CePeD
Centre Population & Développement

juin 2009 • 03

Working Papers du CEPED

Estimating effect of non response on HIV prevalence estimates from Demographic and Health Surveys

Joseph Larmarange, Roselyne Vallo, Seydou Yaro, Philippe Msellati, Nicolas Méda, Benoît Ferry

Centre Population et Développement
UMR 196 CEPED, Université Paris Descartes, INED, IRD
http://www.ceped.org/wp

Contact • Corresponding Author

Joseph LARMARANGE, Démographe IRD
CEPED - UMR 196 Université Paris Descartes, INED, IRD
joseph@larmarange.net – joseph.larmarange@ceped.org

Citation recommandée • Recommended citation

Larmarange J., Vallo R., Yaro S., Msellati P., Méda N., Ferry B., « Estimating effect of non response on HIV prevalence estimates from Demographic and Health Surveys », *Working Paper du CEPED*, number 03, UMR 196 CEPED, Université Paris Descartes, INED, IRD, Paris, July 2009.
Available at http://www.ceped.org/wp

Remerciements • Acknowledgements

This study was supported by the French national AIDS research agency (ANRS – project 12114).

CEPED • Centre Population et Développement

UMR 196 CEPED, Université Paris Descartes, INED, IRD
45, rue des Saints-Pères, 75006 PARIS • 221, boulevard Davout, 75020 PARIS
Pour toute correspondance, s'adresser au CEPED, 221 boulevard Davout, 75020 PARIS
http://www.ceped.org/ • contact@ceped.org

Estimating effect of non response on HIV prevalence estimates from Demographic and Health Surveys

Joseph Larmarange[*], *Roselyne Vallo*[†], *Seydou Yaro*[‡], *Philippe Msellati*[&], *Nicolas Méda*[‡], *Benoît Ferry*[*]

Abstract

In most countries in Sub-Saharan Africa, Demographic and Health Surveys (DHS) with HIV testing became the only measure of HIV prevalence in general population. Significant non response rates were often cited to explain differences between DHS results and estimations from sentinel surveillance in antenatal clinics. The objective of this paper consists to predict with multivariate models the prevalence of non tested persons in order to estimate the effect of non response on national HIV prevalence measure.

We used data from 9 DHS, conducted in sub-Saharan Africa, where HIV results could be linked with data from household and individual questionnaires. Logistic regressions were performed for each country, separately for men and women 15-49 years old, with a common set of predictor variables. For each group, adjusted prevalence was calculated by using observed prevalence for tested people and estimated probability to be HIV positive for non tested persons.

The non response rates in these 9 studies vary from 7.9% to 39.3%. Estimated prevalence of non tested persons is usually higher than observed prevalence of tested persons (13 groups on 18) Nevertheless, ratios of adjusted prevalence to observed prevalence remain relatively close to 1 (from 0.970 to 1.109). Differences between adjusted and observed prevalence is less than 0.32. A significant negative correlation was found between non-response rates and ratios of non-tested to tested, but there is no correlation between ratios of adjusted to observed prevalence and proportions of non-tested.

The overall effect of non response biases on national HIV estimates tends to be small and remains inferior to sample variations. If adjustments need to be interpreted with caution due to the limited information available to predict the prevalence of non tested people, we can conclude that national population-based surveys can provide quality and representative national HIV prevalence estimates.

***Keywords**: HIV Infections, Population surveillance, Health Surveys, Developing Countries, Prevalence Estimates.*

* IRD, CEPED – UMR 196 Université Paris Descartes, INIED, IRD.

† Université de Montpellier I – EA 4205.

‡ Centre Muraz, Bobo Dioulasso, Burkina Faso.

& UMR 145 IRD/Université de Montpellier, Centre de coordination et de recherche du site ANRS Cameroun, Yaoundé.

Introduction

In countries with generalized epidemics, most data on HIV prevalence in the adult population are generated by sentinel surveillance of pregnant women attending selected antenatal clinics (ANC). Such a surveillance system has been implemented in most countries of sub-Saharan Africa: 39 out of 43 in 2003 (UNAIDS/WHO, 2003). ANC sentinel surveillance, which was initially developed to monitor epidemic trends (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2000), was expanded and used for national estimates of HIV prevalence by UNAIDS (Schwartlander *et al.*, 1999).

For a few years, several countries have conducted national population-based surveys that include HIV testing in the general population, to improve their epidemiological monitoring. In sub-Saharan Africa, most of those surveys are Demographic and Health Surveys (DHS) or AIDS Impact Surveys (AIS). DHS, which have existed since 1984, have become one of the four components of the "Monitoring and Evaluation to Assess and Use Results" programme (MEASURE) founded in 1997 by USAID and implemented by Macro International (Measure DHS, 2006).

Since 2001, several DHS and AIS have measured HIV prevalence. For many countries, DHS with HIV testing is currently the first and only measurement of national HIV prevalence in the general adult population. In some countries, the results provided by DHS have diverged from estimates based on ANC data. Non-response rates (refusal or absence) in DHS were cited several times to explain those variations (Boerma *et al.*, 2003; UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2005), non-tested persons supposed to present an higher HIV prevalence (Hull, 1988; Jenum, 1988). However, their impact probably remains limited (Bignami-Van Assche *et al.*, 2005). One of the main explanatory factors (Montana, 2006) would seem to be the location of sentinel sites, with rural areas being under-represented (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2005). It is accepted today that population-based surveys provide better estimates of national prevalence at a given moment (UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, 2005). UNAIDS henceforth calibrates its projections on results from that type of surveys (Ghys *et al.*, 2004; UNAIDS/WHO, 2005). In its software EPP 2005, trends are estimated from ANC data, while the level is determined by a population-based survey (Brown *et al.*, 2006).

However, the amplitude of the bias due to the proportion of persons who were not tested in national population-based surveys remains an important interrogation in order to improve HIV estimates at national level. In 2006, two papers tried to estimate the magnitude of this bias. The first paper, in a scenario assuming that non-responders have twice the HIV prevalence of those who fully participated in the survey, suggests that individual non-response could result in an adjusted HIV prevalence 1.03 to 1.34 times higher than the observed prevalence (Garcia-Calleja *et al.*, 2006). In the second paper, HIV prevalence among non-responding males and females of five surveys was predicted using multivariate statistical models for those who where tested. The authors concluded that, although HIV prevalence tended to be higher in non-tested males and females than in those tested, the overall effects of non-response on the observed national estimates of HIV prevalence were insignificant.

The objective of this paper is to enlarge this analysis to nine DHS or AIS, to estimate the prevalence of HIV among non-tested persons and to determine if there is a link between the proportion of non-response and the amplitude of the bias.

Method

DHS sample design

DHS and AIS use a two-stage stratified sample design. Each country is first stratified into administrative regions and then into rural and urban. Often, the capital of the country is considered as a single region. Primary units correspond to the enumeration areas at the last national population census. At the first stage of sampling, enumeration areas or clusters are selected by stratum, with a probability proportional to the number of households at the last census. Consequently, the spatial distribution of clusters can be regarded as a proxy of population density.

After a complete census of households in each selected cluster, a second sample is chosen to determine selected households where the chief of the household is interviewed about all the members of the household. All eligible women (usually aged 15-49) are selected for the women survey. Only a part of these households are then selected for the HIV survey and the men survey. In selected households, all eligible men (usually aged 15-59) and all eligible women are tested for HIV after consent.

Household response rates were relatively good: more than 95% (Mishra *et al.*, 2006). For persons living in a household who were not surveyed, we don't have any information about them. So, it is not possible to estimate their prevalence. In this paper, only persons recorded in the household database, eligible to HIV test and aged 15-49 were retained for the analysis. Household sampling weights were applied. We used eight DHS and one AIS for which data were available and HIV results linkable to individual data.

Non-tested persons

Among eligible persons for HIV test, we can distinct four groups: persons who were tested and interviewed for the individual questionnaire, persons tested but not interviewed for the individual questionnaire, persons interviewed for the individual questionnaire but not tested and persons not tested and not interviewed for which we just have information of the household questionnaire. Table 1 present the proportion of this four groups for nine surveys. A person can be not tested or/and not interviewed for several reasons: absence, refusal, incapacity or technical problem (see Table 1).

HIV prevalence estimation for non-tested individuals

Multivariate models were used to estimate for each non-tested persons the probability to be HIV positive. Prevalence of non-tested was calculated as the average of this probability. The models were built separately by sex and country. Two kinds of models were used. For non-tested and non-interviewed persons, a logistic regression was performed on all tested persons. Several variables from the household questionnaire were introduced in the model: size of the household, type of residence, instruction level, having a radio and/or a television, wealth index, age and region of residence. For non-tested but interviewed persons, a second model was performed on tested and interviewed persons. The same household variables were introduced in the model and variables from the individual survey were added: marital status, working status, having a sexually transmitted infections, a genital sore/ulcer or a genital discharge in the last 12 months, having used a condom at the last sexual intercourse (except for men in Lesotho), willing to care about HIV/AIDS, number of sexual intercourses in the last month, age at the first sexual intercourse, number of sexual partners in the last 12 months, smoking (except for men in Burkina Faso, Cameroon, Senegal and United Republic

of Tanzania and for women in United Republic of Tanzania), male and female circumcision (except for women in Lesotho and Malawi).

Adjusted prevalence was calculated from observed HIV status for tested persons and from probability to be HIV positive, estimated by the models, for non-tested persons. Two ratios were calculated: estimated prevalence of non-tested persons on observed prevalence and adjusted prevalence on observed prevalence.

Results

Ratio of estimated HIV prevalence of non-tested persons on observed prevalence varies from 0.857 to 1.864. In five on nine surveys, this ratio is less than 1 for women. But, for men, estimated prevalence of non-tested is always higher than observed prevalence. Estimated prevalence of non-tested is statistically different from observed prevalence only for men in Cameroon and women in Ethiopia. Figure 1 shows a significant negative correlation between this ratio and the proportion of non-tested persons. This correlation is stronger for men than for women. Less is the proportion of non-tested persons, higher is their over-prevalence (see Table 2 and Figure 1).

Adjusted prevalence depends on two things: the proportion of non-tested persons and the value of the ratio of non-tested to tested. As this ratio is negatively correlated to the proportion of non-tested, these two effects compensate themselves and the ratio of adjusted prevalence to observed prevalence remains close to one (from 0.970 to 1.109). There is no statistically significant correlation between the ratio of adjusted to observed prevalence and the proportion of non-tested persons (see figure 1). Observed prevalence is a good indicator of the national level of HIV prevalence. The differences between observed and adjusted prevalence are inferior to 0.32. The adjusted prevalence is always included inside the 95% confidence interval of the observed prevalence.

Discussion

Our results are close to those of the precedent analysis conducted by Mishra et al. (Mishra *et al.*, 2006). We used the same approach (logistic regression), but with a methodology slightly different (sampling weight used, age range for men and variables included in the models). The ratios, for the five common surveys, are not exactly the same, but the level orders are almost the same. In comparison with this precedent analysis, our paper shows that biases due to non-tested persons are the same for four other surveys and examines the relation between the over-prevalence of non-tested persons and the proportion of them.

The variables included in the models are not necessarily direct determinants of prevalence. But in order to estimate prevalence of non-tested persons, we have to use available data. However, if some variables are not determinants of HIV status, they remain discriminant, highlighting effects of other variables not collected in DHS or AIS.

To asses the capacity of the models to predict HIV prevalence of non tested persons, we can compare estimated prevalence among non tested people who refused the test and people who were absent. According to previous results (Hull, 1988; Jenum, 1988), the first group would show a higher prevalence than observed people but not the second group. In Burkina Faso, predicted prevalence among refusal is 2.9% [95%CI: 1.6-5.2] versus 1.7% [0.5-4.8] among absent people and 1.8% [1.5-2.1] among tested persons. It is respectively 7.7% [5.9-10.1] versus 5.9% [3.6-9.3] and 5.4% [5.0-5.9] in

Cameroon, 2.6% [1.6-3.9] versus 1.6% [0.8-3.2] and 2.2% [1.9-2.5] in Ghana. In Ethiopia, there were no absent people. The predicted prevalence of refusal is 2.1% [1.5-2.9] versus 1.3% [1.1-1.5] for observed prevalence. In Senegal, the numbers of HIV positive persons are too small in the surevy to get significant results. In Kenya, Lesotho, Malawi and Tanzania, the reason of non test is not available in HIV datasets. These results highlight the ability of this kind of approach to produce likely predictions.

We cannot estimate the HIV prevalence of some population groups with this approach: people living in a household not surveyed (absence or refusal for the household questionnaire) and people not living in a household, such as those living on the street or in institutions (e.g. military barracks, prisons, refugee camps, boarding schools…), for which we don't have any information in DHS or AIS. However, it would be possible to estimate biases due to non-households population if specific surveys on this population have been conducted in the country and if it is possible to estimate its size. The methodological approach developed in this paper focuses only on the biases due to non tested persons. So it doesn't take into account the other sources of biases.

This methodology can be applied only when the survey provides other information on people not documented concerning the analysed variable. Predicted prevalence is pertinent to adjust the observed prevalence but it cannot be used to analyse the determinants.

In conclusion, population-based surveys can provide representative and quality national estimates of HIV prevalence levels in countries with generalized epidemics. Biases of non-response remain minor to sample variations and inferior to biases in selection of antenatal clinics in surveillance system. So, the actual approach of UNAIDS in EPP (Estimation and Projection Package) is efficient. It consists to use antenatal clinical data to estimate trends of HIV epidemics and results of national population-based surveys to determine levels of the epidemics.

References

BIGNAMI-VAN ASSCHE, S., J. A. SALOMON & C. J. MURRAY (2005). Evidence from National Population-Based Surveys on Bias in Antenatal Clinic-Based Estimates of HIV Prevalence. *Population Association of America Annual Meeting*, Philadelphia, Pennsylvania (US).

BOERMA, J. T., P. D. GHYS & N. WALKER (2003). Estimates of HIV-1 prevalence from national population-based surveys as a new gold standard. *Lancet*, 362(9399), 1929-31.

BROWN, T., N. C. GRASSLY, G. GARNETT & K. STANECKI (2006). Improving projections at the country level: the UNAIDS Estimation and Projection Package 2005. *Sexually Transmitted Infections*, 82(suppl_3), iii34-40.

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. *Sexually Transmitted Infections*, 82(suppl_3), iii64-70.

GHYS, P. D., T. BROWN, N. C. GRASSLY, G. GARNETT, K. A. STANECKI, J. STOVER & N. WALKER (2004). The UNAIDS Estimation and Projection Package: a software package to estimate and project national HIV epidemics. *Sexually Transmitted Infections*, 80 Suppl 1, I5-I9.

HULL, H. F. (1988). Comparison of HIV-antibody prevalence in patients consenting to and declining HIV-antibody testing in an STD clinic. *Journal of the American Medical Association*, 260(7), 935-938.

JENUM, P. (1988). Anti-HIV screening of pregnant women in south-east Norway. *NIPH Annals*, 11, 53-58.

MEASURE DHS (2006). *Measure DHS Brochure*. Retrieved 2006, july the 4th, from http://www.measuredhs.com/pdfs/DHS%20brochure%20with%20branding-updated.pdf.

MISHRA, V., M. VAESSEN, J. T. BOERMA, F. ARNOLD, A. WAY, B. BARRERE, A. CROSS, R. HONG & J. SANGHA (2006). HIV testing in national population-based surveys: experience from the Demographic and Health Surveys. *Bulletin of the World Health Organization*, 84(7), 537-45.

MONTANA, L. (2006). Comparison of HIV prevalence estimates from population-based surveys and ANC sentinel surveillance data. XVI International AIDS Conference. Toronto (CA).

SCHWARTLANDER, B., K. A. STANECKI, T. BROWN, P. O. WAY, R. MONASCH, J. CHIN, D. TARANTOLA & N. WALKER (1999). Country-specific estimates and models of HIV and AIDS: methods and limitations. *AIDS*, 13(17), 2445-58.

UNAIDS/WHO (2003). *Reconciling antenatal clinic-based surveillance and population-based survey estimates of HIV prevalence in sub-Saharan Africa*. Genève (CH): UNAIDS/WHO.

UNAIDS/WHO (2005). EPP manual for generalized epidemics. Geneva, UNAIDS/WHO**:** 54.

UNAIDS/WHO WORKING GROUP ON GLOBAL HIV/AIDS AND STI SURVEILLANCE (2000). *Guidelines for Second Generation HIV Surveillance*. Genève (CH): UNAIDS/WHO.

UNAIDS/WHO WORKING GROUP ON GLOBAL HIV/AIDS AND STI SURVEILLANCE (2005). *Guidelines for measuring national HIV prevalence in population-based surveys*. Genève (CH): UNAIDS/WHO.

Table 1. Proportion of people aged 15-49 eligible for HIV testing who were tested and interviewed for nine DHS or AIS

Country	Year	Sex	Neither HIV testing nor individual survey	Individual survey without HIV testing	HIV testing without individual survey	HIV testing and individual survey	No. eligible for HIV testing
Burkina Faso	2003	Men	6.9	6.4	2.1	84.6	3501
		Women	2.6	5.3	0.7	91.4	4607
Cameroon	2004	Men	6.4	4.7	1.0	87.9	5146
		Women	3.8	4.6	1.7	89.8	5759
Ethiopia	2005	Men	8.6	9.1	0.1	82.2	6139
		Women	3.4	9.2	0.2	87.3	6963
Ghana	2003	Men	6.2	14.3	0.1	79.4	4636
		Women	4.2	6.5	0.2	89.1	5845
Kenya	2003	Men	12.9	15.0	0.5	71.6	3970
		Women	5.0	16.3	0.3	78.5	4293
Lesotho	2004	Men	16.0	18.2	0.4	65.4	2926
		Women	6.0	15.0	0.3	78.7	3672
Malawi	2004	Men	13.7	25.6	0.0	60.7	3663
		Women	5.6	27.3	0.0	67.1	4057
Senegal	2005	Men	12.6	12.0	1.0	74.5	3997
		Women	5.6	10.9	0.9	82.6	5342
United Republic of Tanzania	2003	Men	9.0	15.7	0.0	75.3	6282
		Women	4.2	13.6	0.0	82.2	7231

Table 2. Observed prevalence, predicted prevalence for non-tested and adjusted prevalence by country and sex (15-49 years old)

Country (year)	Observed prevalence among those tested (95% CI)		Predicted prevalence among those not tested (95% CI)		Ratio of non-tested to tested	Adjusted prevalence among all eligible respondents (95% CI)		Ratio of adjusted to tested	Proportion of non-tested
Burkina Faso 2003									
Men	1.8	(1.3-2.2)	2.1	(1.9-2.4)	1.196	1.8	(1.4-2.2)	1.026	13.4
Women	1.8	(1.4-2.2)	3.1	(2.5-3.7)	1.760	1.9	(1.5-2.2)	1.060	7.9
Both	1.8	(1.5-2.1)	2.6	(2.3-2.8)	1.443	1.8	(1.6-2.1)	1.045	10.3
Cameroon 2004									
Men	4.1	(3.5-4.6)	5.7	(5.2-6.2)	1.406*	4.2	(3.7-4.7)	1.045	11.1
Women	6.6	(6.0-7.3)	8.4	(7.7-9.2)	1.272	6.8	(6.2-7.4)	1.023	8.5
Both	5.4	(5.0-5.9)	7.0	(6.5-7.4)	1.281*	5.6	(5.2-6.0)	1.027	9.7
Ethiopia 2005									
Men	0.9	(0.6-1.2)	1.2	(1.0-1.4)	1.336	1.0	(0.7-1.2)	1.059	17.7
Women	1.7	(1.4-2.0)	3.2	(2.7-3.7)	1.864*	1.9	(1.6-2.2)	1.109	12.6
Both	1.3	(1.1-1.5)	2.1	(1.8-2.3)	1.556*	1.4	(1.3-1.6)	1.083	15.0
Ghana 2003									
Men	1.4	(1.0-1.8)	1.9	(1.6-2.1)	1.320	1.5	(1.2-1.8)	1.066	20.5
Women	2.7	(2.3-3.1)	2.6	(2.3-2.8)	0.949	2.7	(2.3-3.1)	0.995	10.7
Both	2.2	(1.9-2.5)	2.2	(2.0-2.3)	0.990	2.2	(1.9-2.4)	0.999	15.0
Kenya 2003									
Men	4.7	(3.9-5.5)	5.0	(4.6-5.5)	1.074	4.8	(4.2-5.4)	1.021	27.9
Women	8.7	(7.8-9.7)	7.5	(7.0-8.0)	0.857	8.5	(7.7-9.2)	0.970	21.3
Both	6.9	(6.3-7.5)	6.2	(5.8-6.5)	0.894	6.7	(6.2-7.2)	0.974	24.4
Lesotho 2004									
Men	19.0	(17.3-20.8)	19.2	(18.2-20.2)	1.009	19.1	(17.9-20.3)	1.003	34.2
Women	26.0	(24.4-27.5)	25.3	(24.2-26.5)	0.976	25.8	(24.5-27.1)	0.995	21.0
Both	23.2	(22.0-24.4)	21.9	(21.1-22.6)	0.943	22.8	(21.9-23.7)	0.985	26.9
Malawi 2004									
Men	10.1	(8.8-11.3)	10.5	(9.8-11.2)	1.044	10.2	(9.4-11.0)	1.017	39.3
Women	13.9	(12.6-15.2)	12.9	(12.3-13.5)	0.929	13.6	(12.7-14.5)	0.977	32.9
Both	12.2	(11.3-13.1)	11.7	(11.2-12.1)	0.958	12.0	(11.4-12.6)	0.985	35.9
Senegal 2005									
Men	0.5	(0.2-0.7)	0.5	(0.4-0.7)	1.133	0.5	(0.3-0.7)	1.033	24.6
Women	0.9	(0.6-1.2)	0.8	(0.7-0.9)	0.868	0.9	(0.7-1.1)	0.978	16.4
Both	0.7	(0.5-0.9)	0.6	(0.6-0.7)	0.889	0.7	(0.6-0.9)	0.978	19.9
United republic of Tanzania 2003									
Men	6.0	(5.3-6.7)	7.1	(6.7-7.5)	1.181	6.3	(5.8-6.8)	1.045	24.7
Women	7.5	(6.9-8.2)	8.4	(7.9-8.9)	1.119	7.7	(7.1-8.2)	1.021	17.8
Both	6.9	(6.4-7.3)	7.7	(7.4-8.0)	1.123	7.0	(6.6-7.4)	1.026	21.0

CI = confidence interval

* Predicted HIV prevalence among non-tested is statistically different at 5% from observed prevalence (t test).

Figure 1. Correlations between prevalence ratios and proportion of non-tested