A simple analytical approach for estimating prevalence from mortality by HIV sero-status in rural South Africa

Makandwe Nyirenda[1]

Till Bärnighausen [1,2]

Marie-Louise Newell[1,3]

1. Africa Centre for Health and Population Studies, University of KwaZulu Natal, South Africa,
2. Department of Population and International Health, Harvard School of Public Health, Boston, USA,
3. Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, UCL, UK.

Correspondence to: M Nyirenda, Africa Centre for Health and Population Studies, PO Box 198, Mtubatuba 3935, South Africa. E-mail: mnyirenda@africacentre.ac.za,
Telephone: +27-35-550-7526, Fax: +27-35-550-7565.

The Africa Centre Demographic Information System (ACDIS) is supported by the Wellcome Trust through Grant GR082384/Z/07/Z.

Introduction

About 40 million people were estimated to be HIV infected worldwide in 2006 (UNAIDS, 2006); subsequent advancement in estimation methodology and improved survey data necessitated revision, resulting in an estimated 33.2 million HIV infected people (range 30.6 - 36.1 million) alive globally in 2007 (WHO, 2008). Southern Africa, and South Africa in particular, continues to bear the brunt of the HIV epidemic, with the HIV prevalence rate among adults aged 15-49 years estimated to be 16.9% (Shisana, Rehle, Simbayi, Zuma, Jooste, Pillay-van-Wyk et al., 2009).

Despite the relative ease of diagnosing HIV, the exact number of HIV infected people is unknown even in developed countries because testing uptake in the general population is not universal (WHO, 2007). The numbers in the early years of the pandemic were estimated largely from testing pregnant women attending antenatal clinics from select sentinel surveillance sites (UNAIDS/WHO, 2003; Walker, Stanecki, Brown, Stover, Lazzari, Garcia-Calleja et al., 2003). In addition, a host of other more specific sources have been utilised such as regional or national household surveys (Marston, Harriss & Slaymaker, 2008), surveys among high-risk populations (Islam & Conigrave, 2008; Nardone, Mercey, Johnson & McCarthy, 1999), and population-based surveillance studies (Garcia-Calleja, Gouws & Ghys, 2006). A common thread in many of these survey data sources is the incomplete coverage, non-response (Garcia-Calleja et al., 2006) or other selection bias. Survey participation is by design largely voluntary. A major concern for analysis and generalisation is how big the non-response rate is and whether the bias introduced is substantial (Garcia-Calleja et al., 2006). This is of particular concern if differential response rates are associated with specific characteristics of the population or high-risk groups (Boerma, Ghys & Walker, 2003) and if these data are used as inputs in deriving demographic, social and economic impacts of HIV.

An earlier study examining mortality patterns and levels by HIV infection status in rural South Africa found mortality among HIV-infected individuals to be upto 20 times higher than among HIV uninfected individuals of similar ages (Nyirenda, Hosegood, Bärnighausen & Newell, 2007). Further, mortality among people not participating in the HIV surveillance was shown to be about 7 times higher than that of people who tested HIV-negative. That study utilised data from a longitudinal population-based HIV surveillance conducted between 2003 and 2006, with mortality rates derived from recorded deaths and aggregated person-years of exposure for calendar years 2004, 2005 and 2006 from a parallel household survey.

We now explore a simple approach for estimating the likely HIV status composition of the unknown HIV status group. By using the mortality levels and patterns of the HIV-positive, HIV-negative and the unknown HIV status group, it is possible to estimate the HIV prevalence among the latter group; in turn an improved estimate of the overall prevalence in this population can be obtained. The resultant approach could be applied elsewhere in similar studies suffering from problems of non-response, and provides a means to assess the representativeness of the non-testing group with regard to the general population. The approach in addition would appear to hold even in the context of HIV treatment (ART).

Methods

Data sources

Data used in this analysis, and the initial work that motivated it, come from the Africa Centre Demographic Information System (ACDIS), a longitudinal demographic surveillance system located in a largely rural Umkhanyakude district in northern KwaZulu- Natal, South Africa (www.africacentre.ac.za). Demographic, social and economic data have been collected since 2000 from a key informant reporting

on all individual household members whether resident or non-resident in the geographically well-defined surveillance area.

In addition to the routine collection of demographic data (household surveillance), a parallel prospective population-based individual surveillance has been conducted in which information on health and sexual behaviours and a sample for HIV sero-status testing has been collected since 2003 (Bärnighausen, Hosegood, Timaeus & Newell, 2007; Welz, Hosegood, Jaffar, Batzing-Feigenbaum, Herbst & Newell, 2007) from women 15-49 and men 15-54 years (from January 2007 the upper age limits were lifted, surveillance is now among all adults 15+). A random sample of 12.5% of individuals not resident within the surveillance area is in addition included. During household visits with eligible individuals, written informed consent is obtained and a dried blood spot is prepared from a finger prick. A broad based HIV-1/HIV-2 ELISA test (Vironostika, Organon Teknika, Boxtel, The Netherlands) was used to determine HIV status at the Centre's virology laboratory in Durban. All positive test results were confirmed by a second ELISA (GAC-ELISA, Abbott, Abbott Park, Illinois, USA) on the same sample. HIV infection was defined by positive antibody status on both ELISAs, HIV negative status was defined by a negative first ELISA, HIV unknown defined as not providing a sample.

Ethical approval for both the demographic and HIV surveillances was obtained from the Research Ethics Committee of the University of KwaZulu Natal. Details about ACDIS can be found elsewhere (Hosegood, Benzler & Solarsh, 2005; Hosegood & Timæus, 2005; Tanser, Hosegood, Bärnighausen, Herbst, Nyirenda, Muhwava et al., 2007).

Study sample

This analysis is based on data from all individuals who were age-eligible (women 15-49 and men 15-54 years) for inclusion in the 2005 survey round, and resident in the area at the time of drawing the sample. There were 21,492 resident individuals surveyed in 2005, with median age of the sample of 25 years (range 15-54 years). Person-years of exposure were estimated from 1st January 2005, or a date thereafter at which an individual was first tested or visited in the 2005 HIV survey round and right-censored on 31st December 2006 or by death, out-migration, household membership end or refusal to participate in demographic surveillance before end of 2006. Three distinct strata were identified in the analyses, those who were: 1) HIV-negative, 2) HIV-positive and 3) HIV status unknown (did not get tested in the HIV surveillance). Mortality rates were calculated by dividing the number of deaths by the person-years of exposure by strata.

Summary of Nyirenda et al earlier findings

Nyirenda et al. (2007) reported that 21% (n=1773) of those participating in the 2004 survey round were HIV infected. In the 2005 and 2006 survey rounds, 20% of the sampled population tested HIV positive in each round. The overall mortality rate for the three-year period 2004-2006 was 18.1 deaths per 1000 person-years lived (1252/69,330). Mortality was highest in 2005, 22.5 deaths per 1000 person-years (530/23,528) compared to 16.0 (348/21,783) in 2004 and 15.6 (374/24,019) in 2006. Mortality in general and by age was several times higher among HIV infected persons compared to the uninfected persons (Table 1). The adjusted hazards of dying among persons aged 35 years and over were three to seven times higher than those in the 15-24 year age category. Those aged 25-34 years were about twice as likely to die as those in the 15-24 age group. Further adjusting for age, place of residency and socio-economic status, men had significantly higher adjusted hazard ratios of dying than women, (aHR 2.2, 95% CI 1.3-3.8

for HIV uninfected and aHR 1.7, 95% CI 1.1-2.3 for HIV infected). For details of the methods and results see (Nyirenda et al., 2007).

Presentation of a simple analytical model

We now want to explore whether it is possible to use the information on the mortality levels and patterns by HIV status (positive, negative and unknown) to infer the HIV prevalence among those not participating in the HIV surveillance.

Were individuals with unknown HIV status to be tested some would be found to be HIV infected and others uninfected. Thus, the total person-years lived in this group is contributed partly by HIV infected persons and partly by uninfected people.

Using this understanding, the simple analytical model can be presented as follows. Let us start by allowing for mortality in the HIV negative and in the HIV positive as well as in the unknown HIV status population to vary by sex, s, age group, a, and time period, t. Denoting mortality rates by M, and using the subscripts pos, neg and unk to denote the rates in infected, uninfected and persons for whom HIV status is unknown respectively, we have:

$$M_{unk}(s,a,t) = h_{unk}(s,a,t) \cdot M_{pos}(s,a,t) + \left[1 - h_{unk}(s,a,t)\right] \cdot M_{neg}(s,a,t)$$

where $h_{unk}(s,a,t)$ is the proportion of the total person-years lived in the HIV unknown population (of sex s, age a, in time period t) contributed by infected persons. Another name for h is the period HIV prevalence in the HIV unknown population. Of course $h_{unk}(s,a,t)$ is unknown, but all the mortality rates are known because M_{pos} and M_{neg} have been measured in the population that did have HIV tests. So we can solve for $h_{unk}(s,a,t)$,:

$$h_{unk}(s,a,t) = \frac{M_{unk}(s,a,t) - M_{neg}(s,a,t)}{M_{pos}(s,a,t) - M_{neg}(s,a,t)}$$

We can also compare the period HIV prevalence in the untested population $h_{unk}(s,a,t)$, with period HIV prevalence in the tested population, $h_{knw}(s,a,t)$, because the latter is simply the person-years lived by those who tested positive as a fraction of the person-years lived by positive and negative.

$$h_{knw}(s,a,t) = \frac{PY_{pos}(s,a,t)}{PY_{pos}(s,a,t) + PY_{neg}(s,a,t)}$$

The best estimate of the HIV prevalence in this population can be obtained as a sum-product of the estimated period HIV prevalence, the observed HIV prevalence among those who tested and the person-years lived among the tested and the untested population.

$$h_{best}(s,a,t) = \frac{(h_{unk}(s,a,t) * PY_{unk}(s,a,t) + h_{knw}(s,a,t) * (PY_{pos}(s,a,t) + PY_{neg}(s,a,t)))}{PY_{pos}(s,a,t) + PY_{neg}(s,a,t) + PY_{unk}(s,a,t)}$$

Results

In this analysis we present findings for the year, t=2005 for illustration purposes of the analytical model. Table 1 presents the recorded deaths, person-years of exposure contributed over the two-year period 2005-2006 and mortality rates of age-eligible resident individuals surveyed in the 2005 surveillance round, to which the simple analytical model was applied. The analytical model produced equally consistent results for the years 2004 and 2006.

Table 1: Number of deaths, person-years and mortality rates by age, sex and year, 2005-2006

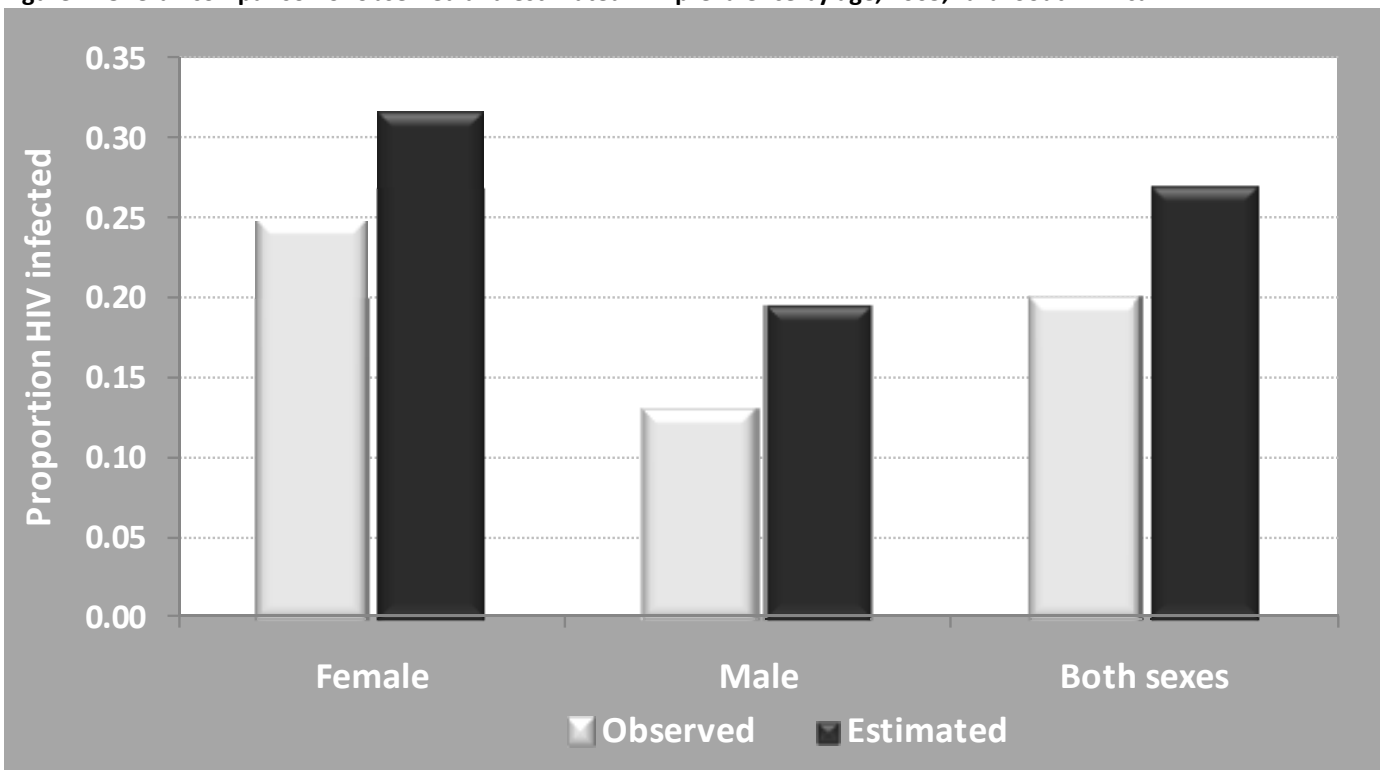
	HIV-Negative		HIV-Positive		Not tested		HIV-Negative		HIV-Positive		Not tested	
	Female						Male					
Age 2005	D/PY	MR	D/PY	MR	D/PY	MR	D/PY	MR	D/PY	MR	D/PY	MR
15-24	13/5325	2.4	19/888	21.4	37/6260	5.9	12/5333	2.3	3/166	18.1	30/6719	4.5
25-34	4/1211	3.3	46/1078	42.7	103/4229	24.4	5/779	6.4	17/416	40.8	71/2793	25.4
35-44	5/1710	2.9	33/774	42.6	70/3745	18.7	7/582	12.0	28/299	93.6	74/2162	34.2
45-54	7/1250	5.6	14/330	42.4	41/1967	20.8	13/642	20.2	18/176	102.0	74/1678	44.1
Total	29/9497	3.1	112/3071	36.5	251/16201	15.5	37/7337	5.0	66/1058	62.4	249/13352	18.6

D=Deaths; PY=Person-years of follow-up; M=mortality rate per 1000 person-years of follow-up

Prevalence estimates for both sexes, 2005

Applying the model to data in Table 1, for both sexes combined the overall estimated HIV prevalence was 27.0%, whereas the observed among resident adults 15-49 years who participated in the 2005 survey was 19.7%. Figure 1 shows that given the mortality rates observed by HIV sero-status, and in those of unknown HIV status in particular, the overall HIV prevalence rate from the observed data likely underestimates the true population HIV prevalence rate. Though the observed 2005 prevalence is lower than the estimated, the age patterns were very consistent.

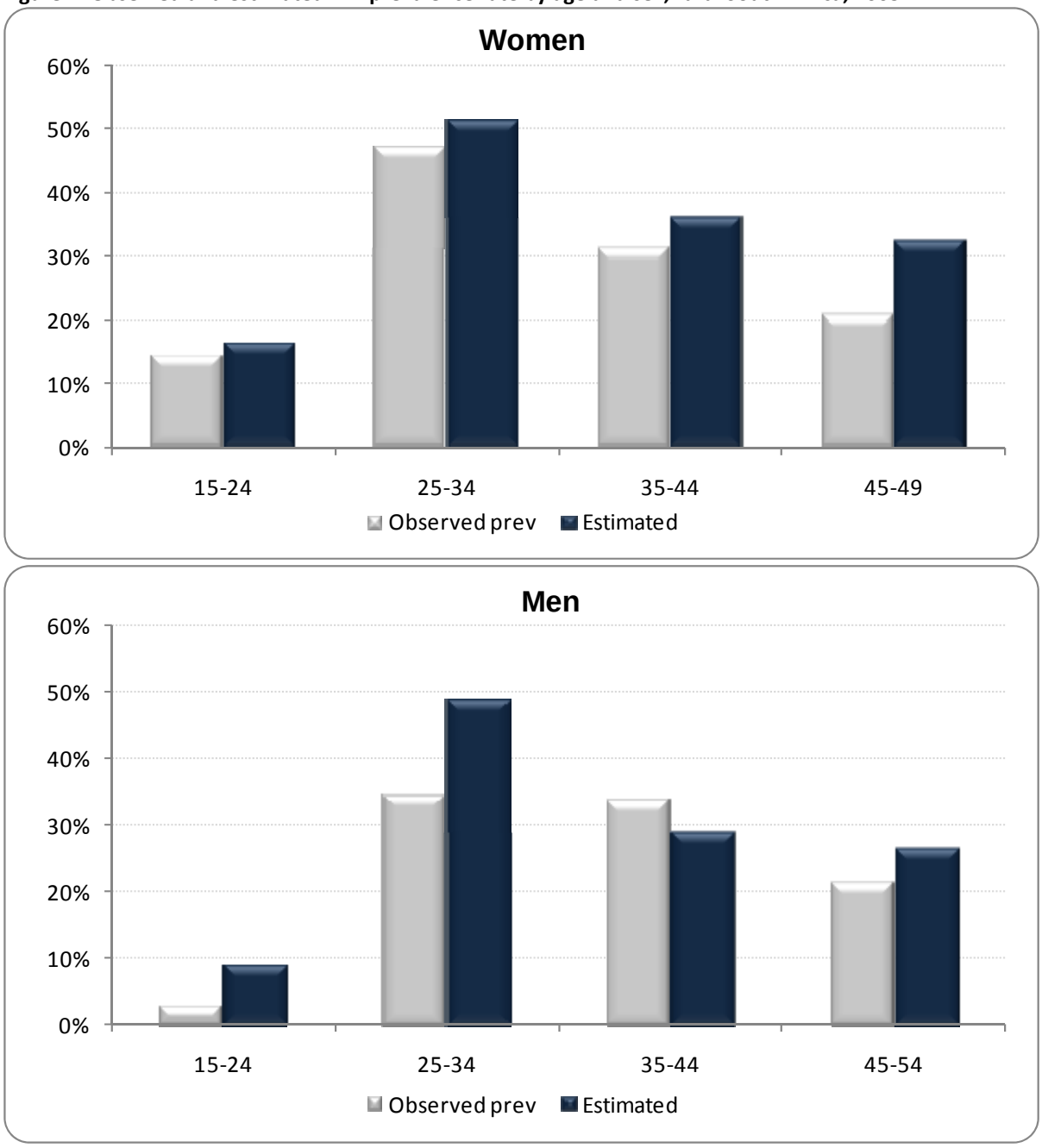
Figure 1: Overall comparison of observed and estimated HIV prevalence by age, 2005, rural South Africa



Prevalence estimates by age and sex, 2005

The best estimate of HIV prevalence from the analytical model for females aged 15-49 year in 2005 was 31.6%. The observed prevalence among participating resident adult females in 2005 was 24.4%. Males aged 15-54 years showed differences between the observed and the estimated prevalence 12.6% vs 19.4%. Males were observed, as well as estimated, to however have a lower HIV prevalence rate than females. Though the levels may differ, HIV prevalence patterns by age were consistent between males and females; prevalence was highest in the 25-34 age range (Figure 2). The only other difference was that among males the estimated prevalence was lower than the observed in age group 35-44 unlike at other ages and among females.

Figure 2: Observed and estimated HIV prevalence rate by age and sex, rural South Africa, 2005



Discussion

The purpose of this analysis was to explore a simple mathematical approach that could be used to obtain a best fit estimate of the level of HIV in a population affected by non-response to HIV surveillance. The approach presented here makes use of computations that are readily done in many population surveys. While information such as births and deaths are easily obtained in surveys, the same cannot be said of HIV testing. Despite improved and easy technologies for HIV testing (WHO, 2007), testing uptake particularly in surveillance studies remains a challenge. In the simple approach we present, data on mortality among persons who are HIV positive, HIV negative and among those not consenting to participate in the surveillance were used to provide an estimate of the HIV profile of those who do not test. This estimate was then used to adjust the overall observed prevalence and derive a best estimate of the population HIV prevalence. Our findings suggest this approach can produce fairly sound results. We stratified our analysis by age, sex and year. Other co-variants that were considered and could be used in this simple approach are education, place of residency, employment status and socio-economic status. These co-variants are associated with HIV incidence (Bärnighausen et al., 2007) as well as with mortality levels (Nyirenda et al., 2007). Analyses by these variants run into the problem of small numbers hence lacked the necessary statistical power to significantly improve the model and were thus not further utilised.

In a more detailed analysis of the HIV prevalence rate of the same study population, it was reported that the HIV prevalence rate among persons participating in the 2003/4 HIV survey round was 13.5% among males 15-54 years and 26.8% among women 15-49 years (Welz et al., 2007). From the analytical model presented here, for 2005 we estimate the HIV prevalence rate among women 15-49 years was 31.6% and 19.4% for men aged 15-54. According to the 2006/7 Swaziland demographic and health survey, a neighbouring country very close to the surveillance area and with comparable HIV rates, 31% and 20% of women and men aged 15-49 were HIV infected (SDHS, 2008; Shisana et al., 2009). Our estimated prevalence rates are further plausible when compared to another DHS from Lesotho, another neighbouring country close to the study area and similarly facing a high HIV burden. The 2004 Lesotho demographic and health survey found that 26.4% of women 15-49 and 19.3% of men 15-54 were HIV infected (LDHS, 2008). The overall adult HIV prevalence rate from Lesotho was 23.2% and from Swaziland 26%. Our estimated overall prevalence rate was 27.6% for 2005 for adults 15-49 years. The HIV prevalence rate for adults 15-49 years in KwaZulu-Natal from nationally representative cross-sectional population-based household surveys that used a multi-stage cluster sampling methodology conducted in 2002, 2005 and 2008 was estimated to be 15.7%, 21.9% and 25.8% respectively (Shisana et al., 2009). Although the cited sources are not directly comparable to our estimates due to methodological differences, it is interesting to note the comparability of our findings from this simple approach to these other sources.

Provided demographic data of age, death and exposure are available, and thus mortality rates can be computed for persons consenting as well as those not consenting to HIV testing, this approach can hence be easily applied. The other advantage of this method is that it does not rely on complex computer simulations or extensive data sets. This is a simple approach for addressing non-response bias where mortality data are available. Other more complex but robust ways to address the same problem exist such as multiple imputations (Bärnighausen, Tanser, Gqwede, Mbizana, Herbst & Newell, 2008). Regression equations may also be used with independent variables such as age, sex, education, place of residency and economic status to predict the likelihood of being HIV infected among the population not tested in the HIV surveillance, and thus obtaining an estimate of the population HIV prevalence.

This simple approach is dependent on the population tested not being statistically significant different from those not consenting. Non-response or non-consent is a common feature of population-based HIV surveys (Boerma et al., 2003; Garcia-Calleja et al., 2006; Mishra, Barrere, Hong & Khan, 2008). A limitation of the approach is thus that the magnitude of non-response, particularly if it is among a select group or persons with particular characteristics is likely to bias survey findings. For HIV surveillance in particular it might be predominantly high-risk groups not consenting (Bärnighausen et al., 2008). As such using the prevalence levels among the low risk groups who participate to estimate prevalence in this high-risk group and in turn the overall prevalence using the approach presented here may potentially under estimate the population prevalence. For our study area there is no indication that persons who do not consent to testing are of peculiar characteristics, to have a significant bias on the findings presented.

This simple approach is easy to use and produces seemingly reliable results. It produces very plausible and consistent results to findings from the same area and comparable populations. Population-based surveys, which inevitably face non-response problems (Boerma et al., 2003), should consider this simple approach in adjusting their HIV prevalence estimates as well as the demographic, socio and economic impacts of HIV on present and future populations.

In sum, after adjusting for the non-consenting population, there was very little change in the pattern of age specific HIV prevalence for women or men and for both sexes combined, but a significant difference between the overall estimated and observed prevalence. The simple model presented here demonstrates a probable degree of underestimation of HIV prevalence in the study population. It must be pointed out though that some of the upward adjustment in prevalence we show may be confounded by the expected increase in prevalence due to ART impact (Herbst, Cooke, Bärnighausen, KanyKany, Tanser & Newell, 2009); as mortality declines due to HIV treatment, other factors being equal, HIV prevalence will increase. While this simple approach produces reliable measures of the population HIV prevalence, there is a need for more robust tools for predicting prevalence levels among the untested population in order to adjust the observed prevalence towards a true population-level prevalence.

References

Bärnighausen, T., V. Hosegood, I.M. Timaeus, & M.L. Newell. 2007. "The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa." *AIDS* 21 Suppl 7:S29-38.

Bärnighausen, T., F. Tanser, Z. Gqwede, C. Mbizana, K. Herbst, & M.-L. Newell. 2008. "High HIV incidence in a community with high prevalence in rural South Africa: findings from a prospective population-based study." *AIDS* 22(1):139-144.

Boerma, J.T., P.D. Ghys, & N. Walker. 2003. "Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard." *362* The Lancet(9399):1929 - 1931.

Garcia-Calleja, J.M., E. Gouws, & P.D. Ghys. 2006. "National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates." *Sex Transm Infect* 82 Suppl 3:iii64-70.

Herbst, A.J., G.S. Cooke, T. Bärnighausen, A. KanyKany, F. Tanser, & M.-L. Newell. 2009. "Adult mortality and antiretroviral treatment roll-out in rural KwaZulu-Natal, South Africa." *Bulletin of the World Health Organisation* 87:1-21.

Hosegood, V., J. Benzler, & G. Solarsh. 2005. "Population mobility and household dynamics in rural South Africa: implications for demographic and health research." *Southern African Journal of Demography* 10(1&2):43-67.

Hosegood, V. & I.M. Timæus. 2005. "Household Composition and Dynamics in KwaZulu Natal, South Africa: Mirroring Social Reality in Longitudinal Data Collection." Pp. 58-77 in *African Households. Census data*, edited by E. van der Walle. New York: M.E. Sharpe Inc.

Islam, M.M. & K.M. Conigrave. 2008. "HIV and sexual risk behaviors among recognized high-risk groups in Bangladesh: need for a comprehensive prevention program." *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases [Abstract]* 12(4):363-370.

LDHS. 2008. "Lesotho Demographic and Health Survey 2004." Calverton, Maryland: Ministry of Health and Social Welfare (MOHSW), Bureau of Statistics (BOS) [Lesotho], and ORC Macro.

Marston, M., K. Harriss, & E. Slaymaker. 2008. "Non-response bias in estimates of HIV prevalence due to the mobility of absentees in national population-based surveys: a study of nine national surveys." *Sex Transm Infect* 84(Suppl_1):i71-77.

Mishra, V., B. Barrere, R. Hong, & S. Khan. 2008. "Evaluation of bias in HIV seroprevalence estimates from national household surveys." *Sex Transm Infect* 84(Suppl_1):i63-70.

Nardone, A., D. Mercey, A.M. Johnson, & M. McCarthy. 1999. "Developing surveillance for HIV transmission and risk behaviours among high-risk groups in a central London health district." *J Public Health* 21(2):208-214.

Nyirenda, M., V. Hosegood, T. Bärnighausen, & M.L. Newell. 2007. "Mortality levels and trends by HIV serostatus in rural South Africa." *AIDS* 21 Suppl 6:S73-79.

SDHS. 2008. "Swaziland Demographic and Health Survey 2006-07." Mbabane, Swaziland: Central Statistical Office [Swaziland] and Macro International Inc.

Shisana, O., T. Rehle, L.C. Simbayi, K. Zuma, S. Jooste, V. Pillay-van-Wyk, N. Mbelle, J Van Zyl, W. Parker, N.P. Zungu, S. Pezi, & the SABSSM III Implementation Team. 2009. "South African national HIV prevalence, incidence, behaviour and communication survey 2008: A turning tide among teenagers?". Cape Town: HSRC Press.

Tanser, F., V. Hosegood, T. Bärnighausen, K. Herbst, M. Nyirenda, W. Muhwava, C. Newell, J. Viljoen, T. Mutevedzi, & M.-L. Newell. 2007. "Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey." *International Journal of Epidemiology*:1-6.

UNAIDS. 2006. "AIDS epidemic update: special report on HIV/AIDS: December 2006." Geveva: UNAIDS/WHO.

UNAIDS/WHO. 2003. "Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups ". Geneva: UNAIDS/WHO, Working Group on Global HIV/AIDS and STI Surveillance.

Walker, N., K.A. Stanecki, T. Brown, J. Stover, S. Lazzari, J.M. Garcia-Calleja, B. Schwartlander, & P.D. Ghys. 2003. "Methods and procedures for estimating HIV/AIDS and its impact: the UNAIDS/WHO estimates for the end of 2001." *AIDS* 17(15):2215-2225.

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

WHO. 2007. "World health statistics 2007." Geneva: World Health Organization (WHO).

WHO. 2008. "The global burden of disease: 2004 update." Geneva: World Health Organization (WHO).