The Prevalence of HIV among Adults 15 and Older in Rural South Africa: A Cross-sectional Study

F. Xavier Gómez-Olivé, Samuel J. Clark, Brian Houle,
Kerstin Klipstein-Grobusch, Jane Menken, Nicole Angotti,
Chodziwadziwa Kabudula, Jill Williams, Stephen Tollman

Gaps remain in our knowledge of the HIV burden in rural areas among older adults and migrants. We estimated HIV prevalence and risk factors in a cross-sectional sero-survey of adults aged 15 and older in a rural South African population. A total of 5,037 (65.7%) individuals were located and 4,362 (86.6%) consented to HIV testing. Prevalence was estimated at 19.4%; 10.6% for men and 23.9% for women, peaking at 45.3% among men and at 46.1% among women, both at ages 35–39. Adjusting for non-response using Heckman selection models increased HIV prevalence by 6.3% for men and 1.5% for women. Compared with a study in KwaZulu-Natal, we found higher HIV prevalence among older adults – rates greater than 15% for men and 10% for women until age 70. High HIV prevalence suggests a great need for research, treatment, and prevention, particularly for older adults in rural areas.

Keywords: HIV; South Africa; rural health; selection bias

Introduction

Among world regions sub-Saharan Africa continues to have the highest burden of HIV/AIDS. In 2009 22.5 million people were living with HIV in the region (UN Joint Programme on HIV/AIDS 2010). Within sub-Saharan Africa the epidemic in South Africa remains one of the largest in the world – HIV prevalence in 2008 was 10.6% (Shisana et al. 2009). Evidence from 2002 – 2008 suggests that HIV prevalence has stabilized in South Africa, with a reduction among adolescents from 2005 – 2008. However, national estimates mask regional heterogeneity, with KwaZulu-Natal having the highest estimated regional prevalence of 21.5% (Welz et al. 2007).

Gaps remain in our understanding of the HIV epidemic in South Africa, and detailed information from rural areas remains scarce (Welz et al. 2007). Surveys often ignore the burden of HIV among those over age 50 (Edward J Mills, Anu Rammohan and Niyi Awofeso 2010; Negin and Cumming 2010). A study in KwaZulu-Natal expanded HIV surveillance in 2007 to include all eligible adults aged 15 and older. They found that HIV incidence among older adults was high, and indicated the need for a greater understanding of the burden, treatment, and prevention needs of this population (Wallrauch, Barnighausen and Newell 2010). Finally, HIV stigma remains problematic in South Africa, particularly in rural areas (Daftary, Padayatchi and Padilla 2007; Goudge et al. 2009; Matovu and Makumbi 2007). Stigma and other factors such as cyclical migration for work may increase nonresponse to prevalence surveys and reduce our ability to effectively quantify the burden of HIV and target treatment and prevention resources.

To address these gaps, we estimated HIV prevalence and its association with socio-demographic risk factors in a rural population in South Africa near the Mozambique border. The population has been under annual demographic surveillance since 1992, affording us additional data on individuals/households and a detailed sampling frame. We compare our results with two studies from KwaZulu-Natal (Wallrauch et al. 2010; Welz et al. 2007) and assess the effects of nonresponse on estimated HIV prevalence among the survey sample using Heckman selection models.

Method

This cross-sectional study took place in the Agincourt health and demographic surveillance system site (HDSS). The HDSS has conducted an annual household census since 1992 on a population living in the rural Agincourt sub-district of the Bushbuckridge District, Mpumalanga Province, South Africa. The study recruited men and women from the Agincourt HDSS study site from August 2010 – May 2011. We randomly selected 7,662 individuals from an eligible population of 34,413 using the 2009 HDSS census round as the sampling frame. Inclusion criteria were men and women aged 15 and older who were permanent residents prior to the 2009 census round. We stratified on sex and age and included an oversample of 284 adults over age 50 from a prior cohort study.

A research team visited selected participants in their homes up to two times for enrollment and informed consent. There were no material incentives provided to participate in the study; test results were made available to participants one month after the home visit at the two health facilities in the area offering antiretroviral treatment. Five dried blood spots were taken from consenting study participants. The samples were tested using screening assay Vironostika Uniform 11 (Biomerieux, France) – positives were confirmed by the SD Bioline HIV ELISA test (Standard Diagnostics Inc., Korea). If the screening and confirmatory assays did not agree a third assay was done. This third assay determined the final result following WHO criteria. Other anthropometric information and interview data, such as sexual behaviour, blood pressure, and height/weight were also collected. We obtained household socio-economic status information from the 2009 census data.

We used a probit regression to model socio-demographic risk factors for HIV status among those who were tested. We included sex, five-year age group, quintile of household socio-economic status in 2009, and previous migration status.

The main livelihood for the study population is cyclic labor migration to locations outside of the study site for periods ranging from daily to annual time scales. While both men and women participate in labor migration, most migrants are men. Since this sex-specific labor migration pattern affected nonresponse to the survey, we used two Heckman selection probit models to adjust HIV prevalence for study nonresponse (Barnighausen et al. 2011). We modelled both the relationship between being found for a potential interview and HIV status, and among those found, the relationship between consenting for HIV testing and HIV status. To increase the robustness of the regression results we included an indicator of the interviewer's identity in the selection equations for both models. To calculate a final adjusted prevalence for the whole population, we used the actual HIV status for those tested, the predicted probability from the model for HIV testing for those who were found but did not consent to testing, and the predicted probability from the model for being found for those who were not found. This approach is described in detail elsewhere (Barnighausen et al. 2011).

We performed all analyses using Stata 11.2 (StataCorp 2009). All estimations incorporated probability weights to produce population estimates.

The study received ethical approvals from the University of the Witwatersrand Human Research Ethics Committee and the Mpumalanga Provincial Research and Ethics Committee.

Results

Figure 1 presents the recruitment of the randomly selected participants. Of the 7,662 participants 469 were found to be ineligible for participation. Of the remaining 7,193 potential participants 5,037 (70%) were located for a potential interview. Of the remaining 5,037 located participants 353 refused to participate (7%), 322 consented to the interview but not HIV testing (6%) and 4,362 consented to both the interview and the HIV test (87%).

Table 1 presents the sex-age-specific HIV prevalence rates estimated from those who were tested, and the adjustments and adjusted HIV prevalence rates resulting from the Heckman selection model procedure.

Table 2 presents socio-demographic characteristics of males and females from the eligible sample. The sample is well-balanced on sex and age-group. Prior migration history was high for both males and females. One-quarter of the eligible sample was from the highest SES quintile in 2009.

Table 3 presents HIV socio-demographic risk factors among study participants from a probit regression among those who were tested. An interaction between sex and age improved model fit ($p < 0.001$) and was included. Those in the middle low and high SES quintiles had a lower probability of being HIV positive relative to those in the low quintile. Men at ages 15–19 ($p = 0.001$), 20–24 ($p < 0.001$), and 25–29 ($p < 0.001$) had a lower probability of being HIV positive relative to women in those ages.

Figure 2(A) presents unadjusted HIV prevalence estimates by sex and age. The measured prevalence for all ages was 19.4% (23.9% for females and 10.6% for males). Males had a peak prevalence of 45.3% at ages 35–39 and prevalence remained over 15% until ages 70 and over. Females had a peak prevalence of 46.1% also at ages 35-39, with prevalence remaining over 10% until ages 70 and over.

Figure 2(A) also presents the adjusted HIV prevalence estimates by sex and age using the Heckman selection models. The adjusted prevalence increased by 2.7% over the unadjusted estimate to 22.1% (25.4% for females and 16.9% for males). Prevalence increased by 6.3% for men but only 1.5% for women in the adjusted compared to unadjusted estimates. The main sex-based difference in the sample subgroups occurs for men who were not found: 1,416 men compared to 738 women (Figure 1). It is this distortion in the missing age structure, rather than large differences in age-specific prevalence that appears to be driving the large sex-based difference in the unadjusted HIV prevalence and the larger adjustment to HIV prevalence for men than women.

Adjusting for non-response could have also altered our estimation of socio-demographic risk factors for HIV. In this case the substantive findings are similar between the two regressions based on individuals who were interviewed (see Table 3 and Table 5). However, using the regression on all individuals (Table 4) indicates that younger men have an increased probability of being HIV positive ($p < 0.001$ at ages 20–24) – a finding that corroborates with the adjusted HIV prevalence estimates for men.

Figure 2(B) presents the age-specific prevalence estimates from two studies from KwaZulu-Natal province that also included older ages (Wallrauch et al. 2010; Welz et al. 2007). Overall HIV prevalence estimates are comparable between studies but the age-patterns are different. In KwaZulu-Natal prevalence is skewed to the left, with high HIV prevalence among younger ages that steadily declines with age. In Agincourt peak HIV prevalence is among slightly older ages, with a slower decline with age.

Discussion

Using a cross-sectional survey we estimated HIV prevalence in rural South Africa for adults ages 15 and older. We found high regional prevalence in Mpumalanga comparable to KwaZulu-Natal, the region with the highest estimated prevalence in South Africa (Wallrauch et al. 2010; Welz et al. 2007). Compared to KwaZulu-Natal, HIV prevalence in Agincourt peaks at slightly older ages and is higher at older ages.

A relatively high burden of HIV among those over age 50 raises several questions. First, it is unknown if these individuals are contracting HIV at earlier ages and surviving for long periods, or if they are acquiring HIV at older ages; additional analyses of sexual risk behaviour among older adults are needed. Second, this high prevalence may affect their capacity to care for grandchildren, creating an epidemic that affects both older people and those under their care (Kautz et al. 2010). Third, older people who also suffer from chronic non-communicable diseases (NCD) will need to use health facilities more frequently, seeking chronic care for both NCD and HIV.

Access to prior data from the longitudinal Agincourt HDSS provided a complete sampling frame of eligible participants. However, due to the annual design of the census we found a high non-response rate resulting from migration. We used Heckman selection models to adjust HIV prevalence estimates for non-response. These methods rely on valid exclusion criteria, and we used interviewer identity following the suggestion of Barnighausen et al. (2011). A key assumption of our adjustment is that interviewer identity is uncorrelated with an individual's HIV status. Adjusting for non-response showed that prevalence among those who tested underestimated HIV prevalence, particularly for younger men, which reduced the female-male differential from 13.3% (unadjusted) to 8.5% (adjusted).

The longitudinal nature of the HDSS will allow us to design new studies in the future to link the distribution of disease burden and causes of mortality and how all of this changes over time. Additionally, those who were found HIV negative in this study may be invited to take part in a research cohort to identify HIV incidence in the study population. Similarly, a HIV positive cohort can explore issues of adherence, resistance and other outcomes. The similarity between Africa Centre and Agincourt data, widely separated rural South African settings, may be an indication that these results may be approximately generalizable for rural South Africa.

Prevention activities need to expand to older adults to reduce new infections. As ART rollout becomes more widespread, the healthcare system needs to target treatment among older adults that are living with HIV. Effective treatment will also be complicated by increased risk of non-communicable disease as individuals age and the requirement to coordinate care and follow-up to reduce excess overall disease burden is made more complex by increasing numbers of older people living with HIV (Levitt et al. 2011).

Acknowledgments

R24AG032112 National Institute on Aging (NIA), Partnership for Social Science AIDS Research in South Africa's Era of ART Rollout. K01 HD057246 and R01 HD054511 National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (NIH). Wellcome Trust. Anglo American Chairman's Fund. MRC/Wits Rural Public Health and Health Transitions Unit (Agincourt), University of the Witwatersrand. Ha Nakekela ("We Care") Field Team. Participating communities.

References

Barnighausen, T., J. Bor, S. Wandira-Kazibwe, and D. Canning. 2011. "Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models." *Epidemiology* 22(1):27-35.

Daftary, A., N. Padayatchi, and M. Padilla. 2007. "HIV testing and disclosure: a qualitative analysis of TB patients in South Africa." *AIDS care* 19(4):572-577.

Edward J Mills, Anu Rammohan, and Niyi Awofeso. 2010. "Ageing faster with AIDS in Africa." *The Lancet*(Published online):DOI:10.1016/S0140-6736(1010)62180-62180.

Goudge, J., B. Ngoma, L. Manderson, and H. Schneider. 2009. "Stigma, identity and resistance among people living with HIV in South Africa." *Sahara J* 6(3):94-104.

Kautz, T., E. Bendavid, J. Bhattacharya, and G. Miller. 2010. "AIDS and declining support for dependent elderly people in Africa: retrospective analysis using demographic and health surveys." *BMJ* 340:c2841.

Levitt, N.S., K. Steyn, J. Dave, and D. Bradshaw. 2011. "Chronic noncommunicable diseases and HIV-AIDS on a collision course: relevance for health care delivery, particularly in low-resource settings--insights from South Africa." *The American journal of clinical nutrition* 94(6):1690S-1696S.

Matovu, J.K.and F.E. Makumbi. 2007. "Expanding access to voluntary HIV counselling and testing in sub-Saharan Africa: alternative approaches for improving uptake, 2001-2007." *Trop Med Int Health* 12(11):1315-1322.

Negin, J.and R.G. Cumming. 2010. "HIV infection in older adults in sub-Saharan Africa: extrapolating prevalence from existing data." *Bulletin of the World Health Organization* 88(11):847-853.

Shisana, O., Human Sciences Research Council., Centre for AIDS Development Research and Evaluation., South African Medical Research Council., and National Institute for Communicable Diseases (South Africa). 2009. *South African national HIV prevalence , incidence, behaviour and communication survey, 2008: A turning tide among teenagers*. Cape Town: HSRC Press.

StataCorp. 2009. "Stata Statistical Software: Release 11." College Station, TX: StataCorp LP.

UN Joint Programme on HIV/AIDS. 2010. "Global Report: UNAIDS Report on the Global AIDS Epidemic: 2010."

Wallrauch, C., T. Barnighausen, and M.L. Newell. 2010. "HIV prevalence and incidence in people 50 years and older in rural South Africa." *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde* 100(12):812-814.

Welz, T., V. Hosegood, S. Jaffar, J. Batzing-Feigenbaum, K. Herbst, and M.L. Newell. 2007. "Continued very high prevalence of HIV infection in rural KwaZulu-Natal, South Africa: a population-based longitudinal study." *AIDS* 21(11):1467-1472.

Tables

Table 1: HIV Prevalence (%)

	Measured (95% CI)				Adjustment		Adjusted	
Age	**Female**		**Male**		**Female**	**Male**	**Female**	**Male**
15-19	**5.5**	(2.6 - 8.4)	**0.4**	(0.0 - 1.3)	0.1	0.4	5.6	0.8
20-24	**27.0**	(21.9 - 32.2)	**6.1**	(2.9 - 9.4)	-0.2	2.1	26.9	8.3
25-29	**37.8**	(32.1 - 43.4)	**21.7**	(15.2 - 28.3)	0.5	7.1	38.3	28.8
30-34	**41.8**	(36.2 - 47.3)	**41.8**	(33.7 - 50.0)	-0.4	4.7	41.4	46.6
35-39	**46.1**	(40.7 - 51.6)	**45.3**	(38.1 - 52.6)	0.8	3.0	46.9	48.3
40-44	**34.4**	(28.1 - 40.8)	**41.0**	(31.4 - 50.6)	1.4	4.5	35.8	45.5
45-49	**34.2**	(28.0 - 40.4)	**28.8**	(20.9 - 36.7)	0.9	3.8	35.1	32.6
50-54	**26.9**	(19.4 - 34.4)	**30.6**	(19.9 - 41.2)	0.1	5.3	26.9	35.8
55-59	**26.8**	(19.5 - 34.0)	**34.6**	(24.2 - 44.9)	0.3	0.3	27.1	34.9
60-64	**13.1**	(7.6 - 18.6)	**19.8**	(12.4 - 27.2)	1.5	1.9	14.6	21.7
65-69	**10.3**	(5.2 - 15.4)	**16.5**	(8.9 - 24.1)	1.8	1.6	12.1	18.1
70-74	**11.0**	(4.6 - 17.4)	**5.7**	(0.8 - 10.5)	1.2	0.5	12.2	6.2
75-79	**6.2**	(0.9 - 11.4)	**5.3**	(0.0 - 12.4)	1.2	0.9	7.4	6.2
80-84	**1.3**	(0.0 - 3.8)	**1.8**	(0.0 - 5.3)	0.4	0.9	1.7	2.7
15-84	**23.9**	(22.2 - 25.6)	**10.6**	(9.3 - 12.0)	1.5	6.3	25.4	16.9

Table 2: Sample socio-demographics by sex for randomly selected sample from the universe of adults ages 15 and over in the study population (N = 34,413), stratified by sex and age

	Female (%) n = 3892	Male (%) n = 3770	Total (%) n = 7662
Sex			
Female	100	0	52
Male	0	100	48
Age group			
15 – 19	8	8	8
20 – 24	12	13	12
25 – 29	12	13	13
30 – 34	12	13	12
35 – 39	12	13	12
40 – 44	9	8	9
45 – 49	8	8	8
50 – 54	4	5	5
55 – 59	5	4	5
60 – 64	5	5	5
65 – 69	4	4	4
70 – 74	3	3	3
75 – 79	3	1	2
80 – 84	3	2	2
SES quintile			
low	15	15	15
middle-low	19	19	19
middle	21	20	21
middle-high	21	21	21
high	24	26	25
Previous migration history			
No	35	45	40
Yes	65	55	60

Table 3: Probit regression of HIV status by socio-demographic characteristics among those tested for HIV (n = 4,343)

	Beta	Lower 95% CI	Upper 95% CI
Male	-1.066	-1.764	-0.369
Age			
15 – 19	–	–	–
20 – 24	1.007	0.702	1.311
25 – 29	1.346	1.045	1.648
30 – 34	1.402	1.103	1.701
35 – 39	1.502	1.206	1.798
40 – 44	1.229	0.914	1.543
45 – 49	1.221	0.907	1.536
50 – 54	1.025	0.675	1.375
55 – 59	1.012	0.665	1.360
60 – 64	0.523	0.153	0.893
65 – 69	0.373	-0.022	0.768
70 – 74	0.354	-0.081	0.789
75 – 79	0.032	-0.470	0.533
80 – 84	-0.669	-1.468	0.131
Sex *X* Age			
Male *X* 20 – 24	0.095	-0.659	0.850
Male *X* 25 – 29	0.584	-0.163	1.332
Male *X* 30 – 34	1.078	0.346	1.809
Male *X* 35 – 39	1.095	0.366	1.824
Male *X* 40 – 44	1.236	0.483	1.989
Male *X* 45 – 49	0.937	0.176	1.699
Male *X* 50 – 54	1.174	0.379	1.968
Male *X* 55 – 59	1.300	0.520	2.081
Male *X* 60 – 64	1.337	0.546	2.128
Male *X* 65 – 69	1.343	0.522	2.163
Male *X* 70 – 74	0.777	-0.113	1.668
Male *X* 75 – 79	1.008	-0.045	2.061
Male *X* 80 – 84	1.216	-0.092	2.524
Village			
1	–	–	–
2	0.154	-0.216	0.524
3	0.112	-0.130	0.354
4	-0.071	-0.365	0.223
5	-0.127	-0.396	0.141
6	0.022	-0.263	0.307
7	-0.060	-0.361	0.241
8	-0.090	-0.341	0.160
9	-0.108	-0.356	0.140
10	-0.208	-0.449	0.033
11	0.059	-0.165	0.282
12	0.112	-0.189	0.413
13	-0.013	-0.295	0.269
14	0.019	-0.326	0.365
15	0.049	-0.227	0.326
16	-0.379	-0.659	-0.100
17	0.134	-0.173	0.441
18	0.232	-0.162	0.627
19	0.197	-0.230	0.623
20	-0.263	-0.693	0.168
21	0.690	0.312	1.069
Prior migration history	-0.023	-0.139	0.093
SES Quintiles			
low	–	–	–
middle-low	-0.166	-0.326	-0.005

middle	-0.085	-0.251	0.081
middle-high	-0.116	-0.292	0.061
high	-0.415	-0.584	-0.247
Constant	-1.442	-1.799	-1.085

Table 4: Heckman contact selection probit model among those eligible for potential interview

	Beta	Lower 95% CI	Upper 95% CI
HIV Status			
Age			
15 – 19	–	–	–
20 – 24	0.886	0.618	1.154
25 – 29	1.198	0.929	1.468
30 – 34	1.246	0.981	1.51
35 – 39	1.386	1.13	1.643
40 – 44	1.157	0.883	1.431
45 – 49	1.114	0.843	1.386
50 – 54	0.901	0.596	1.207
55 – 59	0.914	0.614	1.214
60 – 64	0.583	0.272	0.893
65 – 69	0.57	0.256	0.883
70 – 74	0.511	0.156	0.866
75 – 79	0.324	-0.063	0.711
80 – 84	-0.098	-0.558	0.363
Male	-0.179	-0.505	0.146
Sex *X* Age			
Male *X* 20 – 24	-0.43	-0.842	-0.018
Male *X* 25 – 29	0.031	-0.382	0.444
Male *X* 30 – 34	0.368	-0.055	0.79
Male *X* 35 – 39	0.265	-0.146	0.677
Male *X* 40 – 44	0.432	-0.014	0.879
Male *X* 45 – 49	0.196	-0.236	0.628
Male *X* 50 – 54	0.532	0.05	1.013
Male *X* 55 – 59	0.368	-0.113	0.85
Male *X* 60 – 64	0.406	-0.053	0.866
Male *X* 65 – 69	0.285	-0.192	0.763
Male *X* 70 – 74	-0.129	-0.669	0.411
Male *X* 75 – 79	0.021	-0.65	0.692
Male *X* 80 – 84	0.483	-0.172	1.137
Village			
1	–	–	–
2	0.167	-0.123	0.457
3	0.124	-0.086	0.334
4	0.125	-0.135	0.385
5	-0.109	-0.337	0.119
6	0.148	-0.103	0.399
7	-0.047	-0.33	0.236
8	-0.07	-0.282	0.141
9	0.145	-0.074	0.364
10	-0.281	-0.488	-0.075
11	0.007	-0.185	0.198
12	-0.013	-0.269	0.243
13	0.046	-0.186	0.279
14	0.118	-0.223	0.459
15	0.003	-0.254	0.26
16	-0.115	-0.349	0.118
17	0.146	-0.135	0.426
18	0.097	-0.237	0.432
19	0.104	-0.252	0.459
20	-0.184	-0.542	0.175
21	0.521	0.177	0.865
Constant	-1.428	-1.705	-1.15
Survey participation			
Age			
15 – 19	–	–	–

20 – 24	-0.301	-0.513	-0.088
25 – 29	-0.353	-0.564	-0.142
30 – 34	-0.257	-0.468	-0.045
35 – 39	-0.031	-0.246	0.185
40 – 44	-0.161	-0.388	0.066
45 – 49	0.031	-0.207	0.268
50 – 54	0.229	-0.069	0.528
55 – 59	0.174	-0.119	0.467
60 – 64	0.595	0.261	0.928
65 – 69	0.498	0.16	0.837
70 – 74	0.425	0.037	0.812
75 – 79	0.642	0.205	1.079
80 – 84	0.394	0	0.788
Male	0.183	-0.065	0.432
Sex *X* Age			
Male *X* 20 – 24	-0.672	-0.976	-0.369
Male *X* 25 – 29	-0.81	-1.112	-0.508
Male *X* 30 – 34	-0.934	-1.237	-0.631
Male *X* 35 – 39	-0.932	-1.237	-0.626
Male *X* 40 – 44	-0.967	-1.292	-0.641
Male *X* 45 – 49	-0.973	-1.305	-0.641
Male *X* 50 – 54	-1.008	-1.413	-0.604
Male *X* 55 – 59	-0.963	-1.363	-0.562
Male *X* 60 – 64	-0.937	-1.372	-0.502
Male *X* 65 – 69	-0.796	-1.25	-0.343
Male *X* 70 – 74	-0.702	-1.205	-0.199
Male *X* 75 – 79	-0.817	-1.444	-0.19
Male *X* 80 – 84	0.078	-0.578	0.733
Interviewer			
1	–	–	–
2	-1.049	-1.367	-0.732
3	-0.746	-1.082	-0.409
4	-1.541	-1.872	-1.209
5	-1.192	-1.514	-0.869
6	-1.301	-1.62	-0.983
7	-1.156	-1.473	-0.839
8	-1.141	-1.46	-0.822
9	-1.295	-1.611	-0.979
10	-1.118	-1.434	-0.801
11	-0.948	-1.266	-0.629
Constant	2.019	1.687	2.351

Total = 7191, censored = 2154; ρ = 0.219, 95% CI = [-0.103, 0.549]; Wald test of exclusion restrictions on survey participation, $\chi^2 = 1.80$, $p = 0.180$

Table 5: Heckman consent selection probit model among those interviewed

	Beta	Lower 95% CI	Upper 95% CI
HIV status			
Age			
15 – 19	–	–	–
20 – 24	1.024	0.722	1.325
25 – 29	1.388	1.084	1.691
30 – 34	1.423	1.127	1.718
35 – 39	1.534	1.239	1.828
40 – 44	1.269	0.952	1.587
45 – 49	1.249	0.936	1.562
50 – 54	1.031	0.684	1.378
55 – 59	1.028	0.685	1.372
60 – 64	0.554	0.181	0.927
65 – 69	0.429	0.011	0.848
70 – 74	0.396	-0.046	0.839
75 – 79	0.079	-0.439	0.598
80 – 84	-0.622	-1.433	0.189
Male	-1.027	-1.715	-0.339
Sex *X* Age			
Male *X* 20 – 24	0.098	-0.64	0.835
Male *X* 25 – 29	0.631	-0.114	1.376
Male *X* 30 – 34	1.139	0.412	1.866
Male *X* 35 – 39	1.116	0.405	1.827
Male *X* 40 – 44	1.27	0.533	2.007
Male *X* 45 – 49	0.958	0.213	1.702
Male *X* 50 – 54	1.237	0.444	2.03
Male *X* 55 – 59	1.292	0.528	2.055
Male *X* 60 – 64	1.319	0.545	2.094
Male *X* 65 – 69	1.288	0.47	2.106
Male *X* 70 – 74	0.72	-0.16	1.601
Male *X* 75 – 79	0.949	-0.098	1.996
Male *X* 80 – 84	1.202	-0.079	2.483
Village			
1	–	–	–
2	0.172	-0.19	0.534
3	0.114	-0.125	0.352
4	-0.035	-0.336	0.267
5	-0.136	-0.402	0.13
6	0.049	-0.238	0.336
7	-0.065	-0.363	0.232
8	-0.091	-0.338	0.156
9	-0.058	-0.337	0.221
10	-0.231	-0.483	0.021
11	0.05	-0.175	0.275
12	0.088	-0.221	0.397
13	-0.002	-0.28	0.276
14	0.04	-0.304	0.384
15	0.029	-0.257	0.315
16	-0.35	-0.639	-0.062
17	0.141	-0.163	0.445
18	0.218	-0.174	0.611
19	0.18	-0.244	0.603
20	-0.262	-0.687	0.164
21	0.668	0.279	1.058
Prior migration history	-0.014	-0.13	0.103
SES Quintiles			
low	–	–	–
middle-low	-0.164	-0.324	-0.004

middle	-0.066	-0.239	0.106
middle-high	-0.074	-0.293	0.145
high	-0.359	-0.615	-0.103
Constant	-1.43	-1.789	-1.071
Survey participation			
Age			
15 – 19	–	–	–
20 – 24	-0.416	-0.779	-0.054
25 – 29	-0.676	-1.025	-0.327
30 – 34	-0.522	-0.868	-0.176
35 – 39	-0.672	-1.022	-0.322
40 – 44	-0.702	-1.065	-0.338
45 – 49	-0.566	-0.934	-0.198
50 – 54	-0.334	-0.766	0.098
55 – 59	-0.428	-0.838	-0.018
60 – 64	-0.505	-0.91	-0.1
65 – 69	-0.669	-1.07	-0.268
70 – 74	-0.585	-1.038	-0.132
75 – 79	-0.574	-1.053	-0.096
80 – 84	-0.403	-0.915	0.108
Male	-0.241	-0.627	0.145
Sex *X* Age			
Male *X* 20 – 24	0.027	-0.462	0.517
Male *X* 25 – 29	-0.297	-0.767	0.173
Male *X* 30 – 34	-0.54	-1.008	-0.072
Male *X* 35 – 39	-0.235	-0.694	0.224
Male *X* 40 – 44	-0.316	-0.811	0.178
Male *X* 45 – 49	-0.21	-0.705	0.285
Male *X* 50 – 54	-0.609	-1.171	-0.047
Male *X* 55 – 59	-0.089	-0.685	0.507
Male *X* 60 – 64	0.008	-0.538	0.553
Male *X* 65 – 69	0.301	-0.272	0.873
Male *X* 70 – 74	0.504	-0.136	1.144
Male *X* 75 – 79	0.42	-0.374	1.214
Male *X* 80 – 84	-0.01	-0.721	0.702
Village			
1	–	–	–
2	-0.097	-0.425	0.231
3	-0.014	-0.277	0.249
4	-0.274	-0.588	0.04
5	0.116	-0.178	0.409
6	-0.226	-0.525	0.073
7	0.065	-0.303	0.432
8	0.019	-0.243	0.281
9	-0.361	-0.622	-0.099
10	0.357	0.087	0.627
11	0.074	-0.165	0.314
12	0.276	-0.065	0.618
13	-0.078	-0.358	0.202
14	-0.188	-0.609	0.233
15	0.21	-0.147	0.567
16	-0.161	-0.425	0.103
17	-0.08	-0.501	0.341
18	0.161	-0.304	0.626
19	0.193	-0.274	0.66
20	0.07	-0.391	0.532
21	0.139	-0.314	0.591
Prior migration history	-0.076	-0.2	0.048
SES Quintiles			
low	–	–	–

middle-low	0.027	-0.173	0.228
middle	-0.147	-0.346	0.052
middle-high	-0.359	-0.554	-0.163
high	-0.435	-0.626	-0.245
Interviewer			
1	–	–	–
2	-0.201	-0.489	0.088
3	-0.266	-0.552	0.02
4	0.008	-0.352	0.368
5	0.044	-0.224	0.311
6	-0.085	-0.394	0.224
7	-0.385	-0.662	-0.107
8	-0.207	-0.473	0.059
9	-0.306	-0.622	0.01
10	-0.273	-0.55	0.004
11	-0.108	-0.387	0.171
Constant	2.295	1.842	2.749

Total = 5013, censored = 672; $\rho = -0.342$, 95% CI = [-0.868, 0.546]; Wald test of exclusion restrictions on survey participation, $\chi^2 = 0.52$, $p = 0.471$

Figures

Figure 1: Flowchart of randomly selected participants based on eligibility, being located for potential interview, consenting to interview, and consenting to HIV testing, by sex

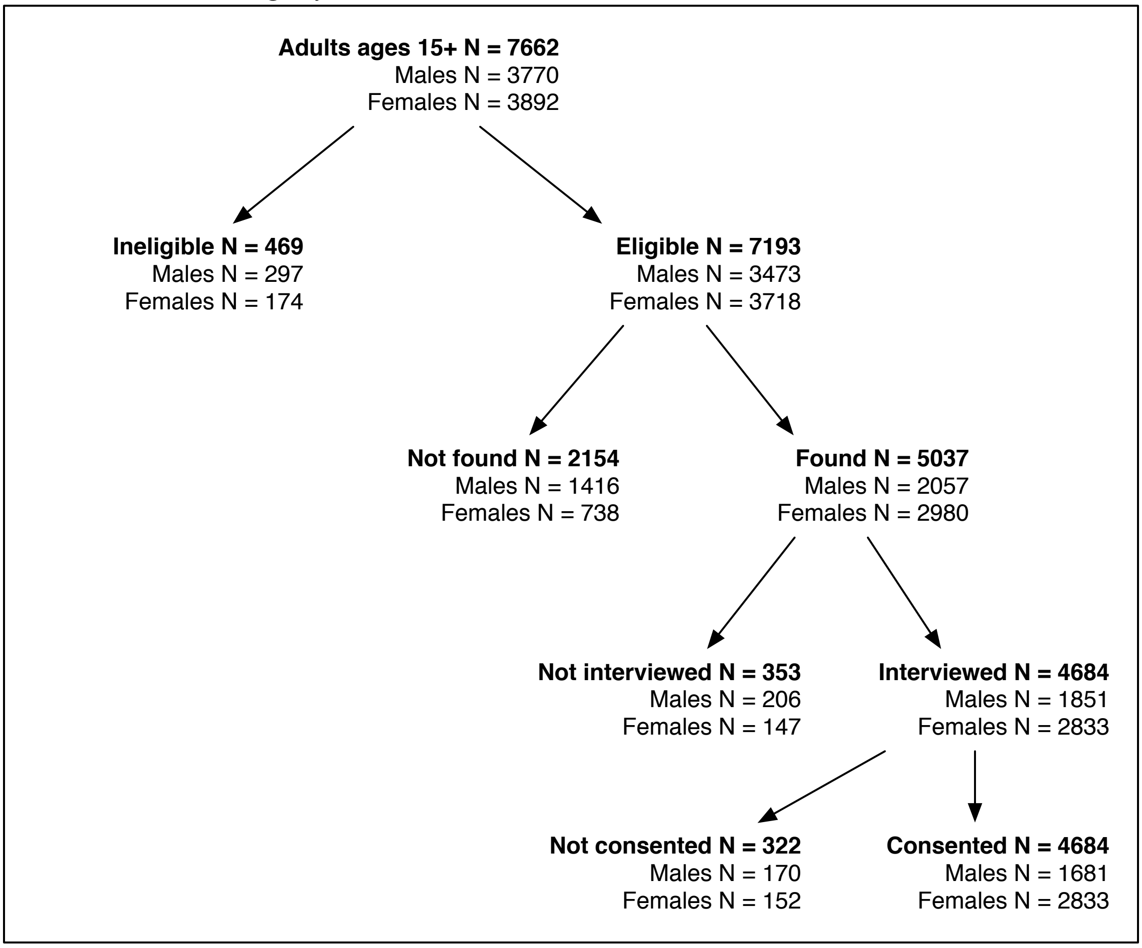

Figure 2: HIV prevalence by sex and age of: (A) Agincourt 2010 unadjusted and adjusted estimates; and (B) KwaZulu-Natal estimates