SNP WORKING PAPERS

Social Networks Project Working Papers

No. 7, May 2005

The Effect of Non-Response on Population-Based HIV Prevalence Estimates: The Case of Rural Malawi

Francis Obare

Social
·N·e·t·w·o·r·k·s·
project

University of Pennsylvania

Social Networks Project Working Papers *is an unreviewed and unedited prepublication series reporting on research in progress.*

Social Networks Project Working Papers are circulated for comment and to inform interested colleagues about the work in progress on the survey data collected by the Project in Kenya and Malawi. The Social Networks project is funded by National Institute of Health Grants HD/MH41713 on "Gender, Conversational Networks and Dealing with AIDS" and HD044228 on "AIDS/HIV risk, marriage and sexual relations in Malawi." The Project has also received support from NIH/ORA; NIH/NIA; The Rockefeller Foundation; The Transnational Cooperation Foundation, Government of Germany; The Research Foundation, University of Pennsylvania; and the Carolina Population Center/USAID Evaluation Project.

The views expressed are those of the authors and not necessarily those of the Project, the National Institutes of Health or any other of the Project's funding agencies.

For more information about the Social Networks Project and to submit manuscripts for publication, contact:

Social Networks Project

239 McNeil Building
University of Pennsylvania
3718 Locust Walk
Philadelphia, Pennsylvania 19104

Email: social_networks@pop.upenn.edu
Website: www.pop.upenn.edu/Social_Networks

Introduction

> *"All population-based surveys, including local studies, have non-response bias, and this may affect the validity of the HIV prevalence estimates…The main challenges are to obtain representative sampling of all [young] people, sound testing procedures and good response rates" (WHO and UNAIDS 2003: 6 & 9).*

Population-based surveys that include the collection of biological and clinical data-biomarkers- are increasingly being conducted in developing countries as a means of obtaining population-based health statistics. Various factors have contributed to this development. For instance, the health facility data have been inadequate for assessing the health status of the population owing to low utilization of health services (Boerma et al 2001; Fisher et al. 1996). Advances in medical technology that offer simple, rapid and relatively inexpensive test devices has made it cheaper to conduct population-based biomarker studies (Boerma et al. 2001; Giles et al. 1999; WHO and UNAIDS 2003). The HIV/AIDS pandemic has increased the need for prevalence data that is representative of the whole population to allow for a more detailed evaluation of the magnitude and distribution of the disease (Fylkesnes et al. 1998; WHO and UNAIDS 2003). This is because country estimates of HIV prevalence have hitherto been derived from pregnant women attending antenatal clinics (ANC) which have their shortcomings as noted below.

Despite the promises of population-based biomarker surveys, they face a number of limitations that include logistical issues, ethical challenges and selection biases (Boerma et al. 2001; Fisher et al. 1996; WHO and UNAIDS 2003). The logistical challenges include proper methods of collecting specimen that ensure the safety of the survey teams and the laboratory personnel, maintaining the cold chain[1] from the field to the laboratory, putting in place effective testing procedures, and sound disposal of specimens and waste once the testing is done (Boerma et al. 2001; Orroth et al. 2003). Ethical issues revolve around obtaining informed consent from the study participants, ensuring confidentiality of both HIV positive and HIV negative respondents,

[1] The cold chain is the system of transporting and storing specimens and reagents used for testing within a safe temperature range (Nayda et al. 2001). The purpose is to prevent them from being less effective or being destroyed. In the context of MDICP, maintaining the cold chain for specimens started right from the field the moment the specimens were collected and was maintained through to the central holding location where all the collected specimens were received before shipping to the laboratory, during the transfer from the central location to the laboratory, and in the laboratory itself.

communicating the results to the study participants, and facilitating the provision of treatment or of resources to those in need of treatment (WHO and UNAIDS 2003). Selection biases arise from the geographical coverage owing to the sampling procedure used, and non-response attributable to refusal or non-contact.

Recent results from the Demographic and Health Surveys (DHS) that included testing for HIV show consistently lower HIV prevalence rates than the rates obtained from ANC data. This trend is shown in Table 1 which compares HIV prevalence from antenatal clinics with those from DHS surveys for selected sub-Saharan African countries. The response rates for HIV testing in these surveys are also given. The problem of selectivity may account for the high HIV prevalence rates observed in the ANC data. Sources of selection bias include the location of clinics that carry out testing for HIV and self-selection into the study by those who attend the clinics. This problem has, however, been recognized in the literature and methods of adjustment have been proposed in extrapolating such data to the general population (see for instance Changalucha et al. 2002; Fabiani et al. 2003; Gregson et al. 2002; Zaba et al. 2000; also Orroth et al. 2003 for sexually transmitted disease prevalence). Non-response could, on the other hand, account for the low HIV prevalence rates observed in the DHS surveys. In deed a number of community-based studies (e.g. Glynn et al. 2001; Fylkesnes et al. 1998; Saphonn et al. 2002) have acknowledged this possibility (see also Boerma et al. 2003; WHO and UNAIDS 2003).

<Table 1 about here>

HIV prevalence estimates from a population-based biomarker study in rural Malawi also show that HIV prevalence is lower than the published ANC figures from the nearest sites. My aim is therefore to examine the effect of non-response on HIV prevalence estimates derived from this study which was part of the Malawi Diffusion and Ideational Change Project (MDICP). First, I explore whether and how non-response due to refusal, absence, death or movement biased the MDICP HIV prevalence estimates. Death and movement are possible sources of non-response in this case because MDICP is a longitudinal study. Biomarkers were included in the third round of the project. Second, I compare the MDICP HIV estimates with those from the adjacent antenatal clinics with

and without taking non-response into account. Finally, I examine if selection due to non-response exert some downward bias in the MDICP HIV prevalence rates.

An important question at this point is why such an exercise is necessary. In the first place, HIV/AIDS, just like any other public health problem, calls for reliable statistics vital for monitoring the progress of the disease, planning control programmes and monitoring the impact of interventions (Fisher et al. 1996; Glynn et al. 2001; Zaba et al. 2000; cf Changalucha et al. 2002). At the very least, this can partly explain the amount of effort that has been invested in developing adjustments for HIV prevalence rates obtained from ANC data. This also leads to the second justification for the exercise carried out in this paper, that is, methodological relevance. In particular, exploring whether non-response is a source of bias in the population-based HIV prevalence estimates could help determine if adjustments are necessary either in the estimates so derived or in future survey designs to ensure good response rates.

The third justification for examining non-response is what I refer to as "the power of statistics". In most of the developing countries, HIV prevalence data from antenatal clinics are produced by the national AIDS councils or commissions. The biomarker surveys are more often than not conducted in collaboration with national statistical offices. In other instances, for example in Tanzania, the national AIDS commission was involved in the DHS survey as well. When the same national agency or two different agencies come up with discrepant HIV prevalence rates, this is likely to generate intense interests from the general populace, the civil society, civil servants and politicians. Anecdotal evidence in the case of Kenya indicates that the government was put under extreme pressure by the civil society and politicians to explain the discrepant HIV prevalence rates.

Background of the Study

Malawi is divided into three administrative regions (Northern, Southern and Central regions). Each region is divided into a number of administrative districts- 27 districts in all according to the 2000 Malawi Demographic and Health Survey (MDHS 2000). Each district is further subdivided into Traditional Authorities (TAs) each headed by a chief. A

group of villages, each under a village headman, form a Traditional Authority (National Statistics Office and ORC Macro 2001). The North and some parts of the South are inhabited by patrilineal and patrilocal while the rest by matrilineal and matrilocal societies (WLSA Research Trust 2000). With a population close to 12 million by mid-2004, about 86% live in the rural areas (Population Reference Bureau 2004).

The national adult HIV/AIDS prevalence rate in 2003 was estimated at 14% (Republic of Malawi 2003). HIV prevalence was higher in the urban and semi-urban areas than in the rural areas (about 22% and 21% versus 15% in the rural areas) and in the South (about 24%) than in the North (20%) or Central (16%) (Republic of Malawi 2003). These rates are based on data from pregnant women attending antenatal clinics. Attendance of antenatal clinics is high both in the rural and urban areas. Estimates from MDHS 2000 show that about 92% of women in rural areas who had a live birth five years before the survey attended antenatal clinic (National Statistics Office and ORC Macro 2001). However, not all antenatal clinics test for HIV/AIDS. The 2003 HIV/AIDS prevalence data come from 19 sentinel sites. Five of these sites were from the North and seven each from Central and Southern regions. Of the five sentinel sites in the North, two were designated as rural, another two as semi-urban, and the remainder as urban. In each of the remaining two regions, three sites were designated as rural, another three as semi-urban, and one as urban (Republic of Malawi 2003)[2].

The patterns of non-response in population-based biomarker surveys can be gleaned from a few of the studies that have so far been conducted in sub-Saharan Africa. More generally, Groves and Couper (1998) have developed a conceptual framework for survey participation. According to the framework, participation in a household survey is viewed as being influenced by the social environment, the household characteristics, the survey design, and the interviewer characteristics. At the empirical level, Huygens et al. (1996) found that the research model, the meanings of research questions, and personal factors affecting the interview relationship influenced the results of a study of sexual behavior in relation to HIV/AIDS in Uganda. Miller et al. (2001), in a study of husband-

[2] The Principal Investigator in the MDICP study took time to visit the Gawanani site in Balaka District and found that it was quite rural as indicated in the 2003 HIV/AIDS report. However, there were suggestions that the sites in Mchinji and Rumphi Districts which are designated as rural might actually be peri-urban. If this is true, it might partly account for the higher rates observed in these sites compared to the MDICP rates.

wife responses in surveys in Malawi, found that respondents tended to tailor their responses based on their perception of what they stood to gain by participating in the study. Boerma et al. (2003) point out that the process of obtaining consent was partly responsible for the high refusal rate in a population-based HIV/AIDS study in South Africa.

Demographic and Health Surveys for five sub-Saharan African countries (Burkina Faso, Ghana, Kenya, Tanzania and Zambia) that have given detailed response rates for HIV tests show no clear pattern of refusal by gender. Whereas a higher proportion of males than females refused HIV tests in Burkina Faso, Ghana and Tanzania, the opposite was the case in Kenya and Zambia (CBS, MOH and ORC Macro 2004; Central Statistical Office, Central Board of Health, and ORC Macro 2003; GSS, NMIMR and ORC Macro 2004; INSD et ORC Macro 2004; TACAIDS, NBS and ORC Macro 2005). More males than females were, however, likely to be absent for HIV test in all the five countries. The possible explanation for this is the gendered division of labor within households in much of sub-Saharan Africa. Men are more likely to be engaged in tasks outside the household while women in tasks within households. Table A1 in the appendix shows the patterns of refusal and absence by urban-rural residence and by gender for the five countries.

Both refusal and absence are higher in the urban than in the rural areas for males and females combined in all the five countries. Absence is likely to be associated with increased risk of HIV infection given that it is those who are absent for household interview that are more likely to be mobile (for example, labor migrants, job-seekers, traders and business people). Mobility has been associated with increased risk of HIV infection in parts of sub-Saharan Africa (e.g. in the study by Crampin et al. 2003). Higher rates of absence in the urban areas compared to rural areas can be attributed to involvement in wage labor or job-seeking. On the other hand, lower socio-economic status has been found to be associated with the tendency of racial and ethnic minorities to cooperate in surveys (Groves and Couper 1998). This might explain higher refusal rates in urban compared to rural areas given that most urban areas in much of the developing world tend to be better off economically than rural areas. Overall, respondents were more likely to refuse than to be absent for HIV test in both urban and rural areas and for males and females.

Similarly, especially in the context of a longitudinal study of the MDICP type, death and movement are other sources of non-response. Death is more likely to be associated with those at the highest risk of HIV infection. In the context of an evolutionary framework, population subgroups endowed with the highest levels of frailty are likely to die out fast leaving behind the robust subgroups (Carey 2003)[3]. The characteristics of the less frail then become the dominant feature of the population as a result. If this were the case then it could be possible that out of the original MDICP sample (from the first round of the survey) only those with the lowest levels of frailty were able to survive to the third round when the testing for HIV was done. Selection would ensure that the MDICP study population in the third round of the survey was made up of those who had the lowest risk of being infected with HIV/AIDS/STIs. This could therefore explain the low prevalence rates. Thus, to the extent that non-responders during testing constituted the highest risk group, HIV prevalence estimates so derived may be biased, hence the need to examine the extent of such bias.

Data

Data for this paper come from the MDICP project, a longitudinal study in rural Malawi that is part of the Social Networks Project of the Population Studies Center, University of Pennsylvania. Its general aim is to examine the role of social networks in changing attitudes and behavior regarding family size, family planning and HIV/ AIDS in Malawi (see http://malawi.pop.upenn.edu for further details). It is conducted in three rural sites in three distinctive districts selected from each of the three regions in the country. These are Rumphi District in the Northern region, Mchinji District in the Central Region, and Balaka District in the Southern region. Though the sampling design was not meant to be representative of the national rural population of Malawi, the sample characteristics have been shown to closely match the characteristics of the rural population of the 1996 Malawi Demographic and Health Survey (Watkins et al. 2003).

[3] Carey (2003) uses the concept of demographic heterogeneity to refer to the endowment of subgroups with different levels of frailty. Demographic selection is then the winnowing process by which populations become more selected as they age because the more frail subgroups exit the population fast. The population is thus transformed into one consisting mostly of the less frail individuals.

The first and second waves of the project (MDICP-1 and MDICP-2) were carried out in 1998 and 2001 respectively and involved survey data collection. The third wave (MDICP-3) was conducted between March and August 2004 and had two components: the survey component and the STI component. Besides following up respondents from previous rounds, MDICP-3 also included new husbands to women interviewed in previous rounds and a sample of adolescents aged 15-24 years. The sample of adolescents included about 500 female respondents in Balaka for the audio-computer assisted self-interviewing (ACASI) study carried out by the Population Council. The present analysis is based on data from Balaka and Rumphi sites only. Out of the possible 3,122 respondents from the two districts (excluding the ACASI sample), 180 cases were discarded when the survey and the STI outcomes were either missing or coded as unknown. This left a total of 2,942 respondents split almost into fifty-fifty between the two districts: 1,461 from Balaka and the other 1,481 from Rumphi. Females make up slightly more than half of the analysis sample (about 53%) while adolescents (both male and female) constitute about 32%. Only 102 new spouses were identified in the two settings.

The STI component involved the collection of biomarkers for testing sexually transmitted infections (STIs) and HIV/AIDS. The aim was to provide population-based information on HIV/AIDS/STI prevalence for the three regions. The information on HIV/AIDS was expected to provide an opportunity for validating the estimates of HIV/AIDS prevalence obtained from data collected from pregnant women attending antenatal clinics. For STI testing, urine samples were collected from male respondents and self-administered vaginal swabs from females. The samples were tested for Chlamydia and gonorrhea for males and Chlamydia, gonorrhea and trichomoniasis for females. For HIV testing, saliva samples were collected from both males and females. A team of trained nurses was responsible for collecting the specimens. They usually visited respondents about two to three days after the visit by the survey team[4].

The process of collecting specimens involved pre-test counseling of respondents, administering of a brief STI questionnaire, obtaining the respondent's informed consent,

[4] An exception was the Rumphi site where, for logistical reasons, about half of the respondents were visited by the nurses' team before being interviewed by the survey team.

and if consent was granted, taking of the specimens. This meant that the respondent could grant an interview for the STI questionnaire but refuse to give specimens for either or both tests. The specimens were then refrigerated over a night or two before being transferred to the laboratory run by the University of North Carolina (UNC) at Chapel Hill based in Lilongwe Central Hospital where the analysis was done[5]. Linked to the process of specimen collection was a randomized experiment focusing on incentives for voluntary counseling and testing (VCT) uptake. Conditional on accepting to give specimens, the respondent was given an opportunity to randomly choose an amount of money written on bottle tops put in plastic bags. The amount picked was then recorded on a voucher which the respondent was to present when he/ she came for his/ her STI or HIV test results. This was followed by post-test counseling of the clients, whether infected or not, so as to avoid any suspicion that only those who were found to be infected were being given post-test counseling.

Methods

The first part involves a sensitivity analysis to determine whether non-response is a source of bias for the observed HIV prevalence rates. The procedure involves assuming a prevalence rate for non-responders that is the same as either the observed MDICP rate or the ANC rate from the adjacent antenatal clinic. This is then defined in the form of relative risk for HIV infection among non-responders compared to the risk of those who were tested.[6] The levels of relative risk range from 1.0 to 3.4. The lower bound represents the case where non-responders have the same HIV prevalence as those who were tested. The upper bound of 3.4 reflects the situation for males in Rumphi District where non-responders have the same prevalence as that estimated from the ANC data from the

[5] The longest the specimens could take before being transferred to the UNC laboratory was about 79 hours. This could be the case for specimens collected on a Friday morning at 9.00 a.m. and delivered to the laboratory on the following Monday at 4.00 p.m.

[6] The relative risk is defined as the ratio of expected percentage of non-responders HIV positive divided by the percent of responders who are HIV positive. This form of defining the assumed HIV prevalence for non-responders was adopted for ease of presentation.

adjacent site[7]. In between, we have relative risks of 1.8 for females and 2.0 for males in the Balaka site, and 2.5 for females in the Rumphi site assuming ANC prevalence rates for non-responders.

The assumed prevalence rate for non-responders is multiplied by the number of non-responders to obtain the expected number of non-responders that would be HIV positive under any given assumption. This number is then added to the number of confirmed positive individuals (among those tested). The sum is divided by the total number of individuals (both non-responders and those tested) to obtain the prevalence that would be observed had non-responders experienced the assumed risk for HIV infection. One-sample test of proportion is performed to determine the probability of observing the new prevalence rate among those who were tested. The expectation is that if non-response is a source of bias, we should expect to see significant differences between the observed and the estimated HIV prevalence rates at any level of assumed prevalence rate among non-responders.

The main sources of non-response for HIV tests were mostly refusal, absence, death, movement, and a final category of 'other'. This final category included outcomes like 'too sick/ hospitalized' and 'divorced/ widowed'. The analysis is done sequentially, first, assuming that refusal was the only source of non-response. The other sources of non-response- absence, death, movement, and 'other'- are included in that order one at a time. Separate analyses are done for each site and for males and females. As noted earlier, one of the reasons for the STI component of MDICP-3 was to obtain HIV/AIDS prevalence rates for comparison with the rates derived from women attending antenatal clinics. Comparisons are therefore made between the estimated HIV prevalence rates that take into account non-response and the ANC rates from the adjacent rural sites. A two-sample test of proportion is used in this comparison[8]. The purpose is to determine if the population-based estimates would be similar to the ANC estimates had the project tested every eligible respondent and found the assumed prevalence rate among non-responders.

[7] Since the published ANC rates are only for females, I obtain the male ANC HIV prevalence rate by assuming a female to male HIV prevalence ratio of 1.2 to 1. This is the ratio that UNAIDS uses in HIV/AIDS projections for generalized epidemics that have been on for more than ten years.

[8] I assume that the variance, and hence the standard deviation, for male ANC HIV prevalence rate is the same as that for females for purposes of testing for the significance of differences between MDICP and ANC male HIV prevalence rates.

The second part of the analysis involves estimating a probit regression model with sample selection (Van de Ven and Van Praag 1981; Winship and Mare 1992) to determine if selection is an important source of bias in the MDICP HIV prevalence rates. The basic selection model is of the following form:

$$Y^*_i = \beta X_i + \varepsilon_{1i} \quad (1)$$

$$Y_i = Y^*_i \text{ if } P^*_i > 0 \quad (2)$$

where equation (1) is the latent equation with Y^*_i being the unobserved HIV status of individual *i*, β is a (k × 1) vector of unknown parameters, X_i is a (k × 1) vector of exogenous variables associated with HIV status, and $\varepsilon_{1i} \sim N(0, 1)$. Y_i in equation (2) is the observed HIV status, and P^*_i is the unobserved propensity to participate in HIV testing. It means that we only observe Y_i for those who participated in HIV testing. P^*_i is commonly estimated by means of a binary regression model of the form:

$$P_i^* = \alpha Z_i + \varepsilon_{2i} \quad (3)$$

where α is a (k × 1) vector of unknown parameters, Z_i is a (k × 1) vector of exogenous variables associated with participation in HIV testing, and $\varepsilon_{2i} \sim N(0, 1)$. The inconsistency of the estimates of β based on the sub-sample that was tested for HIV arises when the correlation coefficient, ρ, between ε_{1i} and ε_{2i} is not equal to zero. This is analogous to the omitted variable bias in which the conditional mean of ε_{1i} given X_i and $P^*_i > 0$ is omitted from the regression. We correct for this potential bias by introducing this conditional mean in the regression equation to obtain:

$$\begin{aligned} Y_i &= \beta X_i + E[\varepsilon_{1i} | X_i, P^*_i > 0] + \varepsilon_{1i} \\ &= \beta X_i + \rho \lambda_i + \varepsilon_{1i} \end{aligned} \quad (4)$$

Since Y_i is dichotomous, we estimate equation (4) by means of a probit selection model[9]. If selection is important, we should expect ρ to be significantly different from zero.

[9] This model is estimated by the 'heckprob" command in STATA 8.0.

At the moment, this part of the analysis is confined to adult respondents only (both males and females) whose background characteristics are obtained from MDICP-2 conducted in 2001. Three probit models are estimated. Model 1 predicts the probability of being tested for HIV. The variables hypothesized to be associated with HIV testing are defined as follows: *age* is grouped into less than 30 years, 30-39 years, 40-49 years, and 50 and over years[10]; dummy variables for *site*, *gender*, whether the respondent *stayed outside the district for more than one month* in the past year preceding the survey, *usually stayed outside the village* (the question referred to husband for females), or had *secondary and above level of education* (proxy for socio-economic status). Age and socio-economic status have been found to be associated with survey participation (for example by Groves and Couper 1998). Other variables include *perceived risk of HIV infection* (no risk/ don't know, low risk, medium/ high risk) and the *number of living children* (proxy for household size, which may determine the likelihood of contactability)[11].

Model 2 predicts the HIV status based on *age*, *site*, *gender*, whether the respondent was *Catholic*, *stayed outside the district for more than one month* in the past year preceding the survey, *usually stayed outside the village*, was in *polygynous union*, had ever engaged in *extra marital affair*, or had ever *used or was using abstinence or condoms*. Catholics are known to oppose the use of condoms and to extent that abstinence proves to be expensive, we should expect a higher probability of being HIV positive among Catholics. Model 3 is the selection model in which Models 1 and 2 are estimated simultaneously. It predicts the probability of being HIV positive conditional on participation in HIV testing and includes only the significant predictors from Models 1 and 2.

[10] Age is adjusted for those who were tested to reflect how old they would be at the time of testing in 2004. For the non-responders, their ages were taken as at the time of interview (2001) which is considered to be the time at which they were last observed.

[11] When the data from the 2004 round of fieldwork become available, I expect to replace the number of living children with the household size and include other factors that may influence participation such as prior testing. The analysis will also include the adolescents and new spouses who are currently excluded.

Results

HIV/STI Prevalence Rates

Table 2 gives the MDICP prevalence rates for HIV and sexually transmitted infections (STIs) in Balaka and Rumphi Districts among those who were tested by selected background characteristics. The patterns reflect what has been observed from prevalence rates derived from antenatal clinic data. Prevalence is higher in the South (8.4% in Balaka) than in the North (4.8% in Rumphi), and among females (7.7%) compared to males (5.4%). Differences by site are statistically significant at the 1% level while gender differences are significant at the 5% level.

<Table 2 about here>

The regional differences are also observed when prevalence among males and females is considered in each site. Male HIV prevalence in Balaka is about twice that of Rumphi (7.1% versus 3.6%). Female HIV prevalence, on the other hand, is about four percentage points higher in Balaka than the observed female prevalence in Rumphi (9.5% compared to 5.6%). These differences are statistically significant at the 5% level. Prevalence of STIs also follows almost a similar pattern except that male STI prevalence is slightly higher in Rumphi than in Balaka (0.7% versus 0.4%). This difference is however not statistically significant while all the remaining differences are significant at the 1% level.

Prevalence rates from the adjacent ANC rural sites are given in the lower panel of Table 2. The MDICP HIV prevalence rates are significantly lower ($p<0.001$) than the ANC rates, a pattern that is observed even if we compare prevalence rates among females alone. For STIs, the MDICP rate for Balaka (7.0%) is significantly higher than the ANC rate from the adjacent rural site of Gawanani (1.0%). The combined STI rate for Balaka and Rumphi is also significantly higher ($p<0.001$) than the combined rate for Gawanani and Mbalachanda (3.7% versus 0.5%). These differences could be due to the different number of sexually transmitted infections tested in both cases: gonorrhea, Chlamydia, and trichomonas in the case of the MDICP sample, and syphilis in the case of ANC.

Response and Non-Response Rates

Response and non-response rates by site and gender are given in Table 3. As far as non-response is concerned, there was generally higher refusal rate for STI test than for HIV test for both sites combined (8.1% versus 7.0%) though this difference is not statistically significant. Refusal rate for either HIV or STI test was higher in Balaka than in Rumphi (8.1% compared to 5.9% for HIV test, $p<0.001$; 8.8% versus 7.5% for STI test, $p=0.0684$). While a higher percentage of males (7.5%) than females (6.5%) refused HIV test in both sites combined, the pattern is reversed for STI test (8.6% of females refused STI test compared to 7.6% of males). But the observed differences in refusal rates by gender are not significant at any of the conventional levels.

<Table 3 about here>

Movement was the major source of non-response for both HIV and STI tests and in both sites. Accounting for almost 15% of the overall response rate for either tests, this rate is higher than that of all the remaining sources of non-response combined. Movement was more of a problem in Rumphi (about 17%) than in Balaka (about 12%). It was also higher among females (about 16%) than among males (about 13%) for both tests and in both sites combined. Whereas the observed difference in non-response due to movement by site is statistically significant at the 1% level, the difference by gender is not. But the pattern of non-response due to movement by gender in each site differs from the pattern observed in both sites combined. In Balaka, a higher percentage of males (about 13%) than females (about 11%) were identified as having moved ($p=0.3430$). The opposite was the case in Rumphi (about 20% females compared to 14% males; $p<0.01$). This holds for both HIV and STI tests.

A note on the ‘other’ category would suffice at this point. As shown in Table 3, the percentage of respondents falling into this category for STI test is twice that of HIV test (5.0% versus 2.6%; $p<0.001$) for both sites combined. During the post-test counseling, final attempts were made to contact those respondents who could not be initially contacted for one reason or the other. Where these final attempts were successful, specimens were collected for HIV testing only. For these respondents, their

STI outcome has been coded as 'other' in the present analysis resulting in the unusually large difference in this outcome for HIV and STI tests.

Effect of Non-Response on HIV Prevalence Rates

The purpose of this section is to answer some 'what if' questions. For instance, what would the MDICP HIV prevalence be if the percentage of non-responders HIV positive was the same as the observed rate among those tested? And what if they had the same prevalence as that observed among the antenatal clinic attendees? Table 4 shows the results of this exercise by site and by gender. Panel A of Table 4 gives the estimated HIV prevalence rates that would be obtained assuming that refusal was the only source of non-response. A relative risk of 2.0 for both males and females in Balaka would give prevalence rates among those who refused the tests that closely resemble the estimated male and the observed female ANC rates for Gawanani (i.e. 14.2% for males and 19.0% for females compared to ANC rates of 14.2% for males and 17.0% for females).

As it turns out, the recalculated prevalence rates based on these assumptions do not result in any significantly different rates from the observed rates among those who were tested for both males and females in this site. In Rumphi, we need to assume a relative risk of 3.4 for male and 2.5 for female refusals to obtain prevalence rates that would resemble the male and female ANC prevalence rates for Mbalachanda. As in the case of the Balaka site, the recalculated prevalence rates based on these assumptions for those who refused the tests do not differ significantly from the observed rates. Neither do we obtain any significantly different results when we assume that those who refused had the same prevalence as those who were tested in the MDICP study for both males and females and in both sites.

<Table 4 about here>

Panel B of Table 4 considers the situation in which refusal and absence were the only sources of non-response. The recalculated HIV prevalence rates under this assumption for various levels of relative risk for HIV infection among non-responders lead to qualitatively similar conclusions as for refusal alone. This is true in both sites and

for both males and females. We start noting some significant differences between the recalculated and the observed MDICP rates when we include death as a source of non-response as shown in Panel C of Table 4. This is however true only in the Balaka site for both males and females and when we consider a relative risk of 3.4. Suffice to keep in mind that this relative risk gives a prevalence rate similar to the ANC rate for male non-responders from the Rumphi site only. The significant differences for both males and females in the Balaka site therefore give prevalence rates for non-responders that are higher than the ANC rates observed in the adjacent site. For male non-responders from this site, it implies a HIV prevalence rate of about 24% while for their female counterparts, it implies a HIV prevalence rate of about 32%. These rates are not plausible for the Balaka site.

Movement is added as an additional source of non-response in the results shown in Panel D of Table 4. For the Rumphi site, we find some significant differences between the recalculated and the observed HIV prevalence rates when non-responders are assumed to have the same prevalence rate as that observed in the adjacent ANC site. This is true for both males and females. The observed rate for males is under-estimated by about 2.4 percentage points and that for females by about 2.7 percentage points. For the Balaka site, significant differences are obtained by assuming prevalence rates among non-responders that are higher than those observed in the adjacent ANC site for both males and females. At least for the Balaka site, the recalculated HIV prevalence rates would not be significantly different from the observed rates even if we assume that non-responders due to refusal, absence, death and movement had the same prevalence as that observed in the adjacent ANC site.

Comparison of ANC and MDICP HIV Prevalence Rates

In this section, I consider all sources of non-response- including the ‘other’ category- and examine how different the recalculated rates are from the ANC rates from the adjacent sites. The results are given in Table 5. Comparisons are also made between the recalculated and the observed MDICP rates when all sources of non-response are considered. Overall, such comparisons yield similar results to those in the last panel of

Table 4. Comparisons with the ANC rates also confirm that we would need to assume that all non-responders (both males and females and in both sites) had a higher risk of HIV infection than that observed in the ANC sites to obtain rates that do not differ significantly from the ANC rates.

<Table 5 about here>

Figure 1 compares prevalence rates among women of reproductive ages (15-49 years) from the MDICP and ANC sites without taking non-response into account. The MDICP rates are based on the sample of adult women who had a birth in the last five years preceding the survey, or who were either childless or had the last birth more than five years before the survey but were pregnant at the time of the survey. They also include some 156 adolescent females aged 15-25years with information on age. The ANC figures apply to aggregated data from eight sentinel surveillance sites designated as rural in the HIV sentinel surveillance report.

<Figure 1 about here>

One striking feature in Figure 1 is the near similarity in the shapes of the two graphs. Ignoring the extreme edges of the MDICP age-pattern of HIV infection, the ANC graph appears to be a version of the MDICP graph slightly tilted to the right. Moreover, HIV prevalence rates from the two sources are almost similar for women aged 25-29 and 45-49 years. At the other ages, however, the ANC rates are consistently higher than the MDICP rates[12]. The peaking of HIV infection in the two sources is also different. While the MDICP graph peaks at ages 25-29 years at about 14%, the ANC graph peaks at ages 35-39 years at the rate of about 22%.

[12] The MDICP rates for ages 15-19 and 20-24 years should be interpreted with caution for the moment. The inclusion of the remaining female adolescents, who fall entirely within these age ranges, may change the shape of the graph at these ages.

HIV Testing, HIV Status and Selection Bias

The correlates of participation in HIV testing and of HIV status as well as the role of selection bias are assessed by means of probit models. The results are given in Table 6. Age, gender, whether the respondent stayed outside the district for more than one month in the past year preceding the survey or usually stayed outside the village are significant predictors of survey participation. Compared to those aged 29 years or younger, older respondents were more likely to participate in HIV testing. The same applies to females compared to males. Those who stayed outside the district for one month or more in the preceding year and those who do not usually live in their villages were less likely to participate. All the coefficients are statistically significant at the 1% level.

<Table 6 about here>

For HIV status, the only significant predictors are age and whether the respondent has been married more than once. Those aged 50 years and over were significantly less likely to be HIV positive compared to those aged 29 years or younger. Significant differences also exist between ages 30-39 and 50 and over (χ^2=9.52; p<0.01), and 40-49 and 50 and over (χ^2=6.59; p<0.05). An overall test for age shows that it is significant at the 1% level (χ^2=11.83; p<0.01). Those who had been married more than once were significantly more likely to be HIV positive (p<0.01). The selection model gives $\rho = 0.30$ with p=0.666, implying that at least for adults, selection is not important.

Discussion and Conclusion

In this paper, I examined the extent of bias in population-based HIV prevalence rates due to non-response using data from the Malawi Diffusion and Ideational Change Project (MDICP) collected from rural Malawi. The data show that HIV prevalence rates are consistently lower than the rates from the antenatal clinics (ANC). This trend is similar to what has been observed in other population-based surveys incorporating biomarkers such as the Demographic and Health Surveys (DHS). And just like in most DHS surveys, the percent of eligible respondents that were tested for HIV is in the seventies or approximates seventy percent. This could be a source of bias especially if the non-

responders tend to be at higher risks of HIV infection than those who participate in testing. It is thus not surprising that as country estimates of HIV prevalence are being adjusted to reflect rates from population-based studies (see for instance UNAIDS 2004), there is also a call for caution that these rates may not be the gold standard owing to the low response rates (e.g. by Boerma et al. 2003; WHO and UNAIDS 2003).

The sources of non-response show that respondents were more likely to refuse than to be absent where absence is defined as being away temporarily. This same pattern has been observed in the DHS surveys. But unlike in most of the DHS surveys where refusal turns out to be the major source of non-response, movement is the main source of response in the present study. The longitudinal nature of the study may explain this pattern given that respondents are more likely to change residences in-between survey rounds. The percent of eligible respondents who had moved was higher in the Rumphi site than in the Balaka site. The seasonality of employment in tobacco plantations in the North might have provided a potential explanation for this were it not for the gender pattern in movement observed in the Rumphi site. One might have argued that people would be more likely to move during low season when tobacco is already harvested and there is not much work. But the fact that this is mainly a male-dominated task while the figures show that females were more likely to move than males implies that the explanation might lie elsewhere.

A simple sensitivity analysis reveals that we would come up with significantly different HIV prevalence rates in the Rumphi site than what we observe in the MDICP data if we assume that non-responders had the same prevalence as that observed from the ANC data from the adjacent site. This holds for both males and females. There are two possibilities regarding this finding. First, it could be an indication of genuine downward bias in the observed MDICP HIV prevalence rate in Rumphi. Second, it could be due to differences in the designation of the MDICP and the ANC sites. While the MDICP site in Rumphi is quite rural, we suspect that the ANC site may be peri-urban which implies a higher HIV prevalence rate in the first place. Assuming this prevalence for non-responders in a rural setting might result in false bias. In the Balaka site where the adjacent ANC site has been confirmed to be rural as well, we do not see any bias resulting from assuming the ANC rate for non-responders.

When we compare MDICP and ANC HIV prevalence rates taking non-response into account, the results show that we need to assume implausible prevalence rates among non-responders to obtain comparable rates. In other words, we need to assume that HIV prevalence among non-responders was higher than the rates observed in the antenatal clinics for the MDICP rates to be similar to the ANC rates. This holds for both males and females and in both sites. What emerges instead is that making the plausible assumption that non-responders had the same prevalence as that observed in the antenatal clinics gives significantly lower estimates than the ANC rates. The implication is that while the MDICP rates may understate the true HIV prevalence, the ANC rates significantly overstate it, at least in these rural settings. In fact there have been suggestions that population-based studies may capture the rural HIV prevalence better than ANC data owing to the location of fewer clinics in rural areas (e.g. Boerma et al 2003).

The factors that turn out to be significantly associated with participation in HIV testing include age, gender, whether the respondent had stayed outside the district for more than one month in the year preceding the survey, or usually lived outside the village. The results are in the expected direction. The last two variables are indicators of movement which should also be associated with increased risk of HIV infection. This is not the case in the present study though previous studies have associated movement with increased probability of being infected with HIV (e.g. Crampin et al. 2003). This could reflect a number of possibilities: it could be that these are poor measures of movement, the association does not exist for this sample, or the results are due to pure chance. Overall, selection due to non-response does not appear to exert bias in the MDICP HIV prevalence rates, at least among adults. A similar finding has been noted for Kenya by Bignami-Van Assche et al. (2005). Perhaps this finding does confirm the suggestion that population-based surveys provide better quality HIV prevalence data for rural populations than do ANC data.

References:

Bignami-Van Assche, Simona, Joshua A. Salomon, and Christopher J.L. Murray. 2005. *Evidence from National Population-Based Surveys on Bias in Antenatal Clinic-Based Estimates of HIV Prevalence*. Paper presented at the 2005 Meeting of the Population Association of America, Philadelphia, March 31-April 2.

Boerma, J. Ties, Peter D. Ghys and Neff Walker. 2003. "Estimates of HIV-1 prevalence from national population-based survey as a new gold standard." *The Lancet* 362:1929-1931.

Boerma, J. Ties, Elizabeth Holt and Robert Black. 2001. "Measurement of Biomarkers in Surveys in Developing Countries: Opportunities and Problems." *Population and Development Review* 27(2): 303-314.

Carey, James P. 2003. *Longevity: The Biology and Demography of Life Span*. Princeton: Princeton University Press.

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.

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

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

Changalucha, John, Heiner Grosskurth, Wambura Mwita, James Todd, David Ross, Philippe Mayaud, Abdul Mahamoud, Arnoud Klokke, Frank Mosha, Richard Hayes and David Mabey. 2002. "Comparison of HIV prevalences in community-based and antenatal clinic surveys in rural Mwanza, Tanzania." *AIDS* 16: 661-665.

Crampin, Amelia C., Judith R. Glynn, Bagrey M.M. Ngwira, Frank D. Mwaungulu, Jörg M. Pönnighaus, David K. Warndorff and Paul E.M. Fine. 2003. "Trends and measurement of HIV prevalence in northern Malawi." *AIDS* 17: 1817-1825.

Fabiani, Massimo, Knut Fylkesnes, Barbara Nattabi, Emingtone O. Ayella and Silvia Declich. 2003. "Evaluating two adjustment methods to extrapolate HIV prevalence from pregnant women to the general female population in sub-Saharan Africa". *AIDS* 17:399-405.

Fisher, Gail, Gregory Pappas and Magdalena Limb. 1996. “Prospects, Problems, and Prerequisites for National Health Examination Surveys in Developing Countries.” *Social Science and Medicine* 42(12): 1639-1650.

Fylkesnes, Knut, Zacchaeus Ndhlovu, Kelvin Kasumba, Rosemary Mubanga Musonda and Moses Sichone. 1998. “Studying the dynamics of the HIV epidemic: population-based data compared with sentinel surveillance in Zambia.” *AIDS* 12: 1227-1234.

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

Giles, Ralph E., Keith R. Perry and John V. Parry. 1999. “Simple/Rapid Test Devices for Anti-HIV Screening: Do They Come UP to the Task?” *Journal of Medical Virology* 59: 104-109.

Glynn, Judith R., Anne Buvé, Michel Caraël, Rosemary M. Musonda, Maina Kahindo, Isaac Macauley, Francis Tembo, Léopold Zekeng and the Study Group on Heterogeneity of HIV Epidemics in African Cities. 2001. “Factors influencing the difference in HIV prevalence between antenatal clinic and general population in sub-Saharan Africa.” *AIDS* 15: 1717-1725.

Gregson, Simon, Nicola Terceira, Memory Kakowa, Peter R. Mason, Roy M. Anderson, Stephen K. Chandiwana and Michel Caraël. 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.

Groves, Robert M. and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons, Inc.

Huygens, Pierre, Ellen Kajura, Janet Seeley and Tom Barton. 1996. “Rethinking Methods for the Study of Sexual Behaviour”. *Social Science and Medicine* 42(2): 221-231.

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.

Miller, Kate, Eliya Msiyaphazi Zulu, and Susan Cotts Watkins. 2001. “Husband-Wife Survey Responses in Malawi.” *Studies in Family Planning* 32(2): 161-174.

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

Nayda, Claire, Ann Kempe, and Nan Miller. 2001. *Keeping it Cool: The Vaccine Cold Chain. Guidelines for Immunisation Providers on Maintaining the Cold Chain*. 2nd Edition. Commonwealth of Australia.

Nyblade, Laura, Jane Menken, Maria J. Wawer, Nelson K. Sewankambo, David Serwadda, Frederick Makumbi, Tom Lutalo and Ron H. Gray. 2001. "Population-Based HIV Testing and Counseling in Rural Uganda: Participation and Risk Characteristics." *Journal of Acquired Immune Deficiency Syndrome* 28(5): 463-470.

Orroth, K.K., E.L. Korenromp, R.G. White, J. Changalucha, S.J. de Vlas, R.H. Gray, P. Hughes, A. Kamali, A. Ojwiya, D. Serwadda, M.J. Wawer, R.J. Hayes and H. Grosskurth. 2003. "Comparison of STD prevalences in the Mwanza, Rakai, and Masaka trial populations: the role of selection bias and diagnostic errors." *Sexually Transmitted Infections* 79: 98-105.

Population Reference Bureau. 2004. *2004 World Population Data Sheet*. Washington, DC: Population Reference Bureau.

Republic of Malawi [National AIDS Commission]. 2004. *Malawi National HIV/AIDS Estimates 2003: Technical Report.*

Republic of Malawi [Ministry of Health and Population]. 2003. *HIV Sentinel Surveillance Report 2003*. National AIDS Commission.

Saphonn, Vonthanak, Leng Bun Hor, Sun Penh Ly, Samrith Chhuon, Tobi Saidel and Roger Detels. 2002. "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." *International Journal of Epidemiology* 31: 449-455.

Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS), and ORC Macro. 2005. *Tanzania HIV/AIDS Indicator Survey 2003-04*. Calverton, Maryland, USA: TACAIDS, NBS, and ORC Macro.

UNAIDS 2004. *2004 Report on the Global HIV/AIDS Epidemic*. Joint United Nations Programme on HIV/AIDS (UNAIDS).

Van de Ven, Wynand P.M.M. and M.S. van Praag. 1981. "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection." *Journal of Econometrics* 17: 229-252.

Watkins, Susan C., Eliya M. Zulu, Hans-Peter Kohler and Jere R. Behrman. 2003. "Introduction to Social Interactions and HIV/AIDS in Rural Africa." *Demographic Research Special Collection 1*. Max-Planck Institute for Demographic Research, Rostock, Germany.

Winship, Christopher and Robert D. Mare. 1992. “Models for Sample Selection.” *Annual Review of Sociology* 18: 327-350.

Women and Law in Southern Africa (WLSA) Research Trust [Malawi]. 2000. *In Search of Justice: Women and Administration of Justice in Malawi*. Blantyre: Dzuka Publishing Co.

World Health Organization (WHO) and UNAIDS. 2003. *Reconciling Antenatal Clinic-Based Surveillance and Population-Based Survey Estimates of HIV Prevalence in sub-Saharan Africa*.

Zaba, Basia W., Lucy M. Carpenter, J. Ties Boerma, Simon Gregson, Jessica Nakiyingi and Mark Urassa. 2000. “Adjusting ante-natal clinic data for improved estimates of HIV prevalence among women in sub-Saharan Africa.” *AIDS* 14: 2741-2750.

Table 1: Comparison of antenatal clinic-based and population-based HIV prevalence with corresponding response rates (for HIV-testing in population-based surveys) for selected sub-Saharan African countries.

Country/ Year	ANC-based HIV prevalence	Population-based HIV prevalence	Response rate for population-based studies[a]
Burkina Faso (2003)	4.2%	1.8%	89.3%
Ghana (2003)	3.1	2.2	84.9
Kenya (2003)	8.0	6.7	76.3
Mali (2001)	1.9	1.7	80.7
South Africa (2001-02)	20.1	15.6	73.7
Tanzania (2003-04)	8.8	7.0	80.5
Zambia (2001-02)	21.5	15.6	76.5

Notes: [a]Response rates appertain to the percent of eligible respondents who were tested for HIV in the population-based surveys; ANC- antenatal clinic.

Sources: Boerma et al. 2003; CBS, MOH, & ORC Macro 2004; Central Statistical Office, Central Board of Health, and ORC Macro 2003; CPS/MS, DNSI et ORC Macro 2002; GSS, NMIMR & ORC Macro 2004; INSD et ORC Macro 2004; TACAIDS, NBS and ORC Macro 2005; WHO & UNAIDS 2003; UNAIDS 2004.

Table 2: Prevalence rates for HIV and sexually transmitted infections (STIs) by selected characteristics in Rural Malawi, Balaka and Rumphi Districts, MDICP 2004 and Antenatal Clinic (ANC) 2003.

	MDICP Prevalence Rates					
	HIV Prevalence (percent)			STI Prevalence[a] (percent)		
Characteristics	Males	Females	Total	Males	Females	Total
Site						
Balaka	7.1 (492)	9.5 (580)	8.4 (1072)	0.4 (468)	12.6 (555)	7.0 (1023)
Rumphi	3.6 (476)	5.8 (534)	4.8 (1010)	0.7 (447)	2.4 (492)	1.6 (939)
Sample						
Adults	7.9 (593)	9.6 (800)	8.9 (1393)	0.4 (560)	7.6 (768)	4.5 (1328)
Adolescents	1.3 (375)	2.9 (314)	2.0 (689)	0.9 (355)	8.6 (279)	4.3 (634)
Total	**5.4 (968)**	**7.7 (1114)**	**6.6 (2082)**	**0.6 (915)**	**7.8 (1047)**	**4.4 (1962)**
	Antenatal Clinic (ANC) Prevalence Rates (percent)					
Gawanani (Balaka)	14.2[b]	17.0 (206)			1.0 (206)	
Mbalachanda (Rumphi)	12.1[b]	14.5 (193)			0.0 (193)	
Both sites combined	**13.2[b]**	**15.8 (399)**			**0.5 (399)**	
All rural sites		**14.5 (1627)**			**2.8 (1627)**	

Notes: [a]STIs tested for in the MDICP sample include gonorrhea, Chlamydia, and trichomonas; for ANC, it is only syphilis; numbers of persons tested are given in parentheses in each case.
[b]Prevalence rates are obtained by assuming a female-to-male HIV prevalence ratio of 1.2 to 1.

Sources: MDICP 2004; Republic of Malawi (2003).

Table 3: Response rates for HIV and STI tests by site and gender in Rural Malawi, MDICP 2004.

	HIV Response Rates (percent)								
	Balaka			Rumphi			Both sites combined		
Outcome	Males	Females	Total	Males	Females	Total	Males	Females	Total
Tested	70.4	76.1	73.4	70.0	66.7	68.2	70.2	71.3	70.8
Refused	8.6	7.6	8.1	6.3	5.5	5.9	7.5	6.5	7.0
Absent	2.4	1.2	1.8	4.1	2.1	3.0	3.3	1.7	2.4
Moved	12.6	10.6	11.6	14.0	20.3	17.4	13.3	15.6	14.5
Dead	2.9	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
Other	3.1	1.7	2.4	2.8	2.6	2.7	3.0	2.2	2.6
Cases (N)	699	762	1461	680	801	1481	1379	1563	2942
STI Response Rates (percent)									
Tested	68.1	73.7	71.0	65.6	61.7	63.5	66.9	67.6	67.2
Refused	8.1	9.3	8.8	7.1	7.9	7.5	7.6	8.6	8.1
Absent	2.4	1.2	1.8	4.0	2.1	3.0	3.2	1.7	2.4
Moved	12.6	10.6	11.6	14.0	20.3	17.4	13.3	15.6	14.5
Dead	2.9	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
Other	5.9	2.4	4.0	6.6	5.2	5.9	6.2	3.8	5.0
Cases (N)	699	762	1461	680	801	1481	1379	1563	2942

Note: Percentages may not add up to exactly 100 in some cases due to round-off error.

Table 4: Comparison of observed and estimated MDICP HIV prevalence rates under different assumptions of relative risks for HIV infection among non-responders, MDICP 2004.

Assumed relative risk for non responders[a]	Estimated HIV prevalence rates assuming refusal is the only source of non-response (Panel A)			
	Males		Females	
	Balaka	Rumphi	Balaka	Rumphi
1.0	7.1	3.6	9.5	5.8
1.8	7.7	3.8	10.1	6.1
2.0	7.8	3.8	10.2	6.2
2.5	8.2	4.0	10.6	6.4
3.4	8.8	4.2	11.2	6.7
	Estimated HIV prevalence rates assuming refusal and absence as the only sources of non-response (Panel B)			
1.0	7.1	3.6	9.5	5.8
1.8	7.8	3.9	10.2	6.2
2.0	8.0	4.0	10.3	6.3
2.5	8.5	4.2	10.8	6.6
3.4	9.3	4.6	11.5	7.1
	Estimated HIV prevalence rates assuming refusal, absence and death as the only sources of non-response (Panel C)			
1.0	7.1	3.6	9.5	5.8
1.8	8.0	4.0	10.4	6.4
2.0	8.2	4.1	10.6	6.5
2.5	8.8	4.4	11.2	6.9
3.4	9.8*	4.9	12.2*	7.5
	Estimated HIV prevalence rates assuming refusal, absence, death and movement as sources of non-response (Panel D)			
1.0	7.1	3.6	9.5	5.8
1.8	8.6	4.4	11.1	7.2
2.0	9.0	4.6	11.5	7.6
2.5	10.0*	5.1	12.6*	8.5*
3.4	11.7**	6.0*	14.4**	10.1**

Notes: [a]Relative risk is defined as the ratio of the expected percentage of non-responders HIV positive to the observed percent HIV positive among those tested.
**p<0.01; *p<0.05.

Table 5: Comparison of MDICP and antenatal clinic (ANC) HIV prevalence rates under different assumptions of relative risks for HIV infection among all non-responders, Balaka and Rumphi Districts, ANC 2003 and MDICP 2004.

	Estimated HIV prevalence rates assuming all sources of non-response (Males[b])					
		Balaka			Rumphi	
Relative risk for non responders[a]	Percent HIV positive	Sign and significance of the difference from MDICP	ANC	Percent HIV positive	Sign and significance of the difference from MDICP	ANC
1.0	7.1	(+) NS	(-) **	3.6	(+) NS	(-)**
1.8	8.8	(+) NS	(-) *	4.4	(+) NS	(-) **
2.0	9.2	(+) NS	(-) *	4.6	(+) NS	(-) **
2.5	10.2	(+) *	(-) NS	5.2	(+) NS	(-) **
3.4	12.1	(+) **	(-) NS	6.1	(+) *	(-) **
	Estimated HIV prevalence rates assuming all sources of non-response (Females)					
1.0	9.5	(+) NS	(-) **	5.8	(+) NS	(-) **
1.8	11.2	(+) NS	(-) *	7.3	(+) NS	(-) **
2.0	11.7	(+) NS	(-) *	7.7	(+) NS	(-) **
2.5	12.7	(+) *	(-) NS	8.6	(+) *	(-) *
3.4	14.7	(+) **	(-) NS	10.3	(+) **	(-) NS

Notes: [a]Relative risk is defined as the ratio of the expected percentage of non-responders HIV positive to the observed percent HIV positive among those tested.

[b]Male ANC rate is computed by assuming a female to male HIV prevalence ratio of 1.2 to 1. The tests assume that the standard error for the proportion of males HIV positive is the same as that of females for the ANC data.

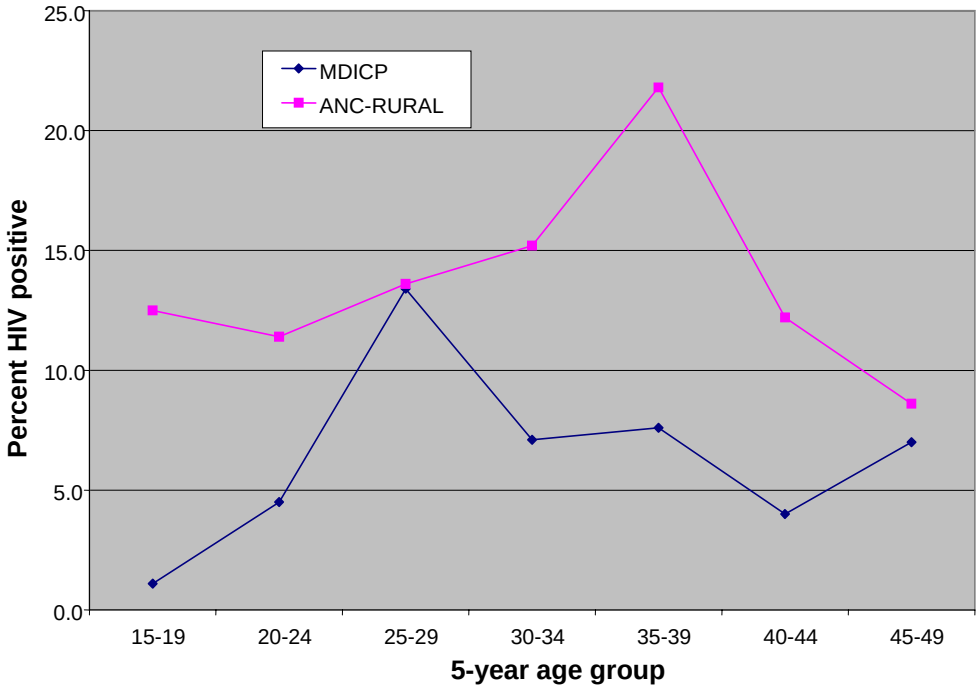
**p<0.01; *p<0.05; NS: not significant; (+): positive difference; (-): negative difference.

Table 6: Results of the probit models predicting participation in HIV testing, HIV status, and HIV status conditional on HIV testing for adult males and females, MDICP 2004.

Covariates	Model 1	Model 2	Model 3
Age group (years)			HIV Status
15-29 [R]			
30-39	0.662** (0.119)	-0.144 (0.178)	-0.079 (0.245)
40-49	0.859** (0.140)	-0.246 (0.187)	-0.046 (0.269)
50+	0.999** (0.170)	-0.759** (0.237)	-0.728* (0.349)
Site (Balaka = 1)	0.016 (0.095)	0.212 (0.128)	
Married more than once		0.557** (0.131)	0.583** (0.125)
			HIV Test
Sex (Females = 1)	0.278** (0.097)	0.074 (0.149)	0.365** (0.082)
Stayed outside district (>1 month)	-0.350** (0.111)	-0.001 (0.182)	-0.282** (0.099)
Husband/man usually stays elsewhere	-0.421** (0.142)	0.324 (0.209)	-0.417** (0.129)
Polygynous union		0.065 (0.140)	
Ever had extramarital sex		0.087 (0.179)	
Use of condoms/ abstinence		0.069 (0.185)	
Religion (Catholic=1)		-0.007 (0.183)	
Perceived risk of infection			
No risk/ don't know [R]			
Low risk	-0.044 (0.112)		
Medium/ high risk	-0.023 (0.114)		
Number of children alive	-0.029 (0.021)		
Highest education level (sec+=1)	-0.199 (0.121)		
Age group (years)			
15-29 [R]			
30-39			0.610** (0.100)
40-49			0.756** (0.105)
50+			0.925** (0.122)
ρ (rho)			0.30 (0.748)
LR test of independence of equations: χ^2 (1 d.f.) = 0.19; P(χ^2 > 0) = 0.666			

Notes: [R]Reference category; d.f. = degrees of freedom; sec+ = secondary and above; $\rho = \text{corr}(\varepsilon_{1i}, \varepsilon_{2i})$; standard errors are in parentheses.
*p<0.05; **p<0.01.

Figure 1: HIV prevalence among women aged 15-49 years, MDICP 2004 and ANC 2003



Notes: The MDICP figures are based on adult women aged 15-49 years who had a birth in the last five years preceding the survey, or were either childless or had a birth more than five years before the survey but were pregnant at the time of the survey. They also include 156 female adolescents aged 15-25 years with information on age. The ANC figures are based on the aggregated ANC rural sites- two in the northern region and three each in the central and southern regions.

Sources: Republic of Malawi 2003; MDICP 2004.

APPENDIX

Table A1: Percent refusing and absent for HIV test in the Demographic and Health Surveys for selected sub-Saharan African countries.

	Urban			Rural			Total		
Country	Males	Females	Total	Males	Females	Total	Males	Females	Total
				Percent refusing HIV test					
Burkina Faso	15.3	11.7	13.5	3.3	2.0	2.6	6.6	4.4	5.4
Ghana	15.1	6.8	10.6	7.9	4.9	6.3	10.7	5.7	8.1
Kenya	16.5	19.2	17.8	11.2	11.9	11.5	13.0	14.4	13.7
Tanzania	21.6	18.4	19.9	11.0	9.9	10.4	13.9	12.3	13.0
Zambia	15.6	15.1	15.3	14.5	15.6	15.1	14.8	15.4	15.1
				Percent absent for HIV test					
Burkina	7.7	2.1	4.9	3.7	1.8	2.6	4.8	1.9	3.2
Ghana	8.8	3.8	6.1	6.2	3.0	4.6	7.2	3.3	5.2
Kenya	20.3	10.6	15.4	7.9	3.5	5.7	12.2	6.0	9.1
Tanzania	13.3	4.5	8.5	6.9	3.9	5.3	8.7	4.1	6.2
Zambia	13.3	3.2	8.1	5.4	3.0	4.2	8.1	3.0	5.5

Sources: CBS, MOH, & ORC Macro 2004; Central Statistical Office, Central Board of Health, and ORC Macro 2003; GSS, NMIMR & ORC Macro 2004; INSD et ORC Macro 2004; TACAIDS, NBS and ORC Macro 2005