Evidence from National Population-Based Surveys on Bias in Antenatal Clinic-Based Estimates of HIV Prevalence

Simona Bignami-Van Assche
Joshua A. Salomon
Christopher J.L. Murray

Harvard University Initiative for Global Health
104 Mt Auburn St, Cambridge, Massachusetts

Paper to be presented at the 2005 Meeting of the Population Association of America, Philadelphia, Pennsylvania

Abstract

National HIV estimates in most developing countries with generalized epidemics (defined as a prevalence of at least 1% in the general population) are based on data generated by surveillance systems that focus on pregnant women attending a selected number of antenatal clinics (ANC). Recently, an increasing number of population-based studies have become available, allowing for critical assessment of the validity of using ANC-based surveillance data to estimate HIV prevalence in the general adult female and male populations. In this paper, we evaluate potential sources of bias in national estimates of adult HIV prevalence based on ANC data, using recent nationally representative, population-based surveys in Burkina Faso, Burundi, Dominican Republic, Ghana, Kenya, Mali, Niger, South Africa, Zambia, and Zimbabwe.

1. Introduction

Data on HIV prevalence may be collected in population-based surveys or by monitoring sentinel populations. In developing countries with generalized epidemics (defined as a prevalence of at least 1% in the general population), national HIV estimates are generally based on data collected through sentinel surveillance systems that focus on pregnant women attending a selected number of antenatal clinics (ANC). In estimating HIV prevalence at the national level from antenatal clinic data three main assumptions are made (WHO/UNAIDS, 2003). First, it is assumed that prevalence among pregnant women offers a reasonable approximation of prevalence among the general adult population. Second, to account for under-representation of more remote rural clinics, estimates outside of major urban areas are adjusted downward by a factor of 20% in most countries. Third, for computing sex-specific estimates of HIV prevalence, it is assumed that the female-to-male ratio of HIV prevalence is 1.2 to 1 in mature epidemics.

An increasing number of community-based studies have revealed the limitations inherent in extrapolating from antenatal clinic-based surveillance data to adult HIV prevalence in the general female population (e.g. Borgdorff *et al*., 1993; Kigadye *et al.*, 1993; Kilian *et al.*, 1999; Fylkenes *et al*, 1998; Kwesigabo *et al.*, 2000). Comparisons of HIV prevalence between ANC and population-based samples in selected communities show that women attending antenatal clinics may have higher HIV prevalence than the general female population in the same age groups, mostly because of reduced fertility among HIV-infected women (Zaba, Boerma and White, 2000). More limited information is available for assessing the validity of standard assumptions on urban/rural and

male/female differentials in HIV prevalence. Consequently, it is currently not possible to fully assess possible biases in ANC-based estimates of HIV prevalence.

In this paper, we evaluate potential sources of bias in national estimates of adult HIV prevalence from ANC-based surveillance data, by using recent nationally representative, population-based surveys in Burkina Faso, Burundi, Dominican Republic, Ghana, Kenya, Mali, Niger, South Africa, Zambia, and Zimbabwe. First, we compare HIV prevalence at the national level from these surveys with estimates inferred from ANC data. Then, by focusing on Kenya (where it is possible to link the individual HIV results with the survey results), we explore potential causes for upward biases in ANC-based estimates of HIV prevalence. By using a probit model with sample selection, we are also able to consider whether nonresponse exerts a significant downward bias on population-based estimate of HIV prevalence.

2. Background

Biases in ANC-based estimates of HIV prevalence have been extensively documented using population-based community surveys carried out in the 1990s in Tanzania (Borgdorff *et al*., 1993; Kigadye *et al.*, 1993; Kwesigabo *et al.*, 1996; Kwesigabo *et al.*, 2000; Changalucha, *et al.* 2002), Uganda (Kilian, *et al*. 1999), Zambia (Fylkenes *et al.*, 1998; 2001), Ethiopia (Fontanet, *et al.* 1998), Zimbabwe (Gregson, *et al*., 2002), Malawi (Crampin, *et al*. 2003) and other countries in sub-Saharan Africa (Glynn, *et al.*, 2001).

The general conclusion from these studies was that HIV prevalence in pregnant women attending ANC underestimates HIV prevalence in the general female population

in all but the youngest age group, where ANC prevalence tends to exceed general population prevalence.[1] These studies also indicated that biases in ANC-based estimates of HIV prevalence depend both on socio-demographic differences between women attending antenatal clinics and the general female population, and on the association of HIV and fertility. First, the catchment population of ANC clinics is often ill-defined, and discrepancies with population prevalences may sometimes reflect a mismatch between the two populations. Second, pregnant women attending ANC services tend to have different socioeconomic characteristics from the pregnant women who do not attend. For example, women with higher educational achievement are likely to be underrepresented in ANC-based data, mainly because of their lower pregnancy rates. Third, women who are infertile, not sexually active or using contraceptives will be under-represented in ANC clinics, and the prevalence of HIV may be different in such women. There is good evidence that women with HIV have lower fertility and are less likely to become pregnant, and hence will be underrepresented in ANC surveys (Glynn, *et al.* 1999). A marked selection bias in higher age-groups due to lower pregnancy rates in HIV-positive than in HIV-negative women appears to be the most important contributing factor in discrepancies between ANC and population-based estimates of HIV prevalence. Despite these problems, existing studies indicate that, when employing the general population of men and women as a comparison, the urban surveillance-based overall estimate matched, although it tended to underestimate prevalence in rural areas (Fylkenes, *et al*., 1998).

Data from an increasing number of national population-based surveys provide a unique opportunity to evaluate biases in ANC-based estimates of HIV prevalence. Recent

[1] This is not the case only in the Ethiopian study, where ANC-based estimates are higher than population-based estimates of HIV prevalence at all ages (Fontanet, *et al*. 1998; Table 3).

technological developments (such as the use of blood-spotted filter paper or oral mucosal transudate for sample collection) have facilitated the collection of biological data in sample surveys, and in the past few years several countries have conducted national population-based household surveys that include HIV testing. HIV data collection has been part of specific AIDS surveys (such as in South Africa, Niger, Burundi and Zimbabwe or more general demographic health surveys (such as in Mali, Zambia, Dominican Republic, Kenya, Burkina Faso, and Ghana).[2] HIV data collection in national representative surveys has the main objective to obtain national and sub-national estimates of HIV prevalence, which provide the opportunity of evaluating corresponding ANC-based estimates of HIV prevalence.

3. Data and Methods

Data

The data for the analysis come from recent nationally representative, population-based surveys carried out in Burkina Faso, Burundi, Dominican Republic, Ghana, Kenya, Mali, Niger, South Africa, Zambia, and Zimbabwe between 2000 and 2003 (Table 1).

[Table 1 about here]

Burundi conducted a national HIV prevalence survey of all persons of at least 12 years of age in 2001. The survey sample design was based on three strata (urban, semi-urban and rural) and HIV tests were performed from venous blood with ELISA and

[2] In almost all surveys data are collected on other aspects, including risk behavior, program coverage and AIDS attitude and knowledge. Information on other sexually transmitted infections (STIs) has also been sought.

Determine/Genscreen as confirmatory tests (CEFORMI, 2002). In Zimbabwe, a national young adult survey was conducted in 2001/2002 among people age 15-29 years. Two ELISA tests were used on dried blood spots, Thermo Lab Systems and Welllcozyme HIV1+2 GacELISA. Tests were repeated on discordant results and repeat discordant results were tested with Western Blot. In South Africa a national survey of people aged two years and over was conducted in 2002. Oral fluid samples for HIV were collected using Orasure HIV-1 and tested using a single Vironostika test (Shisana and Simbayi, 2002). Niger also conducted a national survey on HIV prevalence in 2002 among the adult population age 15-49 years. Tests were performed on dried blood spots with Genscreen HIV1+2, plus Vironostika HIV Uniform II+O for confirmation (Louboutin-Croc, *et al*., 2002).

In Mali, Zambia and the Dominican Republic a Demographic and Health Survey (DHS) was carried out in 2001/2002; in Kenya, Burkina Faso, and Ghana a national DHS was completed in 2003. HIV tests were performed from dried blood spot samples using ELISA tests and confirmatory Western Blot for indeterminate cases, with the exception of the Dominican Republic, where oral fluid samples for HIV were collected using Orasure HIV-1 and tested using a single Vironostika test (Cellule de Planification et de Statistique du Ministère de la Santé, Direction Nationale de la Statistique et de l'Informatique et ORC Macro, 2002; Achecar, *et al*., 2003; Zambia Central Statistical Office, Zambia Central Board of Health, and ORC Macro, 2003; Ghana Statistical Service, Noguchi Memorial Institute for Medical Research, and ORC Macro, 2004; Institut National de la Statistique et de la Démographie et ORC Macro, 2004; Kenya Central Bureau of Statistics, Kenya Ministry of Health, and ORC Macro, 2004). In

Burkina Faso and Ghana, HIV testing included all eligible respondents (i.e. all eligible women age 15-49 and all eligible men 15-59). In the Dominican Republic, HIV testing included all eligible women 15-49 in half of the survey households, and all eligible men 15-59 in the other half. In Mali and Zambia HIV testing included all women age 15-49 and men age 15-59 in one-third of the survey households.

In most cases, tests were performed anonymously, so that it is not possible to link the individual HIV results with the corresponding survey data to explore what individual characteristics and behaviors affect the probability of being HIV positive. A notable exception is represented by the three most recent DHS. Among them, Kenya is a particularly interesting case because of its high HIV prevalence, which allows statistical analyses at high levels of disagreggation.[3]

Methods

The analysis is divided in two parts. The first part is descriptive, and reviews the evidence that suggests the existence of biases in national estimates of adult HIV prevalence from ANC-based surveillance, by using data from all population-based surveys described above. In the second part, we take advantage of the design of the Kenyan survey to investigate in detail the underlying reasons for discrepancies between ANC- and population-based estimates of HIV prevalence. Specifically, we fit probit models with sample selection to the Kenyan data to identify significant predictors of the probability of being HIV positive.

The probit model with sample selection (Van de Ven and Van Pragg, 1981) assumes that there exists an underlying relationship:

[3] Data from the South Africa National HIV study (which would have the same advantages of the Kenyana data) were not available to us.

$$y_j^* = x_j\beta + u_{1j} \qquad \text{(latent equation)}$$

such that we observe the binary outcome:

$$y_j^{\text{probit}} = (y_j^* > 0) \qquad \text{(probit equation)}$$

The dependent variable, however, is not always observed. Rather the dependent variable for observation *j* is observed only if:

$$y_j^{\text{select}} = (z_j\gamma + u_{2j} > 0) \qquad \text{(selection equation)}$$

where: $u_1 \sim N(0, 1)$; $u_2 \sim N(0, 1)$; $corr(u_1, u_2) = \rho$. When $\rho \neq 0$, standard techniques applied to the first equation yield biased results. The probit model with sample selection provides consistent, asymptotically efficient estimates for all the parameters in such model.

In our case, the first equation describes the probability of being HIV positive. Respondents chose whether to be tested for HIV in the survey, and thus, from our point of view, whether we observed their HIV status. If respondents made this decision randomly, we could ignore the fact that the HIV status of the respondent is not observed in all cases and fit an ordinary probit model to the data. However, the assumption of random participation in HIV testing is unlikely to be true. Indeed, it has been emphasized that survey-based estimates of HIV prevalence might be biased downwards because of refusal of HIV testing and non-inclusion of more mobile individuals (Boerma, *et al*., 2003). Given the observed higher risk of infection in mobile individuals (Nunn, *et al*, 1995; Decosas, *et al*. 1995), the most likely scenario is that absentees are found with a higher HIV prevalence than survey participants. Accordingly, the survey might underestimate male HIV prevalence and thus overestimate the male–female differentials. (Fylkenes et al. 1998). To account for non-random participation in HIV testing, we

therefore use a probit model with sample selection.[4] The survey data include respondents who were interviewed but not tested for HIV (because they were away or they refused), so that the outcome of the probit model with sample selection can be interpreted as if we observed HIV status for all respondents in the sample.

4. The existing evidence on discrepancies between ANC-based and population-based estimates of HIV prevalence

In general, HIV prevalence at the national level estimated from the population-based surveys described earlier has been found to be considerably lower than that inferred from ANC data (Table 2 and Figure 1). This finding has stirred a debate about whether UNAIDS and WHO have so far overestimated the size of the epidemic in these countries, or even in sub-Saharan Africa as a whole (Boerma, *et al.* 2003). HIV testing in population-based surveys has a higher rate of non-response than does testing for other outcomes, and is likely to miss many members of mobile or immigrant populations and other groups at an increased risk of HIV-1. As a result, population-based surveys may underestimate true prevalence. It has therefore been emphasized that surveys should be regarded as part of surveillance systems, and a means to improve estimates of HIV-1 prevalence and associated trends. Indeed, recently UNAIDS has begun to revise its national estimates of HIV/AIDS on the basis of the results of population-based surveys (Anne Cross, personal communication).

[Table 2 and Figure 1 about here]

[4] The probit model with sample selection was applied to the Kenyan data by using the heckprob command in STATA 8.1.

Two main points have to be noticed in Table 2. First, although ANC sites are supposed to be representative of urban areas, for half of all countries considered in the table (i.e. Burundi, Mali, Zambia, and Zimbabwe) the difference between ANC- and population-based estimates in urban areas is actually larger than in rural areas. Second, rural prevalence in the survey is always more than 20% lower than the rural prevalence from ANC data. This suggests that the adjustment factor used by UNAIDS and WHO to correct for urban-rural differences in ANC-based estimates of HIV prevalence might be inadequate.

In the next sections, we explore in more detail potential causes of these discrepancies using the example of Kenya.

4. Factors accounting for biases in ANC-based estimates of HIV prevalence: the Kenyan case

There are two main sources of potential upward bias in ANC-based estimates of HIV prevalence. First, HIV-positive women may be more likely to use antenatal clinics than HIV-negative women. Second, antenatal clinics used for sentinel surveillance might be placed in areas with higher HIV prevalence. We discuss each of these possibilities below.

Characteristics of ANC users

When HIV prevalence is tabulated separately for ANC users and non-users in Kenya (Table 3), it is evident that average HIV prevalence for ANC users is higher than for non-users (+2.5%), all sampled women (+1.6%) and, especially, all sampled men (+5.4%). When selected characteristics associated with ANC use are considered, the largest difference in average HIV prevalence between ANC users and non-users is found by age,

urban/rural place of residence, current marital status, number of unions and age at first sex. Differences in average HIV prevalence by region of residence, age at first marriage and having had a birth in the five years preceding the survey are, on the contrary, negligible.

[Table 3 about here]

Given these considerations, the crucial issue in order to be able to correct ANC-based estimates of HIV prevalence is to identify which covariates of ANC use are also associated with HIV status. A simple probit model for the probability of being HIV positive (Table 4) indicates that for females four covariates have a strong impact on the dependent variable: age, urban/rural place of residence, current marital status, and age at first sex. Living in Nyanza province is also significantly associated with higher HIV risk. Having had a birth in the five years before the survey is not significant in the regression model, which contradicts previous findings on the association between HIV and fertility.

[Table 4 about here]

In order to take into account non-response bias, we now turn to a probit model with sample selection. We assume that the probability of being HIV positive is a function of the same covariates as in the standard probit model. We also assume that the probability of being selected for HIV testing depends on individual mobility (measured in terms of the number of years spent by the respondent in her current residence), education, having giving birth in the previous five years, plus the covariates that affect the probability of being HIV positive. The results are shown in Table 5. The results of the probit model with sample selection do not differ considerably from those of the standard

probit model (see Table 4). At the same time, the results of the selection model indicate that there is a selection bias of young female respondents (age 15-19) living in Nyanza province. This suggests that respondents age 30-34 (the reference group) might be more mobile, and thus less easy to be found for the interview and HIV testing, than younger ones in all provinces except Nyanza.[5] It is important to stress that having had a birth during the five years before the survey is not significant in the selection model: in the Kenyan case, it therefore seems that fertility decline associated with HIV infection is not an important factor in explaining biases in ANC-based estimates of HIV prevalence.

[Table 5 and Table 6 about here]

Adjusting for selection bias reduces the predicted probability of being HIV positive by approximately 1.5 percent compared to the standard probit model (Table 6). It is noteworthy that the estimated probability of being HIV positive, but not included in HIV testing, is only 0.6 percent for females. The effect of non-response bias on population-based estimates of HIV prevalence should therefore not be overstated. In addition, according to the probit model with sample selection, the predicted probability of being HIV positive for ANC users in urban and rural areas is, respectively 10.0 and 7.4 percent. These figures are very close to the tabulations presented in Table 3, which suggest that controlling for age, place of residence, current marital status, age at first sex, and non-response does not affect estimates of HIV prevalence in this sample. In other words, these factors do not explain differences between ANC-based and population-based estimates of HIV prevalence in Kenya.

[5] Because of the small number of cases, it is not possible to make a meaningful distinction between respondents who refused to participate in HIV testing and those who were away when the testing was supposed to take place.

Location of ANC clinics

The second possible source of upward biases in ANC-based estimates of HIV prevalence is that antenatal clinics used for sentinel surveillance might be located in areas with higher HIV prevalence. This has been the case in few countries (such as China), where sentinel surveillance sites were initially chosen in areas believed to have higher HIV risk or showing higher HIV prevalence, as an early warning system.

[Table 7 about here]

In Table 7, it can be seen that ANC clinics seem to be quite uniformly distributed across Kenyan provinces. As we have no information about the location of the clinics within each district, we are not able to fully evaluate the magnitude of this source of bias. However, as individual characteristics of ANC users do not seem to explain discrepancies between ANC- and population-based estimates, in the Kenyan case it seems likely that these discrepancies might be due largely to the location of ANC in high HIV prevalence areas.

6. Conclusion

This paper evaluates potential sources of bias in national estimates of adult HIV prevalence from ANC-based surveillance data, by using recent nationally representative, population-based surveys in Burkina Faso, Burundi, Dominican Republic, Ghana, Kenya, Mali, Niger, South Africa, Zambia, and Zimbabwe.

First, we compared HIV prevalence at the national level from these surveys with estimates inferred from ANC data. We found that, in general, HIV prevalence at the

national level estimated from the population-based surveys considered is considerably lower than that inferred from ANC data. In addition, although ANC sites are supposed to be representative of urban areas, for almost half of all countries considered (i.e. Burundi, Mali, Zambia, and Zimbabwe) the difference between ANC- and population-based estimates in urban areas is actually larger than in rural areas. Finally, rural prevalence in the survey is always more than 20% lower than the rural prevalence from ANC data. This suggests that the adjustment factor used by UNAIDS and WHO to correct for urban-rural differences in ANC-based estimates of HIV prevalence might be inadequate.

Then, by focusing on Kenya (where it is possible to link the individual HIV results with the survey results), we explored two main sources of potential upward bias in ANC-based estimates of HIV prevalence. First, HIV-positive women may be more likely to use antenatal clinics than HIV-negative women. We found that this is indeed the case in the descriptive as well as multivariate data analyses. However, we also found that controlling for age, place of residence, current marital status, age at first sex (the four characteristics strongly associated with the probability of being HIV positive) does not explain differences between ANC-based and population-based estimates of HIV prevalence in Kenya. An interesting result of the multivariate analysis by means of a probit model with sample selection is also that, contrary to expectations, in Kenya nonresponse exerts a minimal downward bias on population-based estimate of HIV prevalence. In fact, the estimated probability of being HIV positive, but not included in HIV testing, is only 0.6 percent (for females). The effect of non-response bias on population-based estimates of HIV prevalence should therefore not be overstated. The second possible source of upward biases in ANC-based estimates of HIV prevalence we

explored is that antenatal clinics used for sentinel surveillance might be located in areas with higher HIV prevalence. As we have no information about the location of the clinics within each district, we are not able to fully evaluate the magnitude of this source of bias. However, as individual characteristics of ANC users do not seem to explain discrepancies between ANC- and population-based estimates, in the Kenyan case it seems likely that these discrepancies might be due largely to the location of ANC in high HIV prevalence areas.

References

Achécar MM, Ramírez N, Polanco JJ, Ochoa LH, Lerebours G, García B (2003). *Republica Domicana: Enquesta Demográfica y de salud 2002*. Calverton, Maryland, USA: Centro de Estudios Sociales y Demográficos (CESDEM) and ORC Macro.

Boerma JT, Ghys PD, Walker N (2003). "Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard," *Lancet* **362**(9399):1929-31.

Boisier P, ON, Ouwe Missi Oukem-Boyer ON, Hamidou AA, Sidikou F, Ibrahim ML, Mahamane AE, Mamadou S, Aksenenkova TS, Modibo BH, Chanteau S, Sani A, Louboutin-Croc JP (2004). "Nationwide HIV prevalence survey in general population in Niger," *Tropical Medicine and International Health* **9** (11): 1161–1166.

Borgdorff M, Borongo L, van Jaarsveld E, *et al.* (1993). "Sentinel surveillance for HIV-1 infection: how representative are blood donors, outpatient with fever, anaemia, or sexually transmitted diseases, and antenatal clinic attenders in Mwanza region, Tanzania?," *AIDS* **7**: 567-572.

CEFORMI (Centre de Formation et de Recherche en Medicine et Maladies Infectieuses). 2002. *Enquete national de seroprevalence de l'infection par le VIH au Burundi*. Ministere de la Sante Publique, Ministere a la Presidence Charge de la Lutte contre le Sida et Banque Mondiale: Bujumbura.

Cellule de Planification et de Statistique du Ministère de la Santé (CPS/MS), Direction Nationale de la Statistique et de l'Informatique (DNSI) et ORC Macro (2002). *Enquête Démographique et de Santé au Mali 2001*. Calverton, Maryland, USA: CPS/MS, DNSI et ORC Macro.

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

Crampin AC, Glynn JR, Ngwira BMM, Mwaungulu FD, Ponnighaus JM, Warndorff DK, Fine PEM (2003). "Trends and measurement of HIV prevalence in northern Malawi," *AIDS* **17**: 1817-1825.

Decosas J, Kane F, Anarfi JK, Sodji KDR, Wagner HU (1995). "Migration and AIDS," *Lancet* **346**: 826–828.

Fontanet A, Messele T, Dejene A, *et al*. (1998). "Age- and sex-specific HIV-1 prevalence in the urban community setting of Addis Ababa, Ethiopia," *AIDS* **12**: 315-322.

Fylkenes K, Ndhlovu Z, Kasumba K, Musonda RM, Sichone M (1998). "Studying dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia," *AIDS* **12**:1227-1234.

Fylkesnes K, Musonda RM, Sichone M, Ndhlovu Z, Tembo F, Monze M. (2001). "Declining HIV prevalence and risk behaviours in Zambia: evidence from surveillance and population-based surveys," *AIDS* **15**: 907–916.

Ghana Statistical Service (GSS), Noguchi Memorial Institute for Medical Research (NMIMR), and ORC Macro (2004). *Ghana Demographic and Health Survey 2003.* Calverton, Maryland: GSS, NMIMR, and ORC Macro.

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

Glynn JR, Buve´ A, Carae¨l M, Musonda RM, Kahindo M, Macauley I, *et al*. (2001). "Factors influencing the difference in HIV prevalence between antenatal clinic and general population in sub-Saharan Africa," *AIDS* **15**: 1717–1725.

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

Institut National de la Statistique et de la Démographie (INSD) et ORC Macro (2004). *Enquête Démographique et de Santé du Burkina Faso 2003*. Calverton, Maryland, USA: INSD et ORC Macro.

Kenya Central Bureau of Statistics (CBS), Kenya Ministry of Health (MOH), and ORC Macro (2004). *Kenya Demographic and Health Survey 2003.* Calverton, Maryland: CBS, MOH, and ORC Macro.

Kigadye R-M, Klokke A, Nicoll A, *et al*. (1993). "Sentinel surveillance for HIV-1 among pregnant women in a developing country: 3 years' experience and comparison with a population serosurvey," *AIDS* **7**: 849-855.

Kilian AH, Gregson S, Ndyanabangi B, Walusaga K, Kipp W, Sahlmuller G, *et al*. (1999). "Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda," *AIDS* **13**: 391–398.

Kwesigabo G, Killewo JZJ, Sandstrom A (1996). "Sentinel surveillance and cross-sectional survey on HIV infection prevalence: a comparative study," *East African Medical Journal* **73**:298–302.

Kwesigabo G, Killewo JZJ, Urassa W, *et al*. (2000). "Monitoring of HIV-1 infection prevalence and trends in the general population using pregnant women as a sentinel population: 9 years experience from Kagera region of Tanzania," *J Acquir Immune Defic Syndr* **23**:410-417.

Nunn AJ, Wagner HU, Kamali A, Kengeya-Kayondo JF, Mulder DW (1995). "Migration and HIV-1 seroprevalence in a rural Ugandan Population," *AIDS* **9**:503–506.

Shisana O, Simbayi L (2002). *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.

US Census Bureau. 2003. *HIV/AIDS Surveillance Database, June 2003 Update*. Washington, D.C.: US Census Bureau.

Van de Ven WPMM, BMS Van Pragg (1981). "The demand for deductibles in private health insurance: A probit model with sample selection," *Journal of Econometrics* **17**: 229-252.

World Health Organization (WHO) and UNAIDS (2003). *Reconciling antenatal clinic-based surveillance and population-based survey estimates of HIV prevalence in sub-Saharan Africa*. Geneva, Switzerland: WHO and UNAIDS.

Zaba B, Gregson S. (1998). "Measuring the impact of HIV on fertility in Africa," *AIDS* **12**(suppl 1): S41–S50.

Zaba BW, Boerma JT, White R (2000a). "Monitoring the AIDS epidemic using HIV prevalence data among young women attending antenatal clinics: prospects and problems," *AIDS* **14**:1633-1645.

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

Zambia Central Statistical Office, Zambia Central Board of Health, and ORC Macro (2003). *Zambia Demographic and Health Survey 2001-2002.* Calverton, Maryland, USA: Central Statistical Office, Central Board of Health, and ORC Macro.

Tables and Figures

Table 1: HIV testing sampling strategy, sample characteristics and refusal rates: population-based surveys in 10 countries

Country	Type	Year	Sampling strategy	HIV test	Sample size (women)	Sample size (men)	Refusal rate* (women)	Refusal rate* (men)
Burkina Faso	DHS	2003	All eligible respondents	Anonymously linked	4,575	3,984	7.7%	14.2%
Burundi	HIV survey	2001	Respondents age 12+ years	NA	2,909	2,660	NA	NA
Dominican Republic	DHS	2002	All women in half survey hholds, all men in the other half	Anonymous unlinked	12,514	14,456	11.0%	19.1%
Ghana	DHS	2003	All eligible respondents	Anonymously linked	5,949	5,345	10.7%	20.0%
Kenya	DHS	2003	All respondents in hholds selected for the Men's questionnaire	Anonymously linked	4,303	4,183	13.7%	29.7%
Mali	DHS	2001	One-third of survey hholds	Anonymous unlinked	4,556	4,062	14.8%	24.4%
Niger	HIV Survey	2002	Respondents age 15-49 years	Anonymous unlinked	3,063	2,994	0.9%	1.5%
South Africa	HIV survey	2002	Respondents age 2+ years	Anonymously linked	3,555	2,776	34.9%	41.6%
Zambia	DHS	2001/2	One-third of survey hholds	Anonymous unlinked	2,689	2,418	20.6%	26.7%
Zimbabwe	HIV survey	2001/2	Respondents age 15-29 years	NA	4,263	3,833	11.0%	8.8%

* Among those who were eligible to be tested.

Table 2: Comparison of national population- and ANC-based estimates of HIV prevalence (percent), 10 countries

	Population-based (urban)	**ANC-based (urban)**	**Population-based (rural)**	**ANC-based (rural)**
Burkina Faso	3.1	5.3	1.0	4.2
Burundi	13.0	16.0	2.5	4.5
Dominican Republic	0.9	1.2	1.2	2.2
Ghana	2.3	5.1	2.0	5.1
Kenya	10.0	14.4	5.6	11.6
Mali	2.2	5.8	1.5	3.2
Niger	2.1	2.0	0.6	2.5
South Africa	16.7	27.6	8.3	26.2
Zambia	23.1	26.8	10.8	14.4
Zimbabwe	5.0	30.6	18.0	28.5

Notes: *Burundi*: Semiurban not included.

Sources: *ANC-based*: UNAIDS Epidemiological Factsheets 2004 Update. *Population-based*: Institut National de la Statistique et de la Démographie, et ORC Macro (2004); CEFORMI (2002); Achécar, et al. (2003); Ghana Statistical Service (GSS), Noguchi Memorial Institute for Medical Research (NMIMR), and ORC Macro (2004); Kenya Central Bureau of Statistics, Kenya Ministry of Health, and ORC Macro (2004); Cellule de Planification et de Statistique du Ministère de la Santé, Direction Nationale de la Statistique et de l'Informatique (DNSI) et ORC Macro (2002); Boisier, et al. (2004); Shisana and Simbayi (2002); Zambia Central Statistical Office, Zambia Central Board of Health, and ORC Macro (2003).

Table 3: Average HIV prevalence (%) for ANC users and non-users, and for total female and male population, by selected characteristics (standard deviations in parentheses), Kenya DHS 2003, females

	ANC users		**ANC non-users**		**Total female population**		**Total male population**	
All respondents	10.1	(0.30)	7.6	(0.26)	8.5	(0.28)	4.7	(0.21)
Age group								
15-19	5.2	(0.22)	2.5	(0.16)	3.0	(0.17)	0.3	(0.06)
20-24	11.3	(0.32)	6.6	(0.25)	9.5	(0.29)	2.4	(0.15)
25-29	10.6	(0.31)	18.2	(0.39)	12.8	(0.33)	7.6	(0.26)
30-34	11.7	(0.32)	11.2	(0.32)	11.5	(0.32)	6.7	(0.25)
35-39	8.2	(0.27)	15.9	(0.37)	11.7	(0.32)	8.5	(0.28)
40-44	8.1	(0.27)	10.6	(0.31)	9.8	(0.30)	9.3	(0.29)
45-49	8.5	(0.28)	3.5	(0.18)	4.0	(0.20)	5.6	(0.23)
Place of residence								
Urban	14.3	(0.35)	11.0	(0.31)	12.3	(0.33)	7.6	(0.27)
Rural	9.1	(0.29)	6.5	(0.25)	7.7	(0.27)	3.9	(0.19)
Province								
Nairobi	12.6	(0.33)	11.6	(0.32)	11.9	(0.32)	8.0	(0.27)
Central	7.3	(0.26)	7.9	(0.27)	7.6	(0.27)	2.1	(0.14)
Coast	7.1	(0.26)	6.3	(0.24)	6.6	(0.25)	4.3	(0.20)
Eastern	9.5	(0.29)	2.7	(0.16)	5.9	(0.24)	1.6	(0.12)
Nyanza	18.3	(0.39)	17.9	(0.38)	18.1	(0.39)	12.3	(0.33)
Rift Valley	8.8	(0.28)	4.6	(0.21)	6.8	(0.25)	3.4	(0.18)
Western	6.5	(0.25)	5.1	(0.22)	5.7	(0.23)	3.6	(0.19)
Northeastern	0.0		0.0		0.0		0.0	
Marital status								
Not currently married	18.6	(0.39)	7.6	(0.26)	9.9	(0.30)	2.4	(0.15)
Currently married	8.2	(0.27)	7.6	(0.26)	8.0	(0.27)	7.1	(0.26)
Number of unions								
Married once	9.4	(0.29)	10.4	(0.31)	9.8	(0.30)	6.6	(0.25)
Married twice or more	19.0	(0.39)	12.9	(0.33)	16.5	(0.37)	9.8	(0.30)
Age first marriage								
less than 15	10.0	(0.30)	9.1	(0.29)	9.6	(0.29)	0.0	
15-19	9.4	(0.29)	11.0	(0.31)	10.0	(0.30)	5.4	(0.23)
20-24	12.0	(0.32)	9.9	(0.30)	11.2	(0.31)	8.2	(0.27)
25-29	10.1	(0.30)	9.9	(0.30)	10.0	(0.30)	7.0	(0.26)
30-34	0.0		11.9	(0.32)	9.1	(0.29)	7.7	(0.27)
35-39	0.0		64.1	(0.48)	33.5	(0.47)	10.7	(0.31)
Had birth in past 5 yrs								
No	NA	NA	7.5	(0.26)	7.5	(0.26)	NA	NA
Yes	10.1	(0.30)	7.7	(0.27)	9.9	(0.30)	NA	NA
Age at first sex								
less than 15	16.0	(0.37)	10.3	(0.30)	13.2	(0.34)	6.2	(0.24)
15-19	9.4	(0.29)	11.1	(0.31)	10.1	(0.30)	4.7	(0.21)
20-24	5.0	(0.22)	6.5	(0.25)	5.7	(0.23)	6.6	(0.25)
25-29	10.0	(0.30)	3.0	(0.17)	7.5	(0.26)	5.4	(0.23)
30-34	0.0		0.0		0.0		5.5	(0.23)
Sample size	1446		1825		3271		3578	

Table 4: Probit model for probability of being HIV positive, Kenya DHS 2003, females

Dependent variables	Coef.	SE	p-value
Age group			
15-19	–0.69***	0.14	0.000
20-24	–0.27**	0.11	0.014
25-29	0.03	0.11	0.800
30-34	omitted		
35-39	–0.08	0.12	0.523
40-44	–0.24	0.14	0.093
45-49	–0.57**	0.19	0.003
Place of residence			
Urban	omitted		
Rural	–0.35***	0.09	0.000
Province			
Nairobi	omitted		
Central	0.06	0.14	0.653
Coast	–0.13	0.14	0.362
Eastern	–0.04	0.16	0.821
Nyanza	0.52***	0.13	0.000
Rift Valley	–0.03	0.14	0.838
Western	0.02	0.15	0.911
Northeastern	NA	NA	NA
Marital status			
Not currently married	omitted		
Currently married	–0.39***	0.08	0.000
Number of unions			
Married once	omitted		
Married twice or more	0.24	0.13	0.059
Had birth in past 5 yrs			
No			
Yes	–0.01	0.08	0.885
Age at first sex			
less than 15	0.84***	0.15	0.000
15-19	0.69***	0.14	0.000
20-24	0.36**	0.17	0.034
25-29	omitted		
Ever used modern FP			
No	omitted		
Yes	–0.03	0.07	0.715
Constant	–1.35***	0.17	0.000
Sample size	3115		
LR χ^2	191.93		
Prob (χ^2>0)	[.000]		

Notes: Significant at ***0.1% level; ** 1% level; *5% level.

Table 5: Probit model with sample selection for probability of being HIV positive, Kenya DHS 2003, females: Probit model

Dependent variables	Coef.	SE	p-value
Age group			
15-19	–0.67***	0.14	0.000
20-24	–0.26*	0.11	0.018
25-29	0.03	0.11	0.804
30-34	omitted		
35-39	–0.09	0.12	0.470
40-44	–0.28	0.14	0.042
45-49	–0.54**	0.19	0.004
Place of residence			
Urban	omitted		
Rural	–0.35***	0.07	0.000
Province			
Lives in Nyanza	0.58***	0.09	0.000
Marital status			
Not currently married	omitted		
Currently married	–0.44***	0.08	0.000
Number of unions			
Married once	omitted		
Married twice or more	0.23	0.13	0.069
Age at first sex			
less than 15	0.82***	0.15	0.000
15-19	0.68***	0.14	0.000
20-24	0.35*	0.17	0.037
25-29	omitted		
Constant	–1.38***	0.16	0.000
Sample size	3931		
Wald χ^2	166.67		
Prob (χ^2>0)	[.000]		

Notes: Significant at ***0.1% level; ** 1% level; *5% level.

Table 5 (cont.): Probit model with sample selection for probability of being HIV positive, Kenya DHS 2003, females: Selection model

Dependent variables	Coef.	SE	p-value
Mobility			
Lived at present residence for more than 5 yrs or always	omitted		
Lived at present residence for less than 5 yrs	–0.66	0.41	0.105
Had birth in past 5 yrs			
No	omitted		
Yes	0.42	0.42	0.315
Education			
No education	omitted		
Some education	–0.27	0.48	0.583
Age group			
15-19	1.35*	0.64	0.035
20-24	1.95	1.22	0.110
25-29	1.06	0.58	0.066
30-34	omitted		
35-39	1.24	0.78	0.110
40-44	2.22	2.71	0.412
45-49	1.54	1.25	0.218
Place of residence			
Urban	omitted		
Rural	0.30	0.54	0.576
Province			
Lives in Nyanza	1.88**	0.67	0.005
Marital status			
Not currently married	omitted		
Currently married	0.60	0.47	0.199
Number of unions			
Married once	omitted		
Married twice or more	1.31	1.90	0.491
Age at first sex			
less than 15	-0.85	1.17	0.469
15-19	-1.16	1.14	0.310
20-24	-1.04	1.22	0.395
25-29	omitted		
Constant	1.61	1.16	0.163
ρ	.94	6.23	
LR test of indep. eqns: χ^2	.93		
Prob ($\chi^2>0$)	[.335]		

Notes: Significant at ***0.1% level; ** 1% level; *5% level.

Table 6: Predicted probability of being HIV positive (%): Kenya DHS 2003, females (standard deviations in parentheses)

Model	All females	ANC users	ANC non-users
Standard probit model	8.8 (.07)	9.6 (.07)	8.2 (.08)
Probit model with sample selection			
Selected for HIV testing	7.4 (.07)	8.0 (.07)	6.9 (.07)
Not selected for HIV testing	0.6 (.03)	1.1 (.03)	0.1 (.01)

Table 7: Location of ANC clinics, Kenya 2003

Province (population size[1])	**Location of ANC clinics[2] (U=urban, R=rural)**	**ANC-based estimate of HIV prevalence (median)[3]**	**Population-based estimate of HIV prevalence (females)[4]**
Central Province (3,724,159)	Nyeri district (U, R) Kiambu district (R)	Urban: 15.3 Rural: NA	Urban: 11.5 Rural: 7.1
Coast Province (2,487,264)	Kwale district (U) Mombasa (U)	Urban: 14.4 Rural: ---	Urban: 7.6 Rural: 5.9
Eastern Province (4,631,779)	Kitui district (U) Meru district (U) Embu district (R) Machakos district (R)	Urban: 26.0 Rural: NA	Urban: 13.1 Rural: 5.4
Nairobi (2,143,254)	Nairobi (U)	Urban: 18.4 Rural: ---	Urban: 32.0 Rural: ---
Northeastern Province (962,143)	Garissa district (U)	Urban: 6.3 Rural: ---	Urban: 0.0 Rural: 0.0
Nyanza Province (4,392,196)	Kisii district (U) Kisumu district (U, 2R) Siaya district (2R)	Urban: 25.4 Rural: 31.1	Urban: 28.0 Rural: 16.9
Rift Valley Province (6,987,036)	Nakuru district (U) Kajiado district (R) Uasin Gishu district (R)	Urban: 10.6 Rural: NA	Urban: 11.3 Rural: 5.9
Western Province (3,358,776)	Kakamega district (U) Busia district (U, 2R)	Urban: 17.0 Rural: 14.0	Urban: 14.0 Rural: 4.9

Sources: [1] Government of Kenya (2000). [2,3] US Census Bureau (2003). [4] Kenya DHS (2003).

Figure 1: Comparison of national population- and ANC-based estimates of HIV prevalence (percent), 10 countries