769	558	72.6	1 092	1 030	94.3	550	53.4	70	620	60.2
LP	90	87	96.7	887	623	70.2	1 319	1 261	95.6	679	53.8	218	897	71.1
Total	1 010	970	96.0	10 197	7 249	71.1	14 450	13 518	93.6	8 428	62.3	1 535	9 963	73.7

WC = Western Cape; EC = Eastern Cape; NC = Northern Cape; FS = Free State; KZN = KwaZulu-Natal; NW = North West; GT = Gauteng; MP = Mpumalanga; LP = Limpopo Province.

for the most, similar to reasons advanced by heads of households when they declined to take part in phase 1. Only 1.3% (N = 182 of 13 518) and 0.6% (N = 82) of the respondents specifically mentioned that they were apprehensive of oral fluid being taken, and that it was against their religious beliefs to give oral fluids for testing, respectively.

Overall 16% (N = 1 692) of potential households refused to take part in phase 1. We also tried to understand refusal by tabulating the reasons given by those who refused to participate in phase 2. In some cases, respondents expressed apprehension regarding aspects relating to the HIV testing and to the topic of the survey (Table II).

When respondents who were interviewed and agreed to provide oral fluid specimens for HIV testing (N = 6 371) were compared with those who were interviewed but refused to provide oral fluid specimens (N = 713) on a few selected determinants, the two groups did not differ significantly on key variables such as use of condoms during last sexual intercourse, sexual activity during the last 12 months and sexually transmitted infections (STIs) in the last 3 months (Table III).

Discussion

This was the first-ever South African national, community-based study to simultaneously investigate sociodemographic variables human behaviour and HIV status. Using a community-based, probability sample is clearly advantageous in generating population-level estimates of HIV prevalence owing to inclusiveness of almost all age, sex, race and socioeconomic strata of society.

The sample size was large enough to allow for meaningful analysis of the data as a whole and broken down according to the chosen reporting domain. The fact that a master sample was created using the latest in GPS and aerial photography technology will allow for repeated surveys in the same household or in the same EA, to track changes in population behaviour, exposure to information on HIV prevention and HIV status.

However, because household HIV surveys rely on individual consent, they have lower response rates than antenatal surveys that do not seek consent of pregnant women and this may make the data from a survey requiring consent more subject to bias. Non-response during surveys is a well-documented phenomenon, especially when the research is dealing with a highly stigmatised issue such as HIV/AIDS. Many factors play a role in non-response, but in the South African situation additional factors may also have contributed to high non-response rates. Crime in South Africa is high enough to deter people from opening their doors to strangers, while the racial

Table II. Reasons provided by respondents for not participating in the survey

	Adult		Youth		Child		Guardian		Total	
Reasons for refusing	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Not willing to participate in any survey/interview	303	32.7	149	28.3	33	34.7	70	26.6	555	30.7
Not available now	165	17.8	154	29.2	22	23.1	38	14.4	379	20.9
Too busy to grant interview	122	13.2	38	7.2	5	5.2	30	11.4	195	10.8
Apprehensive of a saliva sample being taken	93	10.0	33	6.3	14	14.7	42	16.0	182	10.0
Objected to the topic of the survey (HIV/AIDS)	50	5.4	32	6.1	7	7.4	12	4.6	101	5.6
Against religious beliefs to provide a saliva sample	44	4.8	18	3.4	2	2.1	18	6.8	82	4.5
Did not want to know HIV status	34	3.7	19	3.6	4	4.2	8	3.0	65	3.6
Government not doing enough for him/her	27	2.9	14	2.6	1	1.1	5	1.9	47	2.6
Objected to being interviewed by the specific interviewer	7	0.8	23	4.4	1	1.1	2	0.8	33	1.8
Afraid	16	1.7	8	1.5	-	-	7	2.7	31	1.7
Unable to provide requested information	6	0.7	14	2.6	-	-	4	1.5	24	1.3
Too late in the evening	13	1.4	5	0.9	1	1.1	2	0.8	21	1.2
Did not trust the interviewers	9	1.0	3	0.6	2	2.1	5	1.9	19	1.1
Feared a breach of confidentiality	8	0.9	5	0.9	2	2.1	3	1.1	18	1.0
Other	4	0.4	1	0.2	-	-	11	4.1	16	0.9
Violence and gangsterism in the area	7	0.8	1	0.2	1	1.1	1	0.4	10	0.6
Refused to continue because respondent was in a hurry	3	0.3	3	0.6	-	-	2	0.8	8	0.4
Objected to providing personal/confidential information	5	0.5	1	0.2	-	-	-	-	6	0.3
Feared a breach of confidentiality	2	0.2	3	0.6	-	-	1	0.4	6	0.3
Refused to continue because respondent became irritated/bored	3	0.3	-	-	-	-	1	0.4	4	0.2
Objected to providing any/some information on the topic	2	0.2	1	0.2	-	-	1	0.4	4	0.2
Refused to continue because respondent became angry	1	0.1	1	0.2	-	-	-	-	2	0.1
Refused to continue because respondent lost interest/got tired	1	0.1	1	0.2	-	-	-	-	2	0.1
HIV test done previously	1	0.1	-	-	-	-	-	-	1	0.1
Total	926	100	527	100	95	100	263	100	1 811	100

Table III. Comparison between respondents who were interviewed and provided oral fluid specimens for HIV testing, and those who were interviewed only on a few selected detreminants

	Response frequencies				
	Questionnaire and test		Questionnaire only		
Variable	*N*	%	*N*	%	*P*-value
Condom use last intercourse					
Yes	1 310	20.6	136	19.1	0.4
No	5 061	79.4	577	80.9	
Total	6 371	100%	713	100%	
Sexual activity in last 12 months					
Virgin	1 033	16.2	128	18.0	0.2
Abstinent	1 242	19.5	126	17.7	
One partner	3 746	58.8	430	60.3	
Multiple partners	350	5.5	29	4.1	
Total	6 371	100%	713	100%	
STI in the last 3 months					
Yes	126	2.1	12	1.7	0.2
No	5 977	97.9	681	98.3	
Total	6 103	100%	693	100%	

profile of interviewers, because of our apartheid history, probably influenced the decision of some respondents not to participate.

Conclusion

We hope that HIV/AIDS researchers in other countries with a great need for detailed research and evidence-based information to guide intervention on the HIV/AIDS epidemics affecting their countries, as is the case in South Africa, will be encouraged to undertake similar studies in their respective countries in the near future. Clearly researchers in other countries who are planning to undertake similar studies will need to adapt these procedures according to their own unique situations. In particular, sampling will be much easier if a more homogeneous population is involved. Similarly, the corresponding response rates may be higher in those countries where racial and political dynamics in general and the HIV/AIDS issue in particular are not as complex as in South Africa.

This study was funded by the Nelson Mandela Foundation, the Nelson Mandela Children's Fund, the Human Sciences Research Council and the Swiss Agency for Development and Cooperation.

References

1. Auvert B, Bollard R, Campbell C, *et al.* HIV infection among youth in a South African mining town is associated with herpes simplex virus-2 seropositivity and sexual behaviour. *AIDS* 2001, **15:** 5885-5989.
2. Colvin M, Abdool Karim SS, Connolly C, Hoosen AA, Ntuli N. HIV infection and asymptomatic sexually transmitted infections in a rural South African community. *Int J STD AIDS* 1998; **9:** 548-550.
3. MacPhail C, Williams B, Campbell C. Relative risk of HIV infection among young men and women in a South African township. *Int J STD AIDS* 2002, **13:** 331-342.
4. Family Health International. *Behavioural Surveillance Surveys: Guidelines for Repeated Behavioural Surveys in Populations at Risk of HIV.* FHI, 2000.
5. WHO/UNAIDS. *Guidelines for Second Generation HIV Surveillance* 2000.
6. Gallo D, George J, Fitchen J, Goldstein A, Hindahl M. Evaluation of a system using oral mucosal transudate for HIV-1 antibody screening and confirmatory testing. OraSure HIV Clinical Trials Group. *JAMA* 1997, **277:** 254-258.
7. SPSS Inc. Chicago, Illinois, USA, 1997.
8. SAS Inc. Cary, NC, USA, 1999.
9. Stata Corp. College Station. Texas, USA, 2002.
10. Kalton G. *Compensating for Missing Survey Data.* Research Report Series. Ann Arbor: Survey Research Centre, Institute for Social Research, University of Michigan, 1983.
11. Lehtonen R, Pahkinen EJ. *Practical Methods for Design and Analysis of Complex Surveys.* New York: John Wiley and Sons 1995.
12. Stoker, DJ. *The Analysis of Complex Sample Data.* Pretoria: Human Sciences Research Council, 1982.

Accepted 12 December 2003.