Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys

Georges Reniers [1][2] & Jeffrey Eaton [3]

[1] Population Program, Institute of Behavioral Science, University of Colorado at Boulder, CO

[2] School of Social Sciences and School of Public Health, University of the Witwatersrand, South Africa

[3] Department of Statistics, University of Washington, WA

Corresponding author:

Georges Reniers

Population Program, Institute of Behavioral Science

University of Colorado at Boulder

Boulder, CO 80309-0484

e-mail: georges.reniers@colorado.edu

Word count abstract: 254

Word count text: 3487

Objectives: We assess the relationship between prior knowledge of one's HIV status and the likelihood to refuse an HIV test in populations-based surveys, and explore its potential for producing bias in HIV prevalence estimates.

Methods: Using longitudinal survey data from Malawi, we estimate the relationship between prior knowledge of one's HIV status and subsequent refusal of an HIV test, and use that parameter to develop a heuristic model of refusal bias that is applied to Demographic and Health Surveys (DHS) where refusal by HIV status is not observed. The model only accounts for refusal bias conditional on a completed interview.

Results: HIV status, prior testing status, and refusal for HIV testing are highly correlated in seroprevalence surveys. Further, the Malawian data indicate that HIV positives are 4.62 (95%-CI: 2.60-8.21) times more likely than HIV negatives to refuse repeat testing in a subgroup of respondents who are aware of their HIV status. Based on that parameter and other inputs from the DHS, our model suggests that a downward bias of 13.3% (95%-CI: 7.2%-19.6%) in national HIV prevalence estimates is possible. In some urban populations, bias exceeds 20%. Because refusal rates are higher in men, seroprevalence surveys tend to overestimate the female to male ratio of infections.

Conclusions: Prior knowledge of HIV status informs decisions to participate in HIV seroprevalence surveys. These informed refusals produce downward bias in HIV prevalence estimates, and leads to biased estimates of the sex ratio of infections. Our results suggest that the recent downward adjustments in HIV prevalence estimates are too optimistic.

Introduction

Historically, most published estimates of HIV prevalence in sub-Saharan Africa are based on sentinel surveillance data in antenatal clinics (ANC). Because of the importance of reasonably accurate HIV prevalence figures for policy formulation and resource allocation, the validity of these estimates have been subject to extensive scrutiny. ANC based estimates typically overestimate true prevalence. This is attributed to the representativeness of women attending antenatal clinics and the under-representation of remote rural areas in surveillance systems [1-15]. The identification of bias has led to the development of correction schemes to improve extrapolations from ANC surveillance data [2, 16-18], but questions continue to surround the uniform applicability of these adjustments in a variety of settings [12].

Expanding resources and progress in medical technology has brought HIV testing increasingly within reach of nationally representative household surveys, and that has generated new prospects of resolving the type and magnitude of bias in ANC sentinel surveillance estimates, or, to provide a new gold standard for HIV prevalence estimates altogether [10, 19-22]. The inclusion of HIV serostatus testing in several Demographic and Health Surveys (DHS) and AIDS Indicator Surveys (AIS) is pushing the agenda in that respect.

Data from population-based surveys are indeed a valuable addition to ANC estimates, but they are also subject to bias due to limitations of the sampling frame (e.g., the exclusion of high risk groups in army barracks, prisons or health facilities) and non-response because of individual mobility and refusal. The association between mobility and HIV infection has been documented extensively [12, 23-31]. In comparison, relatively little is known about the relationship between

refusal and HIV infection in population-based studies [10, 20, 22]. A number of small-scale studies in STD and antenatal clinics concluded that refusals are positively associated with HIV status [32-40], others remain inconclusive about the nature of the relationship or suggest the opposite pattern [41-43].

Overall, population-based seroprevalence surveys are believed to underestimate true HIV prevalence, but the studies that have addressed this issue have failed to identify significant refusal bias [13, 22, 31, 44-47]. One notable exception is a study from a demographic surveillance site in KwaZulu-Natal, South Africa [48]. These studies do not, however, account for the possibility that respondents' decisions to refuse testing are informed by prior knowledge of their HIV status. Cross-sectional seroprevalence surveys usually inquire about prior testing and whether the result of the test was received, but the outcome of the test is not known. We hpothesize that HIV positive individuals who are aware of their HIV status are much less likely to consent to testing in a seroprevalence survey than those who previously tested negative and those who were not previously tested.. Furthermore, we demonstrate that these informed refusals can bias HIV seroprevalence estimates from nationally representative seroprevalence surveys, particularly in settings where HIV prevalence, refusal rates, and HIV testing coverage are relatively high.

Methods

This study consists of three parts. First, we describe levels of prior testing and refusal in African countries with available data, and explore the ecological association between refusal

rates, prior testing rates and HIV prevalence. Second, we investigate the individual-level relationship between prior knowledge of one's HIV status and consent for testing using DHS data, and longitudinal survey data from Malawi. Finally, we develop a heuristic model of bias in HIV seroprevalence surveys that is based on HIV prevalence, the prior testing rate, the refusal rate and assumptions about the relation between prior knowledge of one's HIV status and consent for testing. We apply the model to six DHS to demonstrate potential bias in HIV prevalence estimates stemming from selective refusals by individuals with prior knowledge of their HIV status.

The DHS (http://www.measuredhs.com/) are a widely used source for social science and public health policy research in developing countries. In 2001, the DHS (and AIDS Indicator Surveys, AIS) started administering HIV tests with the objective of providing population-based seroprevalence estimates, and to provide insights in the sociodemographic and behavioral factors associated with HIV infection. Study participants are typically approached for testing following a successful interview, but the protocol in that regard is not uniform. With the exception of Cameroon (2004), the proportion of respondents who were tested but not interviewed was under 1%. We will ignore this category of respondents in this paper. The DHS and AIS, do not usually give feedback to respondents about their test results. Instead, respondents receive referrals for free counseling and retesting at local Voluntary Counseling and Testing (VCT) establishments. In some countries, mobile units have followed the survey to counsel and test interested DHS respondents.

In this paper, we report aggregate data on prior testing, refusal and HIV prevalence from Cameroon (2004), Ethiopia (2005), Ghana (2003), Guinea (2005), Kenya (2003), Lesotho (2004), Malawi (2004), Niger (2006), Rwanda (2005), Tanzania (2003-04), Senegal (2005),

Uganda (2004-05), Zambia (2002) and Zimbabwe (2005-06), and provide more detailed estimates of non-response bias for six countries: Senegal, Ghana, Cameroon, Malawi, Lesotho and Zimbabwe. To assess the hypothesis that HIV positive individuals who are aware of their status are more likely to refuse testing than HIV negative individuals who know their status, we use data from the Malawi Diffusion and Ideational Change Project (MDICP, http://www.malawi.pop.upenn.edu/). These are longitudinal survey data with HIV serostatus testing in waves three (MDICP3) and four (MDICP4). In this dataset we know which respondents were tested in MDICP3, and if so, their HIV status and whether or not they chose to learn their results. In addition, we know whether or not the respondent chose to be retested in MDICP4. We use those data to obtain an empirical estimate of the relationship between HIV status and subsequent consent for testing for respondents with knowledge of their HIV status. The original MDICP sample that was taken in 1998 included around 1,500 ever-married women and their spouses. In MDICP3, the sample was augmented with a group of adolescents (both sexes). In MDICP3, a total of 3,284 individuals were approached for an HIV test using OraSure® saliva swabs [49]. Post-test counseling was offered in VCT tents in or close by the villages of the respondents one to three months after testing. The second round of HIV testing and counseling took place in 2006, this time using a finger-prick rapid tests (Determine® and UniGold). The respondents could choose the testing location (either in the home or in a VCT tent in the village), and post-test counseling was done 20 to 30 minutes after the test. Respondents were given the option to be tested and counseled, or, to be tested without post-test counseling and disclosure of the test results. For individuals who received their test result in MDICP3 and were contacted again in MDICP4, we calculate the relative risk of refusing for HIV positives

compared to HIV negatives using a log-binomial model. Henceforth, we label this relative risk the E parameter.

The possibility that refusals are informed by prior knowledge of HIV positive status does not necessarily imply that refusals produce substantial bias in national or local estimates of HIV prevalence. Bias will also depend on the refusal and prior testing rates. From the DHS we obtain the proportion of the population tested previously, the refusal rate amongst individuals tested previously, the refusal rate of those not tested previously, and, of course, the HIV prevalence in the sample that did not refuse, each stratified by sex and place of residence (urban/rural). Assuming that the risk of refusing for individuals who know they are HIV positive is E times greater than the risk of refusing for individuals who know they are HIV negative, a simple probability calculation yields an estimate of HIV prevalence amongst those who refused, and hence, an estimate of the population-level HIV prevalence (see Appendix). The adjusted population HIV prevalence is estimated for each population subgroup (urban/rural, male/female) separately. The country-level estimate is calculated as a weighted average of the prevalence in each subgroup. For that purpose, the population distribution by place of residence is taken from the UN World Urbanization Prospects database [50], and sex ratios are assumed to be in balance. Estimates of bias are presented in terms of the absolute difference in the observed and adjusted HIV prevalence, as well as by their ratio.

In this calculation three further assumptions are made. The first is that refusal is uncorrelated with HIV status among those who have not been previously tested and counseled for HIV. The second is that prior testing for HIV is independent of HIV status. This means that individuals who are HIV negative are just as likely to know their HIV status as individuals who are HIV positive. This is unlikely to be true: in most DHS surveys, HIV positive individuals are

more likely to have previously received an HIV test (Table 1). In the South African HSRC survey, HIV positive respondents were twice as likely to have received a previous HIV test [51]. We note that this assumption is made to be conservative: if HIV positive individuals are indeed more likely to know their status and also more likely to refuse a subsequent test, then these factors will compound each other to increase bias. Finally, we assume that the only source of bias is from refusal to be tested by those who have been successfully contacted and completed the individual interview portion. In conservative fashion, the model thus ignores potential bias resulting from higher absenteeism in HIV positive individuals and a greater propensity to refuse the survey outright (and not just the HIV test).

Results

Prior testing and refusal. Figure 1 illustrates the prior testing rate, HIV prevalence and the refusal rate by type of place of residence and gender for fourteen African countries. Refusal rates range from under 1% in rural Rwanda to 25% in urban populations of Malawi, Ethiopia, Lesotho and Zimbabwe. Refusal rates vary quite importantly by place of residence and sex: the median refusal rates in urban and rural areas are 16.3% and 8.8%, respectively; the median refusal rates for men and women are 14.6% and 9.2%. Rates of prior testing vary from under 1% for women in rural Guinea and Niger to 43% for women in urban Rwanda. The median rate of prior testing is 11.7%, and is a little higher for men than for women. The difference by place of residence is larger: the median prior testing rates for urban and rural areas are 17.2% and 8.1%, respectively. Figure 1 is also suggestive of a three-way ecological relationship between HIV prevalence, prior

testing and refusal. Rwanda and Uganda stand out with relatively high prior testing rates and relatively low refusal rates. Excluding these two countries, the ecological correlation between either of these variables is greater than 0.5.

The relationship between prior testing status and refusal also holds at the individual level: with the exception of Malawi, the odds for refusing an HIV test are higher in individuals who have been tested before (irrespective of the outcome of the test result, Table 1). This probably means that consent for testing is informed by HIV status and the respondents' prior knowledge of it. The DHS do not, however, allow us to investigate that relationship further because the HIV status of those that refuse the test is unobserved (one just knows whether the respondent has been tested before or not). As an alternative, we estimate this parameter using MDICP data.

Of the 3,284 respondents that were approached for a test in MDICP3, 90.8% consented. Among those, 67.2% came back for post-test counseling. Of the respondents who were tested and received their results in MDICP3, 76.9% were successfully contacted again in MDICP4 (1,462 or 78.5% of the HIV negatives, and 67 or 52.3% of the HIV positives). In that group of respondents, the refusal rate for an HIV test in MDICP4 is 4.5% and the relative risk of refusing is 4.62 times higher in HIV positives (95%-CI: 2.60-8.21) [52]. This indicates statistically significant support for the hypothesis that individuals who know they are HIV positive are more likely to refuse testing than individuals who have previously tested negative for HIV.

Bias in HIV prevalence estimates. Non-response rates from the DHS and model inputs are presented in Table 2. National refusal rates (conditional on a completed individual interview) vary from just over 4.4% in Cameroon to 27.4% in Malawi. As is also shown in Figure 1, refusal rates are usually higher in urban areas and among men, thereby increasing the potential for bias

in these population subgroups. The same is true for the level of prior testing rates. In some settings (e.g., urban Cameroon), however, the prior testing rates are considerably higher for women because of the HIV testing policies as part ANC visits. With the exception of Malawi, non-response on the individual and household interviews combined is of the same order of magnitude than refusal for an HIV test conditional on an individual interview. That implies that our model only accounts for part (roughly half) of the non-response in HIV prevalence surveys. This does not, however, mean that our model accounts for half of the non-response bias, because the relationship between other forms of non-response and HIV status may be either larger or smaller.

Adjusted HIV prevalence estimates accounting for refusal bias based on the inputs in Table 2 are presented in Table 2 itself and Figure 2. The bars represent the observed and adjusted HIV prevalence under the medium scenario where HIV positive individuals who know their status are 4.62 times more likely to refuse than HIV negative individuals who know their status. The red dots represent the ratio of the adjusted over observed prevalence. The whiskers on the ratio correspond to values for the ratio estimated under the assumption that HIV positive individuals are between 2.60 and 8.21 times more likely to refuse, or, to the 95% confidence interval around the estimate of the E parameter.

For some countries such as Senegal, Ghana, Lesotho and Cameroon, the adjusted national-level HIV prevalence estimates are not much higher than the values observed in the survey. In these cases, the medium estimate for the ratio indicates that the adjusted prevalence estimate is between 1.5% and 5.0% higher than the observed values. For countries such as Zimbabwe and Malawi, however, the adjusted prevalence is 8.5% (95%-CI: 4.8-12.0) and 13.3% (95%-CI: 7.2-19.6) higher than the observed values, respectively. Put differently, the adjusted

national-level prevalence estimate will be about 1.5 (95%-CI: 0.8-2.1) and 1.6 (95%-CI: 0.8-2.3) percentage points higher than the observed values in Zimbabwe and Malawi, respectively.

National-level figures sometimes conceal considerable heterogeneity by sex and place of residence: because of the higher prior testing and refusal rates, the bias is usually much higher in urban than in rural areas, and for men compared to women. In urban populations in Zimbabwe and Malawi, for example, the difference between the adjusted and observed HIV prevalence is 3.1 (95%-CI: 1.8-4.3) and 5.2 (95%-CI: 2.9-7.3) percentage points, respectively. Similarly, the estimated difference for the adjusted over observed prevalence is usually larger for men than for women, and particularly so in urban areas. In Zimbabwe, for example, the observed ratio of female to male infections in urban areas is 1.35; the adjusted value is 1.24 (95%-CI: 1.20-1.29). At the national-level, these figures are 1.45 and 1.39 (95%-CI: 1.36-1.42), respectively. In Lesotho, the observed ratio in urban areas is 1.54 compared to an adjusted value of 1.42 (95%-CI: 1.38-1.47). The national-level difference is not as important (1.37 versus 1.33 (95%-CI: 1.32-1.35) because bias is much smaller in rural areas where over 80% of the population lives. The differences are not always large (in Senegal and Cameroon there is none), but they tend to over-represent the ratio of female to male infections, and particularly so in urban areas. Bias in the sex ratio of infections is likely to be compounded by other forms of non-response that we do not account for in our model. In all surveys listed in Table 2, for example, the non-response rate for the individual interview is higher for men than for women.

Discussion

In this paper, we challenge the optimistic inclination in a series of studies that argue that non-response bias in HIV prevalence estimates from population-based surveys is small, if not negligible. Most of these studies acknowledge that refusals correlate with observed socio-demographic and behavioral characteristics, but ignore that they may be informed by prior knowledge of one's HIV status. The latter turns out to be important, and might lead to substantial downward bias in HIV prevalence estimates. Under the medium scenario in our model, we estimate that the adjusted national prevalence can be up to 8.5% (Zimbabwe, 95%-CI: 4.8-12.0) or even 13.3% (Malawi, 95%-CI: 7.2-19.6) higher than the observed survey prevalence. The discrepancies are even higher for sub-populations where prior testing and refusal rates are relatively high. This is often the case for urban populations with high prevalence rates. As the most extreme cases, our results suggest that HIV prevalence estimates are underestimated by 21.9 % (95%-CI: 12.0-33.5) for urban males in Zimbabwe, and by 31.6% (95%-CI: 17.4-44.3) and 29.4% (95%-CI: 16.6-40.4) for men and women in urban areas of Malawi, respectively. Because bias is generally larger for men than for women, data from seroprevalence surveys also tend to overestimate the female to male sex ratio of infections.

Interestingly, these findings indicate that urban areas often weigh less than they should in population-based survey estimates of HIV prevalence, whereas they were traditionally overrepresented in ANC based estimates. Because of the difference in seroprevalence estimates in surveys and ANC data, however, UNAIDS concluded that ANC surveillance data in urban areas overestimate the true prevalence, and recommends adjusting ANC data for urban areas downward by a factor of 0.8 to obtain urban HIV prevalence estimates (previously only rural ANC data were adjusted following such a procedure) [14, 53]. Our results indicate that such an

adjustment is not uniformly appropriate. In addition, our finding that bias is largest in populations where prior testing rates are highest further suggests that the potential for bias in seroprevalence estimates might increase in conjunction with efforts for increasing VCT coverage. We should note, however, that extrapolations from these static observations to trends over time need to be made with necessary caution.

Our results also cast doubt on the magnitude of the recent downward revision of the UNAIDS HIV prevalence estimates which incorporated evidence from nationally representative surveys for the first time [53]. In the case of Malawi, for example, national-level adult HIV prevalence estimates were previously estimated at 14.2 for 2003 and 14.1 for 2005 [54]. The 2004 Malawi DHS reported an observed prevalence of 11.8% (95%-CI: 11.0-12.7), and a non-response adjusted estimate of 12.7% (95%-CI: 12.0-13.3) [55]. Following the new UNAIDS guidelines that value estimates from nationally representative seroprevalence surveys more heavily, the 2003 and 2005 HIV prevalence estimates are now reported by the Malawian government as 12.9% and 12.4%, respectively [56]. Our model establishes the 2004 HIV prevalence estimate at 13.2 % (95%-CI: 12.5-13.9), and there are reasons to believe that this estimate is conservative.

Our model is merely suggestive, however, and should not be used for adjusting the HIV prevalence estimates from nationally representative surveys. One of the limitations of this study is that we had access to only one sample for estimating the relative risk of refusal in HIV positives and negatives conditional on prior knowledge of one's HIV status (i.e., the E parameter in our model). This parameter may depend on a variety of conditions such as gender, place of residence, VCT coverage, access to antiretroviral therapy, and the study protocol for the disclosure of the test results to the respondent [57]. In that respect it is important to acknowledge

that the MDICP testing protocol is different from the protocol used in the DHS or AIS. The level of refusal itself may be important as well: in populations where the refusal rate is higher, refusal may be less selective, and not to the same extent informed by prior knowledge.

Despite the uncertainty around E, modest bias is observed even under a conservative estimate of this parameter (what we have labeled the low scenario). Our estimates of bias are also conservative because of a number of other reasons. First, we assume that HIV positives and negatives are as likely to have ever been tested before. Second, we only account for refusals conditional on a completed survey interview (because we require survey information for estimating one of the parameters in our model). This is conservative because HIV positives who know their status are not only more likely to refuse testing, but also to refuse an interview, particularly if it contains discomforting questions about current and prior sexual behavior. Third, our model does not account for sources of bias related to the sampling frame and non-response for other reasons than refusal (e.g., population mobility). Fourth, we do not account for the potential relationship between perceived risk of infection, true HIV status and refusal in the subgroup that has never been tested before.

References

1. Kigadye RM, Klokke A, Nicoll A, Nyamuryekung'e KM, Borgdorff M, Barongo L, *et al.* **Sentinel surveillance for HIV-1 among pregnant women in a developing country: 3 years' experience and comparison with a population serosurvey**. *AIDS* 1993; **7**:849-855.

2. Fylkesnes 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.

3. Fontanet AL, Messele T, Dejene A, Enquselassie F, Abebe A, Cutts FT, *et al.* **Age- and sex-specific HIV-1 prevalence in the urban community setting of Addis Ababa, Ethiopia**. *AIDS* 1998; **12**:315-322.

4. Glynn JR, Buve A, Carael M, Zaba B. **Adjustment of antenatal clinic HIV surveillance data for HIV-associated differences in fertility**. *AIDS* 1999; **13**:1598-1599.

5. Zaba B, Boerma T, White R. **Monitoring the AIDS epidemic using HIV prevalence data among young women attending antenatal clinics: prospects and problems**. *AIDS* 2000; **14**:1633-1645.

6. Glynn JR, Buve A, Carael M, Musonda RM, Kahindo M, Macauley I, *et al.* **Factors influencing the difference in HIV prevalence between antenatal clinic and general population in sub-Saharan Africa**. *AIDS* 2001; **15**:1717-1725.

7. Saphonn V, Hor LB, Ly SP, Chhuon S, Saidel T, Detels R. **How well do antenatal clinic (ANC) attendees represent the general population? A comparison of HIV prevalence from ANC sentinel surveillance sites with a population-based survey of women aged 15-49 in Cambodia**. *Int J Epidemiol* 2002; **31**:449-455.

8. Gregson S, Terceira N, Kakowa M, Mason PR, Anderson RM, Chandiwana SK, Carael M. **Study of bias in antenatal clinic HIV-1 surveillance data in a high contraceptive prevalence population in sub-Saharan Africa**. *AIDS* 2002; **16**:643-652.

9. Changalucha J, Grosskurth H, Mwita W, Todd J, Ross D, Mayaud P, *et al.* **Comparison of HIV prevalences in community-based and antenatal clinic surveys in rural Mwanza, Tanzania**. *AIDS* 2002; **16**:661-665.

10. WHO/UNAIDS. **Reconciling antenatal clinic-based surveillance and population-based survey estimates of HIV prevalence in Sub-Saharan Africa**. Geneva: WHO/UNAIDS; 2003.

11. Garcia-Calleja JM, Zaniewski E, Ghys PD, Stanecki K, Walker N. **A global analysis of trends in the quality of HIV sero-surveillance**. *Sex Transm Infect* 2004; **80 (Suppl 1)**:i25-30.

12. Crampin AC, Glynn JR, Ngwira BM, Mwaungulu FD, Ponnighaus JM, Warndorff DK, Fine PE. **Trends and measurement of HIV prevalence in northern Malawi**. *AIDS* 2003; **17**:1817-1825.

13. Bignami-Van Assche S, Salomon JA, Murray CJL. **Evidence from national population-based estimates of bias in HIV prevalence**. Paper presented at the Population Association of America Annual Meeting. Philadelphia, March 31-April 2 2005.

14. Gouws E, Mishra V, Fowler TB. **Comparison of adult HIV prevalence from national population-based surveys and antenatal clinic surveillance in countries with generalised epidemics: implications for calibrating surveillance data**. *Sex Transm Infect* 2008; **84 (Suppl 1)**:i17-i23.

15.   Montana LS, Mishra V, Hong R. **Comparison of HIV prevalence estimates from antenatal care surveillance and population-based surveys in sub-Saharan Africa**. *Sex Transm Infect* 2008; **84 (Suppl 1)**:i78-i84.

16.   Nicoll A, Stephenson J, Griffioen A, Cliffe S, Rogers P, Boisson E. **The relationship of HIV prevalence in pregnant women to that in women of reproductive age: a validated method for adjustment**. *AIDS* 1998; **12**:1861-1867.

17.   Zaba BW, Carpenter LM, Boerma JT, Gregson S, Nakiyingi J, Urassa M. **Adjusting ante-natal clinic data for improved estimates of HIV prevalence among women in sub-Saharan Africa**. *AIDS* 2000; **14**:2741-2750.

18.   Walker N, Stanecki KA, Brown T, Stover J, Lazzari S, Garcia-Calleja JM*, et al.* **Methods and procedures for estimating HIV/AIDS and its impact: the UNAIDS/WHO estimates for the end of 2001**. *AIDS* 2003; **17**:2215-2225.

19.   Boerma JT, Holt E, Black R. **Measurement of biomarkers in surveys in developing countries: Opportunities and problems**. *Popul Dev Rev* 2001; **27**:303-314.

20.   Boerma JT, Ghys PD, Walker N. **Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard**. *Lancet* 2003; **362**:1929-1931.

21.   Walker N, Grassly NC, Garnett GP, Stanecki KA, Ghys PD. **Estimating the global burden of HIV/AIDS: what do we really know about the HIV pandemic?** *Lancet* 2004; **363**:2180-2185.

22.   Garcia-Calleja JM, Marum LH, Carcamo CP, Kaetano L, Muttunga J, Way A. **Lessons learned in the conduct, validation, and interpretation of national population based HIV surveys**. *AIDS* 2005; **19**:9-17.

23. Pison G, Le Guenno B, Lagarde E, Enel C, Seck C. **Seasonal migration: a risk factor for HIV infection in rural Senegal**. *J Acquir Immune Defic Syndr* 1993; **6**:196-200.

24. Quinn TC. **Population migration and the spread of types 1 and 2 human immunodeficiency viruses**. *Proc Natl Acad Sci USA* 1994; **91**:2407-2414.

25. Nunn AJ, Wagner HU, Kamali A, Kengeya-Kayondo JF, Mulder DW. **Migration and HIV-1 seroprevalence in a rural Ugandan population**. *AIDS* 1995; **9**:503-506.

26. Decosas J, Kane F, Anarfi JK, Sodji KD, Wagner HU. **Migration and AIDS**. *Lancet* 1995; **346**:826-828.

27. Lurie MN, Williams BG, Zuma K, Mkaya-Mwamburi D, Garnett G, Sturm AW*, et al.* **The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners**. *Sex Transm Dis* 2003; **30**:149-156.

28. Lagarde E, Schim van der Loeff M, Enel C, Holmgren B, Dray-Spira R, Pison G*, et al.* **Mobility and the spread of human immunodeficiency virus into rural areas of West Africa**. *Int J Epidemiol* 2003; **32**:744-752.

29. Lydie N, Robinson NJ, Ferry B, Akam E, De Loenzien M, Abega S. **Mobility, sexual behavior, and HIV infection in an urban population in Cameroon**. *J Acquir Immune Defic Syndr* 2004; **35**:67-74.

30. Coffee MP, Garnett GP, Mlilo M, Voeten HA, Chandiwana S, Gregson S. **Patterns of movement and risk of HIV infection in rural Zimbabwe**. *J Infect Dis* 2005; **191 Suppl 1**:S159-167.

31. Marston M, Harriss K, Slaymaker E. **Non-response bias in estimates of HIV prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys**. *Sex Transm Infect* 2008; **84 (Suppl 1)**:i71-i77.

32.   Hull HF, Bettinger CJ, Gallaher MM, Keller NM, Wilson J, Mertz GJ. **Comparison of HIV-antibody prevalence in patients consenting to and declining HIV-antibody testing in an STD clinic**. *JAMA* 1988; **260**:935-938.

33.   Schwarcz SK, Bolan GA, Kellogg TA, Kohn R, Lemp GF. **Comparison of voluntary and blinded human immunodeficiency virus type 1 (HIV-1) seroprevalence surveys in a high prevalence sexually transmitted disease clinic population**. *Am J Epidemiol* 1993; **137**:600-608.

34.   Jones JL, Hutto P, Meyer P, Dowda H, Gamble WB, Jr., Gunn RA. **HIV seroprevalence and reasons for refusing and accepting HIV testing**. *Sex Transm Dis* 1993; **20**:334-337.

35.   Groseclose SL, Erickson B, Quinn TC, Glasser D, Campbell CH, Hook EW. **Characterization of patients accepting and refusing routine, voluntary HIV antibody testing in public sexually transmitted disease clinics**. *Sex Transm Dis* 1994; **21**:31-35.

36.   Simon PA, Weber M, Ford WL, Cheng F, Kerndt PR. **Reasons for HIV antibody test refusal in a heterosexual sexually transmitted disease clinic population**. *AIDS* 1996; **10**:1549-1553.

37.   Paget WJ, Zwahlen M, Eichmann AR. **Voluntary confidential HIV testing of STD patients in Switzerland, 1990-5: HIV test refusers cause different biases on HIV prevalences in heterosexuals and homo/bisexuals. Swiss Network of Dermatovenereology Policlinics**. *Genitourin Med* 1997; **73**:444-447.

38.   Coulibaly D, Msellati P, Dedy S, Welffens-Ekra C, Dabis F. **Attitudes et comportements des femmes enceintes face au dépistage du HIV à Abidjan (Côte

D'Ivoire) en 1995 et 1996. Raisons du refus du test et indifference face aux résultats. *Sante* 1998; **8**:234-238.

39.   Boxall EH, Smith N. **Antenatal screening for HIV; are those who refuse testing at higher risk than those who accept testing?** *J Public Health (Oxf)* 2004; **26**:285-287.

40.   Mseleku M, Smith TH, Guidozzi F. **HIV seropositive in pregnant South African women who initially refuse routine antenatal HIV screening**. *BJOG* 2005; **112**:370-371.

41.   Meda N, Zoundi-Guigui MT, van de Perre P, Alary M, Ouangre A, Cartoux M, *et al.* **HIV infection among pregnant women in Bobo-Dioulasso, Burkina Faso: comparison of voluntary and blinded seroprevalence estimates**. *Int J STD AIDS* 1999; **10**:738-740.

42.   Mpairwe H, Muhangi L, Namujju PB, Kisitu A, Tumusiime A, Muwanga M, *et al.* **HIV risk perception and prevalence in a program for prevention of mother-to-child HIV transmission: comparison of women who accept voluntary counseling and testing and those tested anonymously**. *J Acquir Immune Defic Syndr* 2005; **39**:354-358.

43.   Fabiani M, Nattabi B, Ayella EO, Ogwang M, Declich S. **Using prevalence data from the programme for the prevention of mother-to-child-transmission for HIV-1 surveillance in North Uganda**. *AIDS* 2005; **19**:823-827.

44.   Obare F. **The effect of non-response on population-based HIV prevalence estimates: the case of rural Malawi**. *Social Networks Project Working Paper*. Philadelphia: University of Pennsylvania; 2005.

45.   Mishra V, Vaessen M, Boerma JT, Arnold F, Way A, Barrere B, *et al.* **HIV testing in national population-based surveys: experience from the Demographic and Health Surveys**. *Bull World Health Organ* 2006; **84**:537-545.

46.   McNaghten A, Herold JM, Dube HM, St Louis ME. **Response rates for providing a blood specimen for HIV testing an a population-based survey of young adults in Zimbabwe**. *BMC Public Health* 2007; **7**.

47.   Mishra V, Barrere B, Hong R, Khan S. **Evaluation of bias in HIV seroprevalence estimates from national household surveys**. *Sex Transm Infect* 2008; **84 Suppl 1**:i63-i70.

48.   Rice BD, Batzing-Feigenbaum J, Hosegood V, Tanser F, Hill C, Barnighausen T, *et al.* **Population and antenatal-based HIV prevalence estimates in a high contracepting female population in rural South Africa**. *BMC Public Health* 2007; **7**:160.

49.   Bignami-Van Assche S, Smith K, Reniers G, Anglewicz P, Thornton R, Chao LW, *et al.* **Protocol for biomarker testing in the 2004 Malawi Diffusion and Ideational Change Project**. *Social Networks Project Working Papers (6)*. Philadelphia: University of Pennsylvania; 2004.

50.   UN. **World urbanization prospects: The 2007 revision population database**. United Nations Population Division, http://esa.un.org/unup/. Accessed 9 November 2007.

51.   Shisana O, Rehle T, Simbayi L, Parker W, Zuma L, Bhana A, *et al.* **South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey**; 2005.

52. Obare F, Fleming P, Anglewicz P, Thornton R, Martinson F, Kapatuka A*, et al.* **Acceptance of repeat population-based voluntary counseling and testing for HIV in rural Malawi**. under review.

53. UNAIDS/WHO. **2007 AIDS epidemic update**. Geneva: UNAIDS and WHO; 2007.

54. UNAIDS. **2006 Report on the global AIDS epidemic**. Geneva: UNAIDS; 2006.

55. NSO and ORC Macro. **Malawi demographic and health survey 2004**. Zomba, Malawi and Calverton, MD: National Statistical Office [Malawi] and ORC Macro; 2005.

56. Rebublic of Malawi. **Malawi HIV and AIDS monitoring and evaluation report**. Lilongwe: Republic of Malawi, Office of the President and Cabinet; 2007.

57. Reniers G, Araya T, Berhane Y, Davey G, Sanders EJ. **Disclosure of HIV test results and non-response bias in seroprevalence surveys**. Boulder: University of Colorado, Population Program Working Papers (POP2007-05); 2007.

58. NSO [Malawi] and ORC Macro. **Malawi Demographic and Health Survey 2004.** Calverton, Maryland: National Statistical Office [Malawi] and ORC Macro; 2005.

59. Central Statistical Office [Zimbabwe], Macro International Inc. **Zimbabwe Demographic and Health Survey 2005-06**. Calverton, Maryland: CSO and Macro International Inc.; 2007.

60. Institut National de la Statistique [Cameroun], ORC Macro. **Enquête Démographique et de Santé du Cameroun 2004**. Calverton, Maryland, USA: INS et ORC Macro; 2004.

61. Ministry of Health and Social Welfare [Lesotho], Bureau of Statistics [Lesotho], ORC Macro. **Lesotho Demographic and Health Survey 2004**. Calverton, Maryland: MOH, BOS, and ORC Macro; 2005.

62.   Ndiaye S, Ayad M. **Enquête Démographique et de Santé au Sénégal 2005**. Calverton,

        Maryland, USA: Centre de Recherche pour le Développement Humain [Sénégal] et ORC

        Macro; 2006.

63.   Ghana Statistical Service, Noguchi Memorial Institute for Medical Research, ORC

        Macro. **Ghana Demographic and Health Survey 2003**. Calverton, Maryland, USA:

        GSS, NMIMR, and ORC Macro; 2004.

Figure 1: Prior testing and refusal rates in 14 sub-Saharan African countries by HIV prevalence (size of the circles), disaggregated by rural/urban residence and sex (in %)
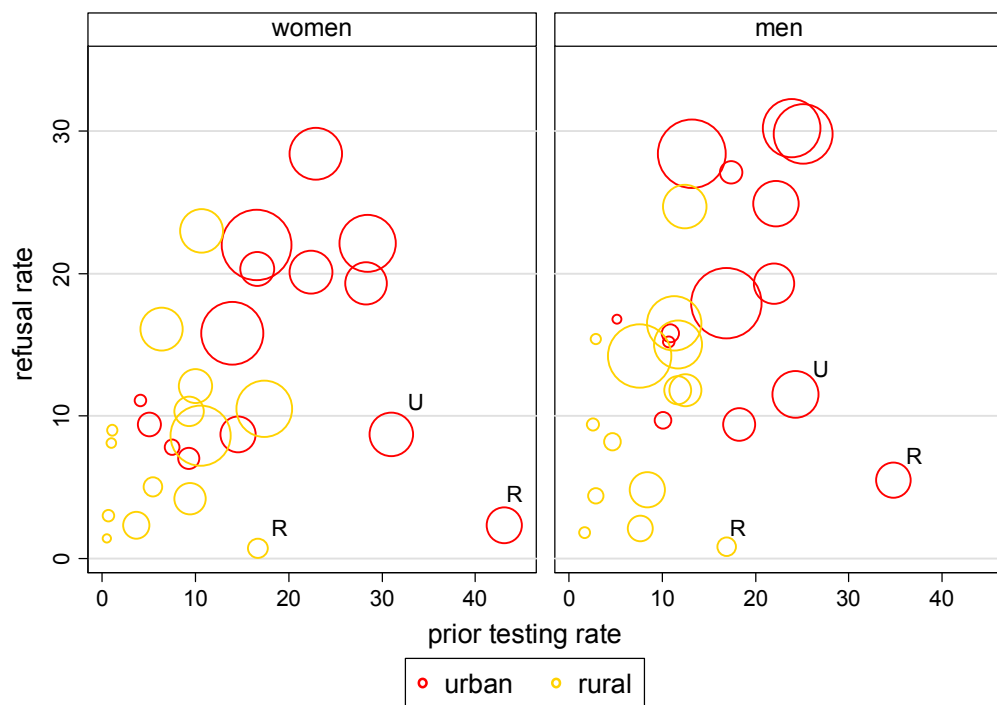
Notes: countries included in the graph are Cameroon (2004), Ethiopia (2005), Ghana (2003), Guinea (2005), Kenya (2003), Lesotho (2004), Malawi (2004), Niger (2006), Rwanda (2005), Tanzania (2003-04), Senegal (2005), Uganda (2004-05), Zambia (2002) and Zimbabwe (2005-06). The circles labeled with an R or U denote subpoplations in Rwanda and Uganda. Compared to the other populations, they are characterized by relatively low refusal rates for the level of prior testing. The prior testing rate is defined as the percentage of respondents who have ever been tested for HIV and received the results of the last test. In Zambia and Tanzania, it is the percentage that has been tested before (irrespective of post-test counseling). The refusal rate is defined as the percentage of respondents that refused the HIV test conditional on having completed the survey interview. Source: Demographic and Health Surveys and AIDS Indicator Surveys.

Table 1: Individual-level relationship between refusal, prior testing status and HIV status

	Refusal by prior testing status OR (95%- Conf. Interval)	Prior testing by HIV status OR (95%- Conf. Interval)
Cameroon (2004)	2.41 (1.90-3.04)	1.80 (1.40-2.31)
Ghana (2003)	1.46 (1.12-1.90)	1.77 (1.05-2.98)
Lesotho (2004)	1.34 (1.00-1.79)	2.09 (1.58-7.66)
Malawi (2004)	0.96 (0.79-1.16)	1.19 (0.88-1.60)
Senegal (2005)	1.97 (1.18-3.30)	2.58 (0.75-8.84)
Zimbabwe (2005-06)	1.15 (1.02-1.30)	1.38 (1.20-1.60)

Notes:
Confidence intervals are based on robust standard errors (adjusted for the non-independence of observations within survey clusters).

Table 2: Distribution of non-response rates and model inputs by type of place of residence and sex (in %)

|	Urban Male	Urban Female	Rural Male	Rural Female	Total	Urban Male	Urban Female	Rural Male	Rural Female	Total
	Senegal (2005)					Ghana (2003)				
Non-response: HH interview [a]	3.5	2.8	0.6	0.6	1.8	2.1	2.1	0.7	0.7	1.3
Non response: Ind Interview [a]	14.5	6.5	13.6	6.2	9.8	7.8	5	5.2	3.8	5.2
Refused HIV test [b]	7.5	7.1	5.9	6.3	6.7	15.3	5.8	8.1	4.2	7.9
Model inputs: [b]										
Prior test [c]	5.4	4.9	3.0	1.5	3.7	10.9	9.4	4.7	5.7	7.3
Refused HIV test - Prior test	10.5	15.7	8.5	7.3	11.9	22.9	4.1	8.4	5.7	10.8
Refused HIV test - No prior test	7.3	6.3	5.7	6.2	6.4	14.3	5.9	8.2	4.1	7.7
Observed HIV prevalence [d]	0.4	1.0	0.5	0.8	0.7	1.6	2.9	1.7	2.5	2.2
Model outputs – Adj. HIV prevalence										
Low	0.4	1.0	0.5	0.8	0.7	1.7	2.9	1.7	2.6	2.2
Medium	0.4	1.0	0.5	0.8	0.7	1.8	2.9	1.7	2.6	2.2
High	0.4	1.0	0.5	0.8	0.7	1.9	3.0	1.7	2.6	2.3
	Cameroon (2004)					Malawi (2004)				
Non-response: HH interview [a]	4.6	3.6	1.5	1.3	2.8	4.9	2.0	4.2	1.9	3.1
Non response: Ind Interview [a]	9.5	7.1	4.3	4.3	6.3	19.8	5.4	13.0	4.2	9.1
Refused HIV test [b]	6.6	6.5	1.6	1.7	4.4	41.0	39.2	25.7	24.5	27.4
Model inputs: [b]										
Prior test [c]	19.0	27.4	8.0	10.2	17.9	25.1	23.3	12.4	11.0	13.6
Refused HIV test - Prior test	10.6	10.0	2.3	2.8	8.38	35.2	39.4	19.6	25.2	27.0
Refused HIV test - No prior test	5.7	5.2	1.5	1.7	3.66	43.3	38.9	26.4	24.1	27.9
Observed HIV prevalence [d]	4.7	8.2	2.9	4.8	5.1	16.4	18.0	8.9	12.5	11.7
Model outputs – Adj. HIV prevalence										
Low	4.9	8.5	2.9	4.8	5.2	19.2	21.0	9.2	13.0	12.5
Medium	5.0	8.8	2.9	4.8	5.4	21.5	23.3	9.6	13.5	13.2
High	5.3	9.2	2.9	4.8	5.6	23.6	25.3	10.1	14.1	13.9
	Lesotho (2004)					Zimbabwe (2005-06)				
Non-response: HH interview [a]	11.7	10.5	3	2.6	4.9	5.9	5.9	4.6	4.6	5.1
Non response: Ind Interview [a]	12.3	4.2	16.3	6.2	10.2	28.1	14.9	11.7	6.6	13.7
Refused HIV test [b]	36.4	28.9	12.6	8.0	15.1	29.4	22.2	15.9	9.9	17.7
Model inputs: [b]										
Prior test [c]	13.3	16.7	8.3	11.5	11.9	24.0	28.8	11.4	17.5	19.5
Refused HIV test - Prior test	39.4	17.5	21.5	11.2	18.8	31.1	20.5	19.0	9.4	19.3
Refused HIV test - No prior test	36.2	31.0	10.9	7.5	14.7	28.7	23.0	15.2	10.0	17.2
Observed HIV prevalence [d]	21.4	33.1	18.2	24.3	22.3	16.0	21.6	13.9	20.8	17.9
Model outputs – Adj. HIV prevalence										
Low	23.1	34.0	18.6	24.6	22.8	18.0	23.2	14.3	21.2	18.7
Medium	24.4	34.6	18.9	24.8	23.2	19.6	24.3	14.7	21.4	19.4
High	25.4	35.0	19.2	25.0	23.5	21.1	25.3	15.1	21.7	20.0

Notes:
[a] Sources: Tables A1 and A2 in [58-63].
[b] Sources: tabulated from raw DHS datasets. Survey weights are used for calculating prior testing and refusal rates, HIV weights are used for calculating HIV prevalence. Reported figures are conditional on a completed interview. The age range for women is from 15-49. For men it is15-59 for Cameroon, Ghana, Lesotho, Senegal, and15-54 for Malawi and Zimbabwe.
[c] Defined as the percentage of respondents who reported to have been tested before and to have received the results of the test.
[d] The national HIV prevalence is calculated as a weighted average of the prevalence of each population subgroup. The population distribution is taken from the UN World Population Prospects [50], and the sex ratios are assumed to be in balance. These are not necessarily realistic assumptions, but they allow for comparison with the adjusted country level HIV prevalence estimate which is calculated in the same fashion.

Figure 2: Observed HIV prevalence, adjusted HIV prevalence, and their ratio, by sex and place of residence
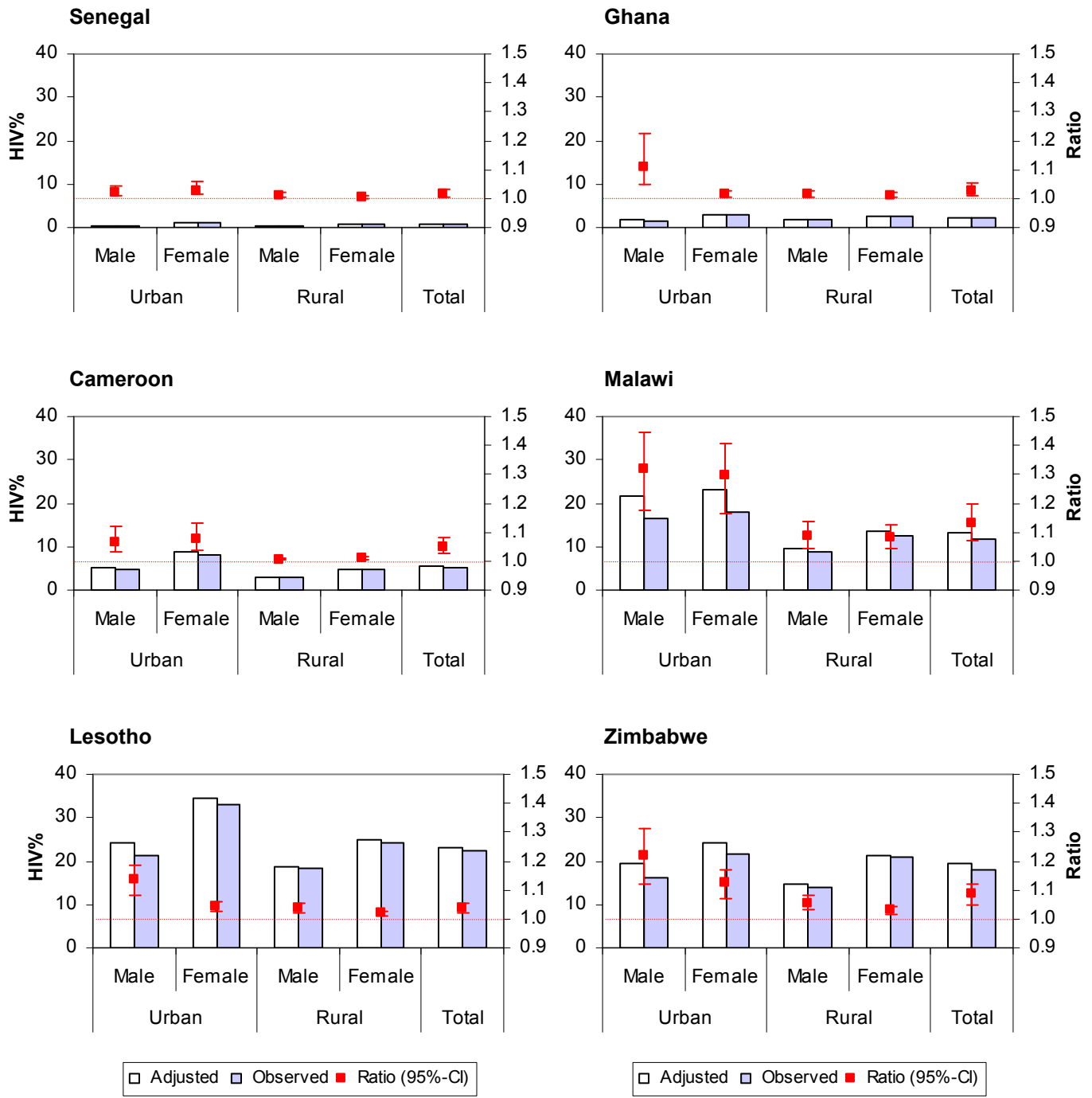
Appendix:

Let us define the event $H = 1$ to be the event that an individual is HIV positive, $T = 1$ be the even that an individual has been tested and knows his or her HIV status, and $R = 1$ be the event that an individual refuses to be tested in a randomized sample. The prevalence estimated in the sero-sample is $P(H = 1 | R = 0)$, that is the probability that an individual is HIV positive given he or she did not refuse, that is he or she participated in the sample. (Note that this ignores individuals who are absent for the sample. Under the *conservative* assumption that absenteeism is independent of HIV status this yields the same results as conditioning on absenteeism.)

We are interested in finding an equation for $P(H = 1)$, the true population prevalence of HIV. In addition to the sample HIV prevalence, from the DHS data, we can estimate the following quantities

$P(T = 1)$, the proportion of the population tested previously and knows their HIV status,

$P(R = 1 | T = 1)$, the probability that an individual refuses given that he or she knows his or her HIV status.

$P(R = 1 | T = 0)$, the probability that an individual refuses given that he or she does not know his or her HIV status.

Recall that from the MDICP data we can estimate the relative risk of refusal for individuals who know that they are HIV positive compared to individuals who know that they are HIV negative, that is

$$E = \frac{P(R = 1 | H = 1, T = 1)}{P(R = 1 | H = 0, T = 1)}. \tag{1}$$

We further assume that being tested previously does not depend on one's HIV status, mathematically that is $P(H = 1 | T = 1) = P(H = 1)$. We note that this is a conservative assumption as several sources have shown that HIV positive individuals are more likely to have been tested. Also we assume that HIV status does not influence refusal for individuals who do not know their HIV status, that is

$$P(R = 1 | H = 1, T = 0) = P(R = 1 | H = 0, T = 0) = P(R = 1 | T = 0).$$

With these assumptions, we can use the rules of conditional probability to find an equation relating our unknown quantity $P(H = 1)$ to known probabilities.

We start using the law of total probability to express

$$P(H = 1 | R = 0) = P(H = 1 | T = 0, R = 0)P(T = 0 | R = 0) + P(H = 1 | T = 1, R = 0)P(T = 1 | R = 0) \tag{2}$$

Now we will write each component of the sum in terms of quantities that we are given. Recalling our assumption that $P(H = 1 | T = 1) = P(H = 1)$ we see that

$$P(H = 1 | T = 0, R = 0) = \frac{P(R = 0 | H = 1, T = 0)P(H = 1 | T = 0)}{P(R = 0 | T = 0)} = P(H = 1). \tag{3}$$

Next, Bayes theorem yields

$$P(T = 0 \mid R = 0) = \frac{P(R = 0 \mid T = 0)P(T = 0)}{P(R = 0 \mid T = 0)P(T = 0) + P(R = 0 \mid T = 1)P(T = 1)}. \qquad (4)$$

For the third term, recalling (1), a bit of algebra shows that

$$P(H = 1 \mid T = 1, R = 0) = \frac{\left[1 - \dfrac{E \cdot P(R = 1 \mid T = 1))}{1 + (E - 1) \cdot P(H = 1)}\right] \cdot P(H = 1)}{1 - P(R = 1 \mid T = 1)}. \qquad (5)$$

And finally Bayes Theorem gives that

$$P(T = 1 \mid R = 0) = \frac{P(R = 0 \mid T = 1)P(T = 1)}{P(R = 0 \mid T = 0)P(T = 0) + P(R = 0 \mid T = 1)P(T = 1)} \qquad (6)$$

.

Substituting (3) - (6) into (2) yields a function including $P(H = 1)$ and quantities which are known from the DHS survey. Rearranging terms in the equation yields a quadratic equation in $P(H = 1)$. It can be shown that exactly one of the roots of the equation will be in the interval $[0,1]$, the estimate of $P(H = 1)$, the population HIV prevalence.