The validity of self-reported HIV infection among the general population in rural Malawi

Simona Bignami-Van Assche, Ph.D., *Harvard University*

Li-Wei Chao, Ph.D., *University of Pennsylvania*

Philip Anglewicz, M.Sc., *University of Pennsylvania*

Paper to be presented at the 2005 Meeting of the International Union for the Study of Population (IUSSP), 18-23 July, Tours, France

Corresponding author: Simona Bignami-Van Assche, Harvard University Initiative for Global Health, 104 Mt Auburn Street, Cambridge, MA 02138. Phone: (617) 495-7773; Fax: (617) 495-8231; Email: simona_bignami@harvard.edu.

Abstract

This paper examines the validity of self-reported human immunodeficiency virus (HIV) infection among the general population in rural Malawi. The data for this study come from the 2004 Malawi Diffusion and Ideational Change Project, a panel household survey that collected both behavioral and biomarker data for a sample of approximately 3000 respondents. To investigate the validity of self-reported HIV infection, the respondents' self-assessed likelihood of having the infection at the time of the survey is compared with their actual HIV status determined by antibody test. The relationship between the validity of self-reports and variables such as the respondents' background characteristics and their perceptions of their own community's HIV prevalence is also examined.

Keywords: HIV; self-reports; antibody testing; developing countries; Malawi

INTRODUCTION

Research on the occurrence and correlates of human immunodeficiency virus (HIV) infection should ideally be based on laboratory-confirmed antibody test results. In most developing countries, however, laboratory confirmation is not generally possible (especially in rural areas) and only self-reports of symptoms or perceived risk are available.[1] The decision of whether to use these reports to measure HIV prevalence and design interventions in developing countries rests on the extent of the bias self-reports will introduce in the population under study.

The literature on biases in self-reports of HIV infection is limited. Evidence from developed-country studies among select populations (i.e. those with HIV infection or at high risk for HIV infection, such as drug users, prostitutes and prisoners) suggests that concurrence between individuals' self-reports of current HIV status and their HIV test results is high (95-99%) for seronegative individuals, but low (40-70%) for seropositive individuals (**Fennema et al. 1995; Thornton et al. 2000; Harrington et al. 2001; Strauss et al. 2001**). For developing countries, one study found a high (60-70%) rate of incorrect self-reports of HIV status among a sample of 122 attendees of a voluntary HIV testing centre in Lusaka, and raised concerns about the accuracy of self-assessed HIV infection among the Zambian population (**Chintu et al. 1997**). On the contrary, a case-control study in Tanzania found no significant association between perceived risk of infection and actual status for both men and women (**Quigley et al. 1997**). Since recent population-based surveys that include HIV testing (such as the Demographic and Health Survey) and national HIV surveys do not collect

[1] The United Nations Group on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) estimate that in low and middle-income countries only 10% of people at risk of HIV infection have access to voluntary counseling and testing (**UNAIDS/WHO 2004**).

information on self-reported *current* HIV infection[2], it is very difficult to reach a clear-cut conclusion on the validity of HIV status derived solely from survey information in developing countries, especially for the general population.

This paper examines the validity of self-reported HIV infection among the general population in rural Malawi, by comparing actual HIV antibody status with self-reported likelihood of currently having the infection for a sample of approximately 3000 respondents. The relationship between the validity of self-reports and variables such as the respondents' background characteristics and their perceptions of their own community's HIV prevalence is also examined.

MATERIALS AND METHODS

Study population

Since 1998, the Malawi Diffusion and Ideational Change Project (MDICP) has collected longitudinal data from a population-based sample of approximately 2500 adult respondents to examine the role of social networks in changing attitudes and behavior regarding HIV/AIDS, family size, and family planning in rural Malawi. The MDICP is conducted in rural areas of three Malawian districts, one in each of the three regions of the country (North, Center, and South). A comparison of the characteristics of the MDICP sample with those of the rural population surveyed in the Malawi Demographic and Health Survey indicates that the MDICP

[2] Population based surveys, such as the Demographic and Health Survey, the 2001 Nelson Mandela/HSRC Study of HIV/AIDS (**Shishana and Simbayi 2002**), the 2003 South Africa National Survey of 15-24 Year Olds (**Pettifor et al 2004**), and a risk perception survey in rural south-west Uganda (**Kengeya-Kayondo et al. 1999**), ask respondents to evaluate their risk of becoming infected with HIV or of getting AIDS in the future, not their current likelihood of being already infected with HIV.

sample is representative of the national rural population (more details on sampling and fieldwork procedures, as well as the survey data, are available for free download from the project's website: http://malawi.pop.upenn.edu).

The first round of the survey (MDICP-1) was carried out in mid-1998, and interviewed 1541 ever-married women of childbearing age and 1065 husbands of the currently married women. In 2001, the second round of the survey (MDICP-2) followed up the same respondents (if present and still eligible), and also interviewed their new spouses if they had divorced and remarried between the two survey waves. The third wave of the survey (MDICP-3), carried out in mid-2004 with the original adult sample and an additional sample of approximately 1500 adolescents, included for the first time a testing component for HIV **(Bignami-Van Assche et al. 2004)**. A group of trained nurses was responsible for approaching MDICP respondents and administering a short questionnaire (henceforth referred to as 'STI questionnaire') on health, sexually transmitted infections (STIs), and HIV knowledge and risk perception. After completing the STI questionnaire, the nurses provided an extensive explanation of the HIV testing process (as most respondents had never been tested for HIV[3]), and requested the respondent's consent to be tested. Respondents who agreed to be tested were provided counseling, after which saliva samples for HIV testing were collected using OraSure® oral swabs[4] (OraSure Technologies Inc, Bethlehem, Pennsylvania,

[3] According to the nationally-representative 2000 Malawi Demographic and Health Survey, 93% of the rural population has never been tested for HIV (**National Statistical Office and ORC Macro 2001**).

[4] The accuracy of saliva for detection of HIV antibodies has been shown to be comparable to serum-based tests (**Holm-Hansen and Constantine 1993; Tamashiro and Constantine 1994; Bruckosa and Stankova 1995; Vall-Mayans et al. 1995; Luo et al. 1995; Frangos-Girard et al. 1996**). For epidemiological purposes in particular, use of saliva seems to offer several advantages over serum. An important motivation for using saliva in population-based surveys is the assumption that a non-invasive method might contribute significantly in reducing selection bias due to non-consent. Studies that have used saliva for detection of HIV antibodies have generally achieved higher consent rates, but data are still lacking to make a sound evaluation of the ways in which saliva and serum compete with regard to acceptability (**Filkenes et al. 1998**).

USA). All specimens were analyzed at the laboratory of the University of North Carolina Project in the capital, Lilongwe. HIV-1 status was ascertained using enzyme-linked immunosorbent assays (ELISA) kits for initial screening and a confirmatory Western Blot test. The survey and biomarker collection protocols were approved both by the Institutional Review Board of the University of Pennsylvania in the United States and by the Research and Ethics Committee of the College of Medicine in Malawi.

Between April and December 2004, the MDICP nurses tested 2907 respondents for HIV. The refusal rates were overall quite low (approximately 10%). The results show that HIV prevalence is approximately 7% for the entire sample, varying from 4.5% in the Northern region to 8.4% in the Southern region (Thornton et al, Reactions to Voluntary Counseling and Testing in Rural Malawi; 2005).

A complete STI questionnaire is available for 2844 (98%) of all MDICP-3 respondents tested for HIV, who represent the sample for the present analysis. Descriptive statistics are summarized in Table 1.

Table 1 about here

Data analysis

The validity of self-reported HIV infection was evaluated by comparing the likelihood of infection reported by the respondent in the STI questionnaire with his/her antibody test result. To this end, we computed standard epidemiological measures used to describe the accuracy of diagnostic tests (sensitivity, specificity, and positive and negative predictive value). The “diagnostic test” in this study refers to the self-reported likelihood of current HIV infection, and the “gold standard” refers to the HIV antibody test.

Sensitivity for self-reported HIV infection was computed as the number of respondents with positive antibody test who also reported having some likelihood (low, medium, or high) of being infected with HIV at the time of the survey divided by the total number of respondents who tested positive for HIV on the antibody test. *Specificity* was computed as the number of respondents with negative antibody test who also reported having no likelihood of being infected with HIV at the time of the survey divided by the total number of respondents who tested negative for HIV. The *positive predictive value* was calculated as the number of respondents with positive antibody test who also reported having some likelihood (low, medium, or high) of being infected with HIV at the time of the survey divided by the total number of respondents who reported some likelihood of being currently infected with HIV. The *negative predictive value* was calculated as the number of respondents with negative antibody test who also reported having no likelihood of being infected with HIV at the time of the survey divided by the total number of respondents who reported no likelihood of being currently infected with HIV.

To determine whether respondents’ background characteristics, self-reported health, and perceived HIV prevalence in the community affect the validity of self-reports, we performed multivariate logistic regression analysis to identify covariates of respondents’ incorrect assessment of their current HIV status. First, among those who think they are infected with HIV, we examined the characteristics of those who are more likely to be actually not infected (false positives) than infected (true positives). Second, among those who think they are not infected with HIV, we examined the characteristics of those who are more likely to be actually infected (false negatives) than not infected (true negatives). Two models were fitted to the data, separately for males and females: one model including only the respondent's background characteristics (age, region, marital status and self-reported health) and another model including respondent's background characteristics as well as his/her perception of HIV prevalence in the community.[5] Odds ratios and 95 percent confidence intervals were calculated using STATA (**Stata Corporation 2004**).

RESULTS

A total of 2825 (99%) respondents in the selected sample gave a non-missing response to the question 'In your opinion, what is the likelihood (chance) that you are infected with HIV now?' Of these, 18% (498) reported they did not know, 59% (1677) reported that there was no chance they were infected, and 23% (650) reported that there was some—low, medium, or high—chance (of these, 5% reported that there was a high chance).

Table 2 about here

Among the 2327 respondents who gave an estimate of their current likelihood of infection (i.e. who did not answer "don't know" to the self-assessed likelihood question above), 71% (1646) accurately assessed their HIV status (Table 2). The majority of those assessments that were incorrect (88%) was due to respondents over- rather than under-estimating their chance

[5] Perceived HIV prevalence in the community was measured by using the respondent's answer to the question "If we took a group of 10 people from this area—just normal people who you found working in the fields or in homes—how many of them do you think would now have HIV?" Answers to this question were on a continuous scale from 1 to 10, but in the multivariate regression analysis they were dichotomized into 0-50% and 50-100% (this dichotomization was chosen in order to maximize sample sizes for the analysis).

of being infected with HIV. Of 1677 respondents who reported no likelihood of being currently infected with HIV, only 5% (84) tested positive, giving a negative predictive value (i.e. the probability of not having HIV among respondents who reported no likelihood of infection) of 95%. Of 650 respondents who reported some chance of being infected (low, medium, or high), 91% (597) tested negative for HIV, giving a positive predictive value (i.e. the probability of having HIV among respondents who reported some likelihood of infection) of 9%. Self-reported HIV status as a “diagnostic test” for HIV infection has a fairly low sensitivity (39%) and not a very high specificity (73%). In other words, among those who are truly HIV infected (as determined by the HIV antibody test) only 39% actually reported that they were infected. Women are less likely than men to correctly assess their HIV status, and this is mostly because women overestimate their risk *more* and underestimate their risk *less* than their male counterparts. Self-reported HIV infection as a diagnostic test, thus, tends to have a lower specificity but higher sensitivity among women (66% and 46%, respectively) than men (80% and 27%, respectively).

Table 3 about here

As respondents tend to overestimate their likelihood of being currently infected with HIV, aggregate levels of self-reported prevalence of HIV infection are 3-8 times higher than that found by HIV antibody testing among respondents who gave an estimate of their current likelihood of infection (Table 3). The ratio between self-reported and actual HIV prevalence is inversely related to HIV prevalence in each region, being highest in the North and lowest in the South.

Table 4 about here

Self-reports are more valid for respondents who tested negative for HIV antibodies than respondents who tested positive (Table 4). Of the 2651 respondents whose HIV test was negative, 60% (1593) had reported having no likelihood of being currently infected with HIV and 17% (461) had reported that they did not know or were uncertain of their HIV status. In contrast, of the 193 respondents who tested positive on the HIV antibody test, 27.5% (53) had reported having some likelihood of being currently infected with HIV (only 7% had reported a high likelihood), and 30% (56) had been uncertain. Overall, HIV prevalence by antibody test was highest among respondents who reported they did not or were uncertain of their HIV status, especially for females.

Table 5 about here

In addition, Table 5 shows that the difference in self-reported likelihood of infection between HIV positive and HIV negative respondents is significant (p=.004), and robust to different dichotomizations of the categorical responses to the self-reported likelihood question. There is also an interaction between HIV status and gender. Among respondents who tested positive on the HIV antibody test, males were less accurate in predicting their test result than females. In contrast, among respondents who had a negative test, males were more accurate than females. However, differences in self-reported likelihood of HIV infection between antibody test-positive and test-negative respondents are significant only for females: males who tested positive are not statistically significantly more likely to think they are infected than males who tested negative, whereas females who tested positive are statistically significantly more likely to think they are infected than females who tested negative.

Table 6 about here

Table 6 compares background characteristics of false positives and true positives, and of false negatives and true negatives. Both male and female respondents who tested positive and accurately reported their current likelihood of HIV infection (true positives) were older, more likely to be married and rated themselves in worse health compared to respondents who reported some likelihood of being currently infected with HIV but whose HIV test was negative (false positives). For both males and females, respondents with negative antibody test who accurately reported their current likelihood of HIV infection (true negatives) were younger, more likely to be unmarried and not to have a partner, rated themselves in better health, and thought that more people in their communities had HIV, compared to respondents who reported no likelihood of being currently infected with HIV but tested positive for HIV (false negatives). In addition, for males, false positives tend to be younger, less likely to be married and reported themselves in better health (excellent, or very good) than false negatives. For females, there is no clear age pattern, but the other patterns tend to be opposite those of males: false positives are more likely to be married and to rate themselves in worse health than false negatives. For both males and females, false positives are more likely than false negatives to perceive a higher HIV prevalence in their communities.

Table 7 about here

In the logistic regression analysis comparing the probability of false positive responses to that of true positive responses (Table 7), the only two significant covariates for females are marital status and self-reported health. Married women are 3.6 times more likely than unmarried ones to report some likelihood of HIV infection but to test negative (95 percent confidence interval: 1.37, 9.46). In other words, among women who believe that they are infected with HIV, married respondents are 3.6 times more likely to be incorrect in their beliefs than unmarried women. Among women who think they are infected with HIV, those who also report a poor health status are only one-fifth as likely to be actually infected than those who think they have excellent health (odds ratio = 0.20; 95 percent confidence interval: 0.07, 0.60). For males, there is a strong region effect (males residing in the Northern region are 5.3 times more likely than those in the Southern region to overstate their likelihood of HIV infection). In addition, there is a small but significant age effect and a very large effect of marital status (married men are 33.7 times more likely than unmarried men to believe they are infected with HIV but to test negative for HIV). For both males and females, perceived community HIV prevalence is not a significant predictor of false positive responses.

Table 8 about here

In the multivariate logistic regression analysis comparing the probability of false negative responses relative to that of true negative responses (Table 8), for both males and females region and age are significantly associated with the underreporting of HIV infection: respondents living in the Northern region are 2.3 times less likely to understate their likelihood of HIV infection than respondents living in the Southern region, and older respondents are 3-5 times more likely to understate it than younger ones. Males who reported good health are also more likely than those who report excellent health to give a false negative response (odds ratio = 3.46, 95 percent confidence interval: 1.45, 8.29).

DISCUSSION

To our knowledge, this is the first study to use population-based results of HIV testing to validate self-reported *current* HIV infection. Because of the extensive and rich data collected by the MDICP, we were able to use a variety of measures to investigate the accuracy of positive and negative reports of HIV at different levels of aggregation. We were also able to assess gender differences in the accuracy of self-reported likelihood of HIV infection. This is particularly important because in Malawi, as in other highly-HIV affected countries, men and women have different socio-demographic characteristics and exposure risks to HIV.

Our main finding is that respondents who estimated their likelihood of current HIV infection were generally accurate in their assessment, with males being more accurate than females. When they were inaccurate, it was primarily because respondents thought they were HIV positive but were, in fact, HIV negative: false positives constitute 88% of all inaccurate self-assessed infections. This result is consistent with the MDICP survey data, which show that respondents vastly overestimate the transmission probabilities of HIV: over 90% of 2001 MDICP respondents believe that HIV transmission is certain or highly likely from a single unprotected act of sexual intercourse with a HIV-infected person.

Our finding that HIV positive respondents are significantly less likely than HIV negative respondents to predict their results correctly confirms what other studies have found in high-risk populations in developed countries. Our study makes a unique contribution by finding that among those who held incorrect beliefs about their own HIV infection status, the majority was due to overestimation of their own likelihood of infection. Among respondents who are not infected, women tend to overestimate their own risk of infection more than men. However, among respondents who are infected, women are better at predicting their own status, as they tend to underestimate their likelihood of infection less than men. Nevertheless, women in rural Malawi are overall less accurate than men in predicting their own HIV infection status.

Another interesting finding is the large disparity in the number of “don’t know” responses to the self-assessed HIV infection risk question given by HIV-positive and HIV-negative respondents. Among HIV negative respondents, approximately 15-20% said they were uncertain about their current likelihood of HIV infection; whereas among HIV positive respondents this proportion is about 30-50%. In addition, overall HIV prevalence is highest among respondents who reported that they did not know or where uncertain about their HIV status. Combined with the low proportion of false negative responses, this finding suggests that respondents (and especially those who were HIV positive) may have given answers affected by social desirability bias. In other terms, respondents who had been promiscuous and thus thought they had a high chance of being infected with HIV may have resented the invasion of privacy by an unknown interviewer or feared stigma associated with admitting that they might have the infection.

Our conclusion is that although many respondents correctly evaluate their likelihood of HIV infection, reliance on self-reports may introduce misclassification bias and potentially lead to false conclusions regarding the efficacy of prevention interventions and behavioral change in the general population. Since respondents more frequently overestimate their likelihood of HIV infection than they underestimate it, HIV prevalence calculated on the basis of self-reports is considerably higher than actual HIV prevalence found by antibody testing. As inferences drawn from self-reports are potentially misleading, such data need to be validated and biologic indicators should be preferred as outcome measures.

The rate of incorrect self-reports of HIV status also raises concerns about the perception of Malawians with respect to the epidemiology of HIV infection. A possibility that deserves attention is that individuals with treatable medical conditions (such as tuberculosis) who perceive themselves as having HIV infection may fail to seek medical attention because they think that they have an untreatable disease. This highlights the importance of improving and expanding access to HIV testing, especially in rural areas.

ACKNOWLEDGEMENTS

The Malawi Diffusion and Ideational Change Project is supported by grants from the Rockefeller Foundation; NICHD (R01-HD4173, R01 HD372-276); NIA (AG1236-S3); and the Center for AIDS Research and the Center on the Demography of Aging, both at the University of Pennsylvania.

REFERENCES

UNAIDS/WHO. UNAIDS/WHO Policy on HIV Testing [*UNAIDS* Web site]. June 2004. Available at: http://www.unaids.org/html/pub/una-docs/hivtestingpolicy_en_pdf.pdf. Accessed May 23, 2005.

Fennema JS, van Ameijden EJ, Coutinho RA, van den Hoek JA. Validity of self-reported sexually transmitted diseases in a cohort of drug-using prostitutes in Amsterdam: trends from 1986 to 1992. *Int J Epidemiol* 1995; 5:1034-41.

Thornton L, Barry J, Long J, et al. Comparison between self-reported hepatitis B, hepatitis C, and HIV antibody status and oral fluid assay results in Irish prisoners. *Commun Dis Public Health* 2000; 3:253-5.

Harrington KF, DiClemente RJ, Wingood G, et al. Validity of Self-reported Sexually Transmitted Diseases Among African American Female Adolescents Participating in an HIV/STD Prevention Intervention Trial. *Sexually Trans Dis* 2001; 28:468-71.

Strauss SM, Rindskopf DM, Deren S, Falkin GP. Concurrence of drug users' self-report of current HIV status and serotest results. *J Acquir Immune Defic Syndr* 2001; 27:301-7.

Chintu C, Baboo KS, Gould SS, et al. False-positive self-reports of HIV infection. *Lancet*, 1997; 349:650.

Quigley M, Munguti K, Grosskurth H, et al. Sexual behavior patterns and other risk factors for HIV infection in rural Tanzania: A case-control study. *AIDS*, 1997; 237-248.

Shisana O, Simbayi L. *Nelson Mandela/HSRC study of HIV/AIDS: South African national HIV prevalence, behavioral risks and mass media. Household Survey 2002*. Cape Town: Human Science Research Council; 2002.

Pettifor AE, Rees HV, Steffenson A, Hlongwa-Madikizela L, MacPhail C, Vermaak K, Kleinschmidt I. *HIV and sexual behavior among young South Africans: a national survey of 15-24 year olds*. Johannesburg: reproductive Health Research Unit, University of the Witswatersrand; 2004.

Kengeya-Kayondo JF, Carpenter LM, Kintu PM, et al. Risk perception and HIV-1 prevalence in 15,000 adults in rural south-west Uganda. *AIDS*, 1999; 2295-302.

STATA [computer program]. Version 8.2. College Station, TX: Stata Corporation; 2004

Bignami-Van Assche S, Smith KP, Reniers G, et al. *Protocol for biomarker testing in the 2004 Malawi Diffusion and Ideational Change Project 2004*. Philadelphia, PA: SNP Working Paper No.6; 2004.

National Statistical Office [Malawi], ORC Macro. *Malawi Demographic and Health Survey 2000*. Zomba, Malawi and Calverton, Maryland, USA: National Statistical Office and ORC Macro; 2001.

Holm-Hansen C, Constantine NT. Saliva and HIV testing [letter]. *Lancet* 1993; 341:382-383.

Tamashiro H, Constantine NT. Serological diagnosis of HIV infection using oral fluid samples. *Bull World Health Organ* 1994; 72:135-143.

Bruckova M, Stankova M. Detection of HIV virus antibodies in saliva. *Epidemiol Mikrobiol Imunol* 1995; 44:127-129.

Vall-Mayans M, Casabona J, Rabella N, De Miniac D. Testing of saliva and serum for HIV in high-risk populations. Ad hoc group for the comparative saliva and serum study. *Eur J Clin Microbiol Infect Dis* 1995; 14:710-713.

Luo NP, Kasolo F, Ngwenia B, du Pont HL, Zumla A. Use of saliva as an alternative to serum for HIV screening in Africa. *South Afr Med J* 1995; 85:156-157.

Frangois-Girard C, Thortensson R, Luton P, et al. Multi-center European evaluation of HIV testing on serum and saliva samples. *Transfus Clin Biol* 1996; 3:89-98.

Filkenes K, Ndhlovu Z, Kasumba K, Mubanga Musonda R, Sichone M. Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia. *AIDS* 1998; 12:1227-1234.

Tables and figures

Table 1: Descriptive statistics of selected sample, 2004 MDICP

	Males	Females
HIV prevalence	5.7	7.7
Region		
South	36.5	37.9
Center	30.3	28.0
North	33.3	34.1
Age		
Less than 15	1.1	1.1
15-19	18.8	18.2
20-24	17.4	14.6
25-29	7.8	13.0
30-34	10.1	12.4
35-39	8.9	12.0
40-44	8.9	9.6
45-49	6.3	6.6
50-54	8.9	5.3
55+	8.3	3.0
Currently married	68.6	73.5
Self-reported health		
Excellent	40.8	32.2
Very good	18.7	20.0
Good	27.5	31.6
Fair	11.1	13.7
Poor	1.0	0.9
Suspect or knows that spouse/partner has HIV		
Yes, know	0.5	0.7
Suspect	3.9	11.7
Probably not	48.4	40.6
Don't know	24.8	29.1
Does not have partner	20.9	16.1
Perceived HIV prevalence in the community		
Mean number of people (out of 10) in the community the respondent thinks have HIV	3.9	4.3
Don't know	15.1	20.3
N	1308	1536

Percentages might not add up to 100 because of missing values.

Table 2: Proportion accurate, false negatives, false positives, positive predictive value (sample sizes in parentheses), negative predictive value, and sensitivity and specificity of self-reported likelihood of current HIV infection, by sex and region*

	% accurate	FN	FP	PPV	NPV	Sensitivity	Specificity
Males	77.5 (1072)	4.4 (855)	93.5 (217)	6.5	95.6	26.9	80.1
Females	64.9 (1255)	5.6 (822)	91.0 (433)	9.0	94.4	45.9	66.3
Both sexes	70.7 (2327)	5.0 (1677)	91.0 (650)	9.0	95.0	38.7	72.7

* Respondents who answered "don't know" to the question on self-assessed likelihood of HIV infection are excluded.

Legend: FN: false negatives; FP: false positives; PPV: positive predictive value; NPV: negative predictive value.

Table 3: Comparison between HIV prevalence found by antibody testing and by self-reported HIV infection for respondents who evaluated their likelihood of current HIV infection, by sex and region*

	Prevalence by HIV test (%)	Prevalence by Self-reported infection (%)	N
BOTH SEXES			
South	7.5	29.5	861
Center	6.9	31.8	680
North	3.4	23.3	786
Total	6.0	28.1	2327
MALES			
	Prevalence by HIV test (%)	Prevalence by Self-reported infection (%)	N
South	6.4	25.4	393
Center	6.1	16.2	314
North	2.2	18.1	365
Total	4.9	20.2	1072
FEMALES			
	Prevalence by HIV test (%)	Prevalence by Self-reported infection (%)	N
South	8.3	32.9	468
Center	7.4	45.1	366
North	4.5	27.8	421
Total	6.8	34.7	1255

* Respondents who answered "don't know" to the question on self-assessed likelihood of HIV infection are excluded.

Note: Prevalence based on self-reported HIV infection is calculated by dichotomizing the categorical responses to the self-reported likelihood question into 'no likelihood' and 'some likelihood' (low, medium, or high).

Table 4: HIV positive respondents, HIV negative respondents and actual HIV prevalence from antibody test, by self-assessed likelihood of current HIV infection and sex

	No likelihood	Low	Medium	High	Don't know	N
HIV POSITIVE						
Males	52.8	11.1	5.6	2.8	27.8	72
Females	38.7	16.0	7.6	9.2	28.6	119
Both sexes	44.0	14.1	6.8	6.8	28.3	191
HIV NEGATIVE						
	No likelihood	Low	Medium	High	Don't know	N
Males	66.6	9.0	4.2	3.3	16.9	1227
Females	55.2	13.4	8.1	6.5	16.8	1407
Both sexes	60.5	11.4	6.3	5.0	16.9	2634
HIV PREVALENCE by HIV TEST						
	No likelihood	Low	Medium	High	Don't know	N
Males	4.4	6.7	7.1	4.8	8.8	1299
Females	5.6	9.2	7.3	10.7	12.5	1526
Both sexes	5.0	8.0	7.2	7.7	10.7	2825

Table 5: Pearson chi-square tests for differences in self-assessed likelihood of current HIV infection between HIV positive and HIV negative respondents, by three different dichotomizations of categorical responses to the self-assessed likelihood question

	No vs. some likelihood		No/low vs. medium/high		No vs. high likelihood	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
Males	1.51	.219	0.01	.922	0.38	.539
Females	5.23	.022	4.09	.043	1.88	.170
Both sexes	8.36	.004	4.14	.042	3.10	.078

Table 6: False negatives, true negatives, false positives and true positives, by sex and background characteristics

	False negatives		True negatives		False positives		True positives	
	M	F	M	F	M	F	M	F
Region								
South	42.1	45.7	33.9	37.9	44.8	34.3	64.3	46.2
Center	42.1	34.8	30.2	24.0	23.6	38.8	21.4	28.2
North	15.8	19.6	35.9	38.1	31.5	26.9	14.3	25.6
Age								
Less than 25	15.8	23.9	43.0	41.9	29.1	26.1	7.1	30.8
25-29	7.9	10.9	7.3	11.9	9.4	15.5	7.1	15.4
30-34	13.2	21.7	10.4	9.4	10.8	15.7	14.3	15.4
35-39	21.1	10.9	8.4	9.3	7.4	14.2	21.4	17.9
40-44	15.8	19.6	6.7	8.6	8.4	11.2	28.6	12.8
45+	26.3	10.9	20.7	14.7	30.5	13.2	21.4	7.7
Currently married	86.8	67.4	62.9	67.1	81.8	83.5	85.7	76.9
Self-reported health								
Excellent	26.3	30.4	45.2	37.2	31.0	25.6	21.4	12.8
Very good	7.9	26.1	18.6	20.7	20.7	17.8	21.4	30.8
Good	50.0	30.4	25.9	29.1	27.6	35.8	28.6	17.9
Fair	15.8	10.9	9.2	11.0	17.2	18.3	28.6	30.8
Poor	0.0	0.0	0.5	0.8	2.0	0.5	0.0	7.7
Perceived HIV prevalence in the community								
Mean number of people (out of 10) in the community the respondent thinks have HIV	3.8	3.7	5.4	5.7	4.2	4.4	5.9	6.0
Don’t know	10.5	17.4	15.8	19.7	6.9	18.8	7.1	10.3
N	38	46	817	776	203	394	14	39

Note: Percentages might not add up to 100 because of missing values.

Legend: M: Males; F: Females

Table 7: Multivariate logistic regression analysis of false positives

	MALES				FEMALES			
	Model 1		Model 2		Model 1		Model 2	
Variable	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Region								
South	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Center	2.18	0.47, 10.00	1.90	0.39, 9.31	1.89	0.82, 4.36	1.71	0.67, 4.41
North	5.95*	0.92, 38.64	5.30*	0.77, 36.48	1.23	0.52, 2.94	1.16	0.48, 2.85
Age								
Less than 25	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
25-29	0.03**	0.00, 1.03	0.04*	0.00, 1.40	1.11	0.37, 3.35	0.87	0.27, 2.78
30-34	0.01**	0.00, 0.51	0.01**	0.00, 0.47	1.09	0.36, 3.30	1.04	0.30, 3.60
35-39	0.00***	0.00, 0.22	0.00***	0.00, 0.22	0.88	0.30, 2.55	0.77	0.23, 2.52
40-44	0.00***	0.00, 0.14	0.00***	0.00, 0.14	1.27	0.39, 4.10	0.84	0.24, 2.91
45+	0.02**	0.00 0.91	0.04*	0.00, 1.61	2.61	0.67, 10.21	2.16	0.52, 8.96
Marital status								
Married	36.46**	1.68, 791.74	33.71**	1.46, 776.37	2.27*	0.93, 5.54	3.60***	1.37, 9.46
Not married	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Self-reported health								
Excellent	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Very good	1.12	0.19, 6.73	1.94	0.26, 14.51	0.26**	0.08, 0.80	0.31*	0.10, 1.03
Good	0.89	0.16, 4.98	1.00	0.17, 5.75	1.08	0.33, 3.55	1.42	0.41, 4.95
Fair/poor	0.76	0.14, 4.22	0.76	0.13, 4.43	0.20***	0.07, 0.60	0.20***	0.06, 0.62
Community prevalence								
0-50% infected			1.00	(reference)			1.00	(reference)
50-100% infected			0.55	0.13, 2.36			0.78	0.35, 1.73
N	204		190		406		329	
LR χ^2	18.3		20.12		23.97		23.78	
prob > χ^2	0.075		0.065		0.013		0.107	

Notes: * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

Table 8: Multivariate logistic regression analysis of false negatives

	MALES				FEMALES			
	Model 1		Model 2		Model 1		Model 2	
Variable	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Region								
South	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Center	1.09	0.51, 2.31	1.08	0.48, 2.46	1.32	0.64, 2.74	1.16	0.50, 2.69
North	0.29**	0.11, 0.79	0.38*	0.14, 1.04	0.39**	0.17, 0.89	0.43**	0.18, 1.01
Age								
Less than 25	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
25-29	2.11	0.42, 10.61	2.22	0.41, 12.03	2.28	0.71, 7.34	1.70	0.42, 6.92
30-34	2.55	0.58, 11.11	2.88	0.61, 13.62	6.53***	2.38, 17.89	5.64***	1.86, 17.05
35-39	5.45**	1.38, 21.45	4.56**	1.01, 20.55	3.13**	0.96, 10.16	3.70**	1.08, 12.63
40-44	4.25**	0.98, 18.45	4.17*	0.84, 20.59	6.12***	2.19, 17.12	6.50***	2.23, 18.97
45+	2.49	0.67, 9.32	2.88	0.70, 11.81	1.56	0.51, 4.78	1.53	0.44, 5.33
Marital status								
Married	1.53	0.42, 5.62	1.61	0.39, 6.72	0.46**	0.21, 0.99	0.51	0.22, 1.17
Not married	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Self-reported health								
Excellent	1.00	(reference)	1.00	(reference)	1.00	(reference)	1.00	(reference)
Very good	0.50	0.13, 1.90	0.67	0.17, 2.65	1.20	0.52, 2.76	1.36	0.54, 3.40
Good	2.92***	1.30, 6.53	3.46***	1.45, 8.29	1.10	0.49, 2.45	1.28	0.53, 3.09
Fair/poor	2.60*	0.89, 7.65	2.58	0.79, 8.39	1.01	0.34, 3.00	1.19	0.34, 4.09
Community prevalence								
0-50% infected			1.00	(reference)			1.00	(reference)
50-100% infected			0.73	0.35, 1.51			0.89	0.45, 1.79
N	812		683		777		628	
LR χ^2	38.93		34.25		27.25		23.24	
prob > χ^2	0.000		0.001		0.004		0.081	

Notes: * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.