

A multistage approach to adjusting for non-response when estimating HIV prevalence using survey data

Samuel OM Manda - *Chief Specialist and Professor of Biostatistics, Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa.*

Most countries in sub-Saharan Africa (SSA) have relied on antenatal clinic (ANC)-based sentinel surveillance systems to estimate HIV population prevalence and to monitor trends and progress made against the epidemic (1). These ANC surveillance data are still widely used in many settings for estimating the HIV burden nationally and at sub-national levels (2-3). However, these data have well-known and documented coverage and representativeness limitations including an excess of sentinel ANC sites in urban areas and other readily accessible locations; the exclusion of men; potential difference in HIV prevalence between pregnant and non-pregnant women; and potentially HIV-associated infertility (4). These limitations are of great concern since these data are used as inputs for deriving a number of national HIV statistics in many countries using the Spectrum/Estimation and Projection Package (EPP) software (5).

Thus, a number of countries have started conducting nationally representative population-based surveys that include HIV testing. For example, Demographic and Health Surveys that include HIV testing have been conducted in over 30 countries in Africa, Asia, Latin America, and Eastern Europe (4). These data are increasingly becoming the main source of HIV prevalence estimates and are a marked improvement as they provide superior estimates of HIV prevalence if accompanied by high HIV testing uptake (6) and even in assessing geographical distribution of adult HIV prevalence (7-9). However, these data have high refusal rates, which might hamper the interpretation and generalisation of HIV burden estimates (10-11). Refusal to HIV testing among women ranged from 1% to 17% in 18 of the 19 countries (10). For instance, in the 2004 Malawi Demographic Health Survey,

30% of women and 37% of men in the HIV subsample refused to consent to HIV testing (12); in South Africa only 67.5% of eligible individuals consented to HIV testing, with Black Africans at 73.3%, having the highest HIV testing response among the four race groups (13). One of the main reasons for a subject to refuse HIV testing is knowledge of their HIV positive status as observed in repeated HIV serosurveys in rural Malawi (14-15). A number of techniques have been used to account for the effects of non-response in HIV prevalence estimates; complete imputation of missing data including HIV status and other data (16) and HIV mortality rates (17). As an alternative, this paper proposes a multi-stage approach to obtain optimal HIV statistics using population-based HIV surveys. These stages range from deriving an appropriate HIV test response regression model; to deriving a HIV prediction equation model to impute HIV status for subjects who refused HIV testing and survey interviews; to fitting recently developed Bayesian multivariate spatial models to obtain smoothed HIV prevalence maps at the appropriate sub-national level by combining population-based and ANC HIV data sources (18).

Multi-stage approach

Firstly, a probit regression equation for HIV testing response is estimated for all HIV selected testing sample. This is used to compute inverse probability weights (IPWs), which are obtained from the inverse of estimated probabilities of accepting HIV testing (18). Secondly, a HIV prediction model using the observed and known HIV status from the HIV consented and tested subjects is developed, also using a probit regression model. For disease mapping

purposes, individual-level HIV status (observed and predicted) are averaged at a desired sub-national geographic level to obtain the estimate of HIV prevalence rates. The individual weights used would vary according to use of a) inverse probability weights for those who were selected for HIV testing; b) individual interview weights for those who were interviewed for the main survey, but were not selected for the HIV testing sample; and c) household weights for those who were not interviewed and were not in the HIV testing sample (10). Finally, the aggregated population-based HIV prevalence are smoothed using shared spatial components modelling techniques (19) to incorporate information from ANC HIV data sources.

The data sources used in this study are from the Malawi Demographic and Health Survey (MDHS) of 2010 (MDHS 2010) and Malawi ANC HIV Prevalence Survey of 2010. Malawi is one of only a few countries where a Demographic and Health Survey has collected nationally representative HIV prevalence data. The DHS sampled a total of 27,000 households and involved nearly 24,000 female and 7,000 male respondents. In every third household, blood specimens were collected for HIV testing from all women aged 15-49 and men aged 15-54 who consented to HIV testing. ANC HIV surveillance in Malawi has been conducted every one to two years since 1994 using a consistent methodology in the same population group. For the 2010 ANC survey, 23,788 pregnant women were enrolled from 28 urban sites and 26 rural sites distributed across the three regions of Malawi. Despite recent declines in HIV prevalence, Malawi remains among the group of countries with the largest HIV epidemics in the world. The country's HIV prevalence in adults (15-49 years) was estimated at 11.8% in 2010; ANC HIV prevalence was estimated at 10.6% in 2010 (20).

The fitting of the shared spatial component model was done using Bayesian based Markov Chain Monte Carlo techniques at the level of a district. There are currently a total of 28

administrative districts in Malawi, but two districts were combined into the original districts from which they separated around 2000. The logarithms of the two district-level HIV prevalence rates were modelled and an asymmetric formulation of the shared component model was adopted (17). Poverty and population density were used as district-level contextual factors for the ecological models. Each of these contextual covariates was partitioned into fourths; this categorization enables effects to be detected at the extremes of the range. The WinBUGS Bayesian package was used code for fitting the bivariate spatial model.

Slightly higher HIV prevalence when adjusting for non-response

A total of 23,020 women and 6,805 men (aged 15-49 years), were interviewed, in the MDHS 2010; resulting in 97% and 92% response rates for women and men, respectively. In the 1:3 households selected for HIV testing, 1,104 of the 7,391 selected men and 673 of the 8,174 selected women did not consent to HIV testing, resulting in an overall 89% consent rate for HIV testing.

Of interest was that females (92%) had higher rates of consenting compared to males (85%), and also those in rural areas (89%) compared to urban areas (85%). Type of employment, religions background and education level were associated with acceptance of HIV testing. Other Christians and Muslims are significantly less likely to accept HIV testing. The district-level distribution of HIV testing refusal rates is shown in Figure 1, where higher refusal rates of HIV testing were observed in the south-eastern and in most northern parts of the country.

Using the MDHS 2010 HIV response weight, the adult HIV prevalence was estimated at 10.61% with a 95% confidence interval (CI) of 9.9% to 11.33%. Using the inverse probability weighting (IPW), the prevalence was 10.19% (9.69%, 10.71%). On the other hand, ANC 2010 HIV prevalence per site had a median of 10.6% (1.85%, 24.09%).

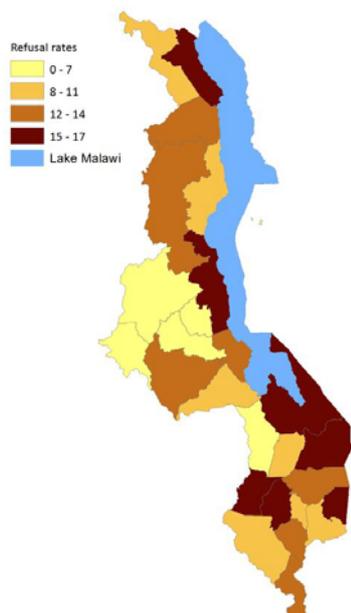


Figure 1. DHS refusal rates to HIV testing, Malawi, 2010



Figure 2a. Observed and imputed DHS HIV prevalence (%), Malawi, 2010

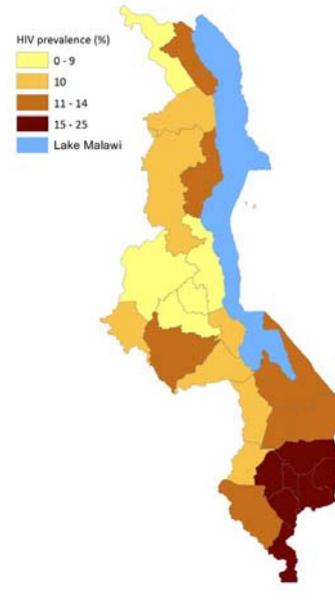


Figure 2b. Antenatal HIV prevalence (5), Malawi, 2010

The district-level MDHS 2010 HIV prevalence had a mean and median of 10.45% and 10.09% and ranged from 3.16% to 18.00%; for ANC 2010, the district-level HIV prevalence had a mean and median of 12.12% and 11.01% and ranged from 5.54% to 25.2%. This showed great variation in HIV prevalence rates between the districts. A number of factors, such as gender, residence, employment, marital status, ethnicity, condom use, multiple sex partners, and risky sex in the past 12 months, are significantly associated with HIV status (Table 1 - downloadable). The discriminative ability of the resultant model had an area under the receiver operating curve (AROC) of 0.7757% (0.7680%, 0.7833%), which is deemed satisfactory. After predicting HIV status for the non-tested) subjects (n =

17,406, and assigning to all the subjects their differing weights as described above, the overall HIV prevalence using all the 31,139 subjects was estimated at 11.05% (10.80%, 11.30%).

Figures 2a-b show total MDHS 2010 weighted and antenatal 2010 HIV prevalence in Malawi 2010. Higher MDHS 2010 HIV prevalence rates were observed in the mostly southern districts, which mirrors the ANC HIV geographical distribution. The district-level distribution for poverty level and population density were evenly spread, but southern districts bear the highest burden of poverty (maps not shown). Thus districts with high HIV prevalence have high levels of poverty and population density.

Table 2: Estimated covariate effects with associated 95% credible intervals using the Bivariate Spatial Analysis: Malawi 2010

Characteristics	DHS HIV	ANC HIV
Poverty fourths		
I (Lowest)	0.0	0.0
II	0.11 (-0.18, 0.40)	0.16 (-0.08, 0.39)
III	0.10 (-0.18, 0.36)	0.24 (0.01, 0.45)
IV (Highest)	-0.00026 (-0.31, 0.31)	0.17 (-0.09, 0.42)
Population density fourths		
I (Lowest)	0.0	0.0
II	-0.09 (-0.41, 0.26)	-0.11 (-0.38, 0.16)
III	0.09 (-0.19, 0.38)	-0.06 (-0.29, 0.18)
IV (Highest)	0.34 (0.02, 0.65)	0.37 (0.11, 0.63)

The results of the effects of the contextual factors from fitting the smoothed risks using the bivariate shared model are shown in Table 2. High population density areas were significantly associated with high HIV prevalence. Increased district-level poverty was associated with increased HIV rates, but this was not statistically significant. The distribution of specific covariate-adjusted smoothed HIV risks from fitting the bivariate spatial model were generally similar to the distribution of the observed and weighted HIV prevalence maps; but clusters of high HIV-prevalent districts were now more apparent (maps not shown). The estimates of the effects of the shared component (which we took to act as a surrogate for high HIV risky behaviours) had a larger effect on HIV incidence in the southern parts of the country around the high population density and urban areas (Figure 3a). Figure 3b, shows the excess risk attributable to ANC HIV, which is much

larger in the central-eastern and northern parts of the country.

We have found that adjusting for non-response in estimating the national HIV prevalence using a representatively population-based survey resulted in a slightly higher HIV rate. This finding confirms the well-held view that HIV testing refusals may have higher HIV prevalence. However, the confirmation should be treated with caution as it may as well depend on the type of adjustment and prediction models and the set of socio-demographic and behavioural characteristics included in the two probit regression models. Another study proposed a similar HIV prevalence adjustment methodology, but the ANC data was used to drive a HIV prediction model (21). The approach presented here enhanced this by combining spatial smoothing techniques based on novel application of multivariate spatial models, inverse probability weighting, and HIV prediction model.

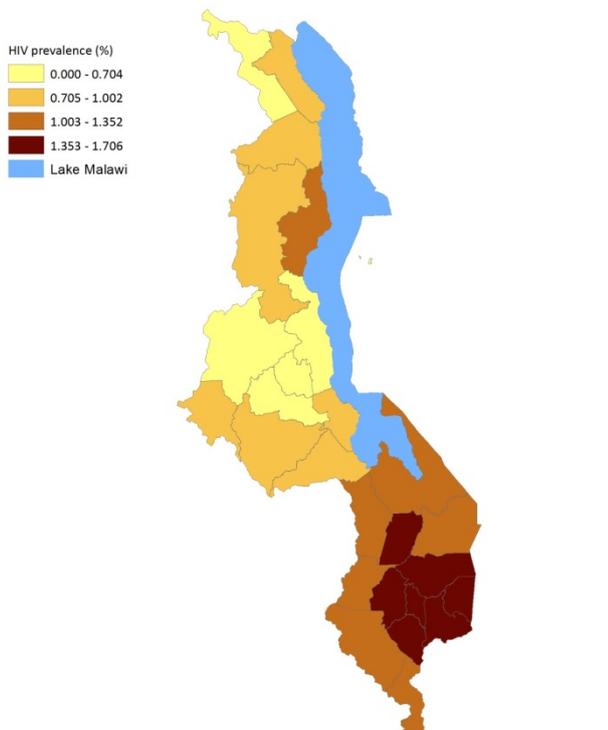


Figure 3a. Shared HIV prevalence (%), Malawi, 2010

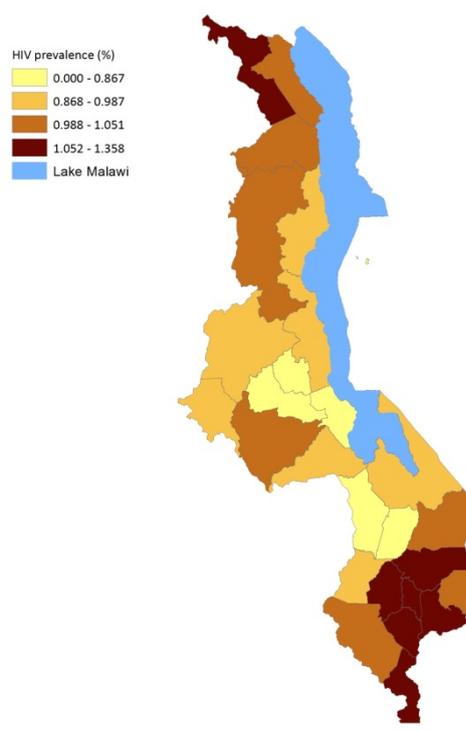


Figure 3b. Differential ANC HIV prevalence (%), Malawi, 2010

Downward bias in the HIV prevalence estimates can also result from many competing factors including migration of HIV risk subjects and non-response due infidelity suspicion (15). Thus, in the absence of a true HIV estimate among non-tested subjects, these conjectures are just indicative of the down effect of non-response on the HIV prevalence estimate

This work presented here is an abridged version of a published article (7) which was co-authored by Lieketseng Masenyetse L, Bo Cai and Renate Meyer. I would like to take this opportunity to thank them for their contributions.

Samuel OM Manda - Chief Specialist and Professor of Biostatistics, Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa. Areas of interest: the methodological and application biostatistics; Markov chain Monte Carlo (MCMC) based Bayesian statistics, multivariate spatial statistics, sample survey, recurrent event history and longitudinal data analyses. samuel.manda@mrc.ac.zac

References:

1. Joint United Nations Programme on HIV/AIDS (UNAIDS) . Global Report: UNAIDS Report on the Global AIDS Epidemic, 2012. Geneva: UNAIDS; 2012.
2. Manda SOM, Lombard CL, Mosala T. Divergent spatial patterns in the prevalence of the human immunodeficiency virus (HIV) and syphilis in South African pregnant women. *Geospatial Healt*. 2012; 6(2):221-231.
3. Malawi Ministry of Health. HIV and syphilis sero-survey and national HIV prevalence and AIDS estimates report for 2010. Lilongwe: Malawi Ministry of Health; 2011.
4. Brookmeyer R. Measuring the HIV/AIDS Epidemic: Approaches and Challenges. *Epidemiological Review*. 2010; 32:26–37.
5. Stover J, Brown T, Marston M. Updates to the Spectrum/Estimation and Projection Package (EPP) model to estimate HIV trends for adults and children. *Sex Transm Infect* 2012; 88: i11-i16.
6. Boerma JT, Ghys DP, Walker N. Estimates of HIV-1 Prevalence from National Population-based Surveys as a New Gold Standard. *Lancet*. 2003;3 63: 1929–30.
7. Manda SOM, Masenyetse L, Bo, C, Meyer R. Mapping HIV Prevalence Using Population and Antenatal Sentinel-based HIV Surveys: A Multi-Stage Approach. *Population Health Metrics*.2015;13:22.
8. Kandala N-B, Brodish P, Buckner B, Foster S, Madise N. Millennium development goal 6 and HIV infection in Zambia: what can we learn from successive household surveys? *AIDS (London, England)*. 2011;25(1):95-106.
9. Niragire F, Achia TNO, Lyambabaje A, Ntaganira J. Bayesian Mapping of HIV Infection among Women of Reproductive Age in Rwanda. *PLoS ONE*. 2015; 10(3): e0119944.
10. Mishra V, Vaessen M, Boerma T, Way A, Barrère B, Cross A, et al.. HIV testing in population-based surveys: Experience from the Demographic and Health Surveys. [Bull World Health Organ](http://dx.doi.org/10.1186/1475-2875-84-7). 2006;84(7):537-45.
11. Garcí a-Calleja JM, Gouws E, Ghys PD. National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates. *Sexually Transmitted Infections*. 2006; 82 Suppl 3: iii64–70.
12. National Statistical Office (NSO) (Malawi) and ORC Macro. Malawi Demographic and Health Survey 2004. Calverton, Maryland: NSO and ORC Macro; 2005.
13. Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Zungu N, et al. South African National HIV Prevalence, Incidence and Behaviour Survey, 2012. Cape Town: HSRC Press; 2014.
14. Floyd S, Molesworth A, Dube A, Crampin AC, Houben R, Chihana M, Price A, Kayuni N, Saul J, French N, Glynn JR. Underestimation of HIV prevalence in surveys when some people already know their status, and ways to reduce the bias. *AIDS*. 2013;27:233-242. t
15. Obare F. Nonresponse in Repeat Population-Based Voluntary Counselling and Testing for HIV in Rural Malawi. *Demography*. 2010. 47(3): 651–665.
16. Chinomona A, Mwambi H. Multiple imputation for non-response when estimating HIV prevalence using survey data. *BMC Public Health*. 2015;15:1059.
17. Nyirenda M, Zaba B, Bärnighausen T, Hosegood V, Newell M-L . Adjusting HIV Prevalence for Survey Non-Response Using Mortality Rates: An Application of the Method Using Surveillance Data from Rural South Africa. *PLoS ONE*. 2010; 5(8): e12370.
18. Jones AM, Koolman X, Rice N. Health-related non-response in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models. *Health Economics and Data Group Working Paper 05/05*, University of York; 2005.
19. Manda SO, Feltbower RG, Gilthorpe MS. Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes. *Eur J Epidemiol*. 2009; 24: 743-752.
20. Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994–2010. *BMC Infect Dis*. 2014;14:285.
21. Ivaschenko O, Lanjouw P. A New Approach to Producing Geographic Profiles of HIV Prevalence: An Application to Malawi. Policy Research Working Paper 5207. Human Development Economics, Europe and Central Asia Region; and Poverty and Inequality Team, Development Research Group, World Bank; 2010.