

The effect of participant non-response on the HIV prevalence estimates in a population based survey in two informal settlements in Nairobi city

Authors: Abdhahah K Ziraba ^{1,2,§}, Nyovani Madise ³, Mwau Matilu ⁴, Eliya Zulu ¹, John Kebaso ¹, Samoel Khamadi ⁴, Vincent Okoth ⁴, and Alex Ezeh ¹

¹ African Population and Health Research Center, Shelter Afrique Centre, Longonot Road, Upper Hill.

P. O. Box 10787, 00100, Nairobi Kenya

² Department of Epidemiology and Population Health, Centre for Population Studies. London School of Hygiene and Tropical Medicine. 49-51 Bedford Square, London, WC1B 3DP, UK

³ University of Southampton, School of Social Sciences, Southampton, SO17 1BJ, United Kingdom

⁴ Kenya Medical Research Institute (KEMRI). P. O. Box 54840, 00200, Nairobi Kenya

§Corresponding author

Email addresses:

AKZ: abdhahah.ziraba@lshtm.ac.uk

NM: n.j.madise@soton.ac.uk

MM: mmwau@kemri.org

EZ: ezulu@aphrc.org

JK: jkebaso@yahoo.com

SK: skhamadi@kemri.org

VO: vokoth@kemri.org

AE: aezeh@aphrc.org

Abstract

Background:

Participant non-response in an HIV sero-survey can affect estimates of HIV prevalence. Non-response can arise from refusals to provide a blood sample or failure to trace a sampled individual. In a sero-survey conducted by the African Population and Health Research Center and Kenya Medical Research Centre in the slums of Nairobi, 46% of the sampled individuals did not provide a blood sample. This paper describes selective participation in the sero-survey and also estimates bias in HIV prevalence figures.

Methods:

The paper uses data derived from an HIV sero-survey nested in an ongoing demographic surveillance system. Non-response was assessed using logistic regression and multiple imputation methods to impute missing data for HIV status using a set of common variables available for all sampled participants.

Results:

Age, residence, high mobility, wealth and ethnicity were independent predictors of a sampled individual not being contacted. Individuals aged 30-34 years, females, individuals from the Kikuyu and Kamba ethnicity, the married and residents of Viwandani were all less likely to accept to be tested. Although men were less likely to be contacted, those found were more willing to be tested compared to females. The overall observed HIV prevalence was overestimated by 2%. The observed prevalence for male participants was underestimated by about 2% and that for females was overestimated by 4%. These differences were small and did not affect the overall estimate substantially as the observed estimates fell within the confidence limits of the corrected prevalence estimate.

Conclusions:

Non-response to HIV sero-survey in informal settlements is high. The effect on the overall prevalence estimate was however minimal.

Background:

Selective participation in a study can potentially skew estimates of the outcome of interest in a study population [1-5]. This is more likely to be the case if the reason or circumstances that influence low participation are in some way related to the main outcome. Non-response in HIV sero-surveys is mainly due to refusal to provide a blood sample for HIV testing or absenteeism of the sampled individual during the survey period. Several population based HIV sero-prevalence studies have reported varying non-response rates for HIV testing ranging from as low as 5% among men in Rwanda to 56% in Lesotho [2, 3]. A moderately high non-response rate (14.4%) for HIV testing for Kenya was reported in an earlier survey [2]. From the studies carried out so far on this topic, it has been shown that the effect of participant non-response on HIV prevalence estimates vary by certain characteristics such as gender and residence among others but generally the overall effect on national estimates are small unless the level of non-response is very high as was the case in Lesotho [1-3, 6].

HIV/AIDS remains a highly stigmatised disease with many people preferring either not to know their status or to keep it a secret [7, 8]. The preference of an individual not to participate in a sero-survey may partly be influenced by the fear of knowing their own HIV sero-status. On the other hand, those who know their status to be positive may participate in an HIV study in the hope that they can be helped or may choose not to participate as they see no immediate benefit. Personal perceived risk may be correlated with actual risk of HIV infection [9]. Perceptions about HIV risk are unlikely to be random among individuals in a population; they are likely to vary by defined individual characteristics such as race, religion,

ethnicity, and past behaviours including experience with drug use or sex work [10]. For that reason if participants who perceive themselves to be at a higher risk of contracting HIV do not to participate, then the prevalence estimates may be biased downwards and might affect the overall estimate.

Interviewers may fail to make contact with a sampled person for a number of reasons including temporary absence, work patterns, inability to locate the household/structure in which the sampled person lives, and out-migration among other reasons. Highly mobile individuals such as long distance truck drivers, security personnel and migrant workers often have a different level of exposure to the risk of HIV [11-14]. In highly mobile populations even though a good random sample is drawn, many sampled individuals may not be contacted. If a study population has a substantial proportion of highly mobile individuals who miss out on a sero-prevalence study, the estimates are likely to be biased downwards as less mobile individuals are over-represented in the effective sample interviewed during field work [2, 3]. On the other hand, if majority of the residents of a community are migrant workers who live away from their families, they are likely to be exposed to higher risks of HIV infection. To the extent that such persons are over-represented in a sero-prevalence survey, estimates are likely to be biased upwards.

The slum context:

Although informal settlements in Nairobi city are home to over 60% of Nairobi population [15], the informal nature of housing is likely to lead to underrepresentation of the slum population in national surveys given the difficulty involved in listing temporary housing structures. Until the project on which this paper is based was conducted, HIV prevalence in the informal settlements was unknown. Kenya has had at least two large population based

HIV testing surveys [16, 17]. The Kenya Demographic and Health Survey of 2003 put the HIV prevalence estimate for Nairobi province at 10%. Nyanza province had the highest prevalence rate of 15% and Kenya had an average prevalence of 6.7% [16]. There were differences in HIV prevalence rates by age, gender, ethnicity, rural-urban residence, educational attainment and wealth status. These differences have been observed in several other surveys in sub-Saharan countries [2, 3, 16]. A more recent survey, the Kenya AIDS Indicator Survey 2007 estimated the national prevalence to be 7% and 9% for Nairobi province [17]. The national surveys are however unable to provide HIV prevalence estimates for slums. Earlier behavioural research indicates that high risk sexual practices are prevalent in the informal settlements of Nairobi [18, 19]. Furthermore, recent work using verbal autopsies to establish causes of death without HIV status showed that HIV/AIDS and tuberculosis accounted for over 50% of adult mortality burden in the slums [20].

Against this background, The African Population and Health Research Center (APHRC) in partnership with The Kenya Medical Research Institute (KEMRI) carried out a survey to estimate the prevalence and risk factors for HIV in two informal settlements in Nairobi city. The two communities where the project was carried out are informal settlements characterized by poor housing, lack of clean water, poor sanitation, unemployment, poverty and overcrowding. Viwandani slum is located very close to the city's industrial area and it is home to many low income youths working in the industries close by. Korogocho is a more established slum settlement with a high proportion of men living with their spouses and children. Korogocho residents are predominantly either very low income earners or unemployed. Additionally residents of Viwandani are relatively more educated than those of Korogocho.

Results from the HIV sero-survey on which this paper is based showed that HIV prevalence in the slums was higher (11.5%) [21] than the current national average (7%) and that for Nairobi province (9%) [17]. The survey however, like many community based surveys faced a challenge of non-response with a sizeable proportion of sampled individuals being non-responders (46%). The desire to understand the effect of non-response on the prevalence estimates was the basis for this paper. We hypothesised that the HIV prevalence estimate in the survey was underestimated due to low participation of highly mobile community members. Specifically this paper aimed at: describing selective participation in the sero-survey by socio-demographic characteristics and also estimating the bias in the estimates of HIV prevalence.

Methodology:

Data used in this paper came from a cross sectional sero-survey carried out from September 2006 to November 2007. The project was nested in the Nairobi Urban Health and Demographic Surveillance System (NUHDSS) covering about 60,000 individuals in two slums; Korogocho and Viwandani. The NUHDSS database provided the sampling frame from which a random sample of eligible participants was drawn. Eligible individuals had to be residents in the demographic surveillance area and registered with the NUHDSS; aged between 15 to 54 years for men and 15 to 49 years for females.

A list of all sampled participants was generated with enough information to enable field workers to positively identify participants in their households. On the other hand, the questionnaires and blood sample filter papers didn't contain any identifiers except a new identification number (ID) to allow linkage to the NUHDSS data. A minimum of three visits

were made for individuals who were not found at home on the first visit and arrangements in terms of security were made to interview individuals who were identified as only available at odd hours (very early in the morning or late in the evening).

Participants were given information about the objectives of the study and information about their rights, potential risks and benefits was readout by the interviewer to those who could not read and those who could read were allowed enough time to read before making a decision. Those who accepted to participate affirmed it by signing the pre-written consent form. Minors (15-17 years old) who agreed to participate accented to participation by signing the Minor's consent form and their guardians also had to confirm their support by appending their signature or thumb prints. Individuals who consented to participate had the option of either responding to the interview only, providing a blood sample only, or providing both.

The survey used a questionnaire to collect data on knowledge of HIV prevention, HIV testing history, marriage and sexual activity, and circumcision. HIV status was determined using HIV serology on dried blood spots obtained from participants through a finger prick using Determine® HIV-1/HIV-2 (Abbott) and Uni-Gold™ Test kits according to manufacturer's instructions. The KEMRI HIV-Particle Agglutination test kit was used as the tie breaker for indeterminate findings. By design, participants could not get to know their HIV status results if they wanted to from the blood sample provided for the study. Those who wanted to know their status were provided standard pre-test counselling, testing and post-test counselling at a sister Voluntary Counselling and Testing Centre. Core variables from the NUHDSS database were linked anonymously to the survey and sero-data results using a linking ID.

Data analysis:

Descriptive and multivariate logistic regression analyses were carried out to describe participation by socio-demographic characteristics, assess determinants of sampled individuals being contacted and determinants for accepting to provide a blood sample for HIV testing or not.

To facilitate assessment of potential bias in the HIV sero-prevalence estimates, the analysis was carried out using multivariate multiple imputation techniques using a set of variables where data were available for the various interview outcome categories. Irrespective of the interview outcome, all sampled individuals had a minimum set of socio-demographic data derived from the Nairobi Urban Health and Demographic Surveillance System (NUHDSS) data base. Using multivariate imputation methods (Multivariate Imputation with Chained Equations-MICE) for missing data as described by van Buuren et al [22] and implemented in Stata software by Royston [23], the HIV prevalence among non-responders was estimated using a common set of variables including age, gender, residence, ethnicity, marital status, educational attainment, mobility index and socio-economic status (using a wealth index constructed from household items). For the category of individuals who were interviewed but not tested, an additional set of variables from the survey questionnaire was included in the imputation models for this subgroup. We used a logistic model to assess how predictive of HIV the socio-demographic characteristics were. This gave a pseudo R squared of about 11%. The model improved to 32% when additional variables from the survey (on HIV knowledge and attitudes and sexual behaviour) were added including; male circumcision; age at first sex; number of partners in last 12 months; high risk sex; *stigma indicator-“keep it a secret if family member is HIV positive*; condom use at last sex; and ever tested for HIV.

The mobility index was derived from NUHDSS records on each individual's movement episodes within and without of the slum per unit time as described later. For this paper non-response to HIV testing refers to the following categories; *i) No contact made with participant, ii) participant contacted but refused to give a blood sample and iii) participant consented to provide the interview but refused to give a blood sample.* Results are presented separately for each gender because descriptive results showed significant differences in contact rates by gender. The different participant response categories and available variables are outlined in Table 1 below:

[Table 1: About here]

As pointed out by Marston et al [3], mobility is an important risk factor for HIV and whenever possible should be factored in the adjustments. Mobility data was available for all individuals as it was derived from the demographic surveillance data base. The mobility index was derived from a count of movement episodes of participants within or without of the surveillance area per unit time. An individual was considered to be highly mobile if they had at least one or more episodes of change of residence per year or at least one outmigration and return episode to the surveillance area in two years.

The missing HIV status (for those in category 2) was imputed against category 1 which had HIV status data, socio-demographic variables and survey data. Missing HIV status data in category 3 and 4 were imputed separately against category 1 using socio-demographic variables and mobility index. All analyses were carried out using Stata version-10 statistical software using a user written program; *ice* [23, 24]. Adjustments for HIV prevalence were carried out separately for each of the three participant categories which had no HIV status

data and by gender. For each category, using the multiple imputation program we created 10 multiple data sets (*5-10 multiple copies are recommended*) with missing data inserted as predicted by the variables in the model. The imputed data files were then combined to get a single analysable data file. From the combined file, prevalence estimates by participant category and the overall estimate were computed. The overall corrected HIV prevalence estimate was taken to be a weighted average of the categories.

Results:

Bivariate Analysis

Overall approximately 70% of the sampled individuals were successfully contacted. Out of 3497 (70% of overall sample) who were contacted, 2721 (54% of overall sample) accepted to provide a blood sample. The percentage of those who were contacted and accepted to be tested constituted 78% among those contacted. Table 2 shows the percentage distribution of those contacted, those who accepted to be tested and HIV prevalence by socio-demographic characteristics. The chi-square statistic and corresponding p-values show that there were statistically significant differences between those contacted and those not contacted by age, gender, residence, ethnicity, mobility and wealth status ($p\text{-value} < 0.05$). Proportionately more women and residents of Korogocho slums were contacted. Individuals in the lowest wealth quintile, the highly mobile and those never married were less likely to be contacted as opposed to those in the wealthier quintiles, the less mobile and those who had ever been married. The percentage distribution of formal education attainment between those contacted and those not contacted was not significantly different.

With regard to accepting to be tested, higher proportions of younger individuals, residents of Korogocho and members of the Luhya and Luo ethnic backgrounds accepted the test than

their counterparts. There were no significant differences between those who accepted the test and those who refused by gender, educational attainment and wealth status. The distribution of HIV prevalence by age, ethnicity, slum of residency, educational attainment, and marital status significantly varied across several variables. Individuals below 20 years of age had the lowest prevalence but one of the highest participation rates, while men had lower participation rate and lower HIV prevalence. On the other hand, residents of Korogocho had higher participation rate and higher HIV prevalence. The Luo and Luhya ethnic groups, the widowed/divorced had higher participation rate and corresponding higher HIV prevalence than their counterparts.

[Table 2: About here]

Multivariate Analysis

Table 3 provides odds ratios derived from a logistic model for sampled individuals being successfully contacted by gender controlling for a set of socio-demographic characteristics.

With regard to age, the results show a general trend for both sexes where the older an individual was the less likely they were to be contacted. There was no significant association between being contacted and educational attainment. Women from the Luhya tribe and men from Kamba tribes were more likely to be contacted compared to their counterparts.

[Table 3: About here]

For both sexes, individuals from the poorest households were about 3 times less likely to be contacted than their richest counterparts. Women from Viwandani were about 28% less likely to be contacted compared to women from Korogocho but the difference was not significant for men. Women and men classified as highly mobile were about 68% and 72% less likely to

be contacted respectively compared to those classified as less mobile. Women and men who had never been married were about 60% less likely to be contacted compared to their married counterparts.

[Table 4: About here]

Table 4 shows the odds ratios for accepting vis-à-vis refusing to provide blood sample once contacted by gender. Regarding age, results show a general trend for both sexes where older individuals were less likely to accept to be tested compared to their younger counterparts although differences were not statistically significant for most categories. Educational level, wealth status and being highly mobile were not significantly associated with accepting or rejecting to be tested. For both sexes, individuals from the Luhya and Luo ethnic backgrounds were more likely to accept to be tested compared to their Kikuyu counterparts. The widowed or divorced men were about 2.3 times more likely to accept testing than the currently married individuals. While for females, the never married were about 1.4 times more likely to accept the HIV test. Residents of Viwandani slum were generally less likely to accept testing compared to residents of Korogocho slum, but there were differences by gender. Women and men of Viwandani were about 59% and 34% less likely to accept to provide a blood sample compared to those of Korogocho respectively.

Table 5 shows observed and adjusted prevalence of HIV for men and women separately and the overall combined estimates. In all interview and test outcome categories, the observed and imputed prevalence for HIV were higher among women. The imputed prevalences for women in all categories were lower than the observed. Women who were not tested had an imputed prevalence lower than the observed prevalence by 8%. The overall corrected

prevalence among women was lower than the observed by 4%. Males who were not tested had an imputed prevalence higher than the observed prevalence by 5%. The overall corrected HIV prevalence for men was higher than the observed by 2%. The overall adjusted prevalence of HIV for both sexes was lower than the observed by about 2%.

[Table 5: About here]

Discussion:

This paper explored non-response to HIV testing in a survey and its impact on HIV prevalence estimates in informal settlements with a relatively mobile and young population. Non-response rate for HIV testing in this study (46%) was quite high relative to other community based HIV testing surveys [1-3]. Absenteeism contributed 65%, while refusals accounted for 35% of non-response. At the time of designing the survey, an estimated non-response rate of 40% based on what has been reported elsewhere and the attrition rates in the NUHDSS, was factored in the sample size estimation.

Bivariate and multivariate assessments of responders and non-responders showed that indeed there were statistically significant differences between the two groups justifying the need to assess the extent to which the observed differences could have affected the overall estimate of HIV prevalence in this population. Age, socio-economic status, residence and mobility index were found to be good predictors of whether an individual was likely to be successfully contacted or not. Older people were not only less likely to be contacted but they were also less likely to accept to be tested. This finding was a bit surprising. One would have expected younger adults to be more mobile and less inclined to spare their time to participate in the survey. It should however be noted that economic survival in the informal settlements relies on a cash economy dominated by informal employment. It might be the case that older people

(up to 49 years for women and up to 54 years for men) have more demanding family responsibilities and as such remain active fending for their families. Similarly residents of Viwandani were less likely to be found at home and if found, they were less inclined to participate. This finding is in line with our expectation. Viwandani slum is predominately inhabited by young adults, with smaller families, more educated and more likely to be working in the nearby industrial estate hence the higher likelihood of not being found at home.

Members of two of the ethnic communities with the highest HIV prevalence in Kenya, the Luo and Luhya, [16] were more likely to be contacted compared to their Kikuyu counterparts and furthermore they were also more likely to accept testing. It is hard to find an explanation for this observation. As we expected, the mobility index predicted the likelihood of being found at home but not necessarily that of accepting or refusing to participate. If all these dynamics were examined in isolation, it would be hard to predict the likely overall impact that the differential participation had on HIV estimates since the odds of participation are in varying directions by socio-demographic characteristics. The odds of participation in the survey were not consistently higher among sub-groups that are characteristically known to have higher or lower HIV prevalence such as age, gender, ethnicity, marital status and socio-economic status. Thus from descriptive results it is difficult to guess which overall direction the results would be biased, if at all, given that both participation rates and observed HIV prevalence varied in both directions along the various categories of the key socio-demographic variables whose sizes (stratum size) also varied.

In the final model of multiple imputations, the overall effect on the estimates was small showing that contrary to our expectation, HIV prevalence appears to have been overestimated

by about 2%. The imputed estimates among females were consistently lower than the observed prevalence while for males the imputed prevalences were higher than those observed except the category of those who accepted to be interviewed but did not give blood. It is important to note that all observed estimates lie within the confidence limits of the adjusted estimates indicating that differences are small in spite of the significant differences in participation rates by socio-demographic characteristics as noted in Tables 2, 3 and 4. The high non-response rate observed in this study notwithstanding, results show that sound estimates can be obtained in a community based HIV sero-prevalence survey in a similar setting. The observed bias in this study is minimal but has a gender angle to it with a tendency to overestimate prevalence among women and underestimate it among males.

Future work in similar settings should take into consideration a number of issues. The informal nature of the housing makes listing of households extremely difficult. In the absence of a dedicated registration and monitoring system such as the demographic surveillance system, having an updated sampling frame is nearly impossible. Ways around this challenge should be carefully considered from the start. Although this study has found that the impact of non-response on the overall estimates was minimal, it is prudent to adequately sample the population, factoring in non-response rates based on attrition rates where available. Extra efforts in reaching the hard to contact individuals must be considered while planning the study, especially in terms of duration of the study and adequacy of field staff.

Limitations:

Although the NUHDSS provided background characteristics for all sampled individuals including non-responders, the set of variables was rather limited for predicting the risk of HIV infection. It is possible that non-responders were significantly and systematically

different from responders on characteristics other than those used in the adjustments.

Mobility, as pointed out in other studies, is a key predictor of HIV infection yet the way the mobility index was measured falls short of capturing short term movements such as absence for days or weeks as happens with long distance truck drivers. Movements involving short durations of absence might actually be more important in exposing individuals to the risk of HIV than movements involving longer periods of absence. Non-return migration can also result into underestimation of HIV prevalence especially if the reason for out-migration is associated with poor health as is the case with terminally ill HIV/AIDS patients. One study noted markedly high HIV/AIDS-related death rates among rural returnees in South Africa [25] , indicating that a significant proportion of rural return migrants were HIV positive. HIV prevalence in the origin population could be affected (lowered) as a result of selective out-migration of infected individuals.

Conclusions:

The estimate of HIV prevalence in slums is higher than that reported for Nairobi province, with women being disproportionately affected. Non-response resulted into minimal overestimates of HIV prevalence overall. We also infer that it is possible to obtain reliable results even in a relatively mobile population under surveillance as long as proper considerations are made at the survey design and implementation stages.

Competing interests:

All authors declare that there are no competing interests

Authors' contribution:

AKZ: took lead in preparing the manuscript, participated in the study design, implementation, data management and analysis and manuscript write-up. NM: was the principal investigator, she participated in the proposal development, study design, project implementation, supervision and manuscript preparation. MM: participated in the proposal development, study design, project implementation, supervision of testing of the blood samples in the laboratory and manuscript preparation. EZ: participated in the proposal development, study design, advised on aspects of implementation, and participated in manuscript preparation. JK: participated in the study design, project implementation, supervision and manuscript preparation. VO: participated in the proposal development, study design, project implementation, supervision, and testing of the blood samples in the laboratory and manuscript preparation. SK: participated in the proposal development, project implementation, study design, supervision and testing of the blood samples in the laboratory and manuscript preparation. AE: participated in the proposal development, study design, project implementation, supervision and overall project management. He also participated in manuscript preparation.

Acknowledgements:

We wish to extend our sincere thanks to Viwandani and Korogocho communities for their continued support and participation in our research projects. Implementation of this project was made possible through the generous financial support from the Rockefeller Foundation under grant 2006AR013. Furthermore we acknowledge the financial support to NUHDSS from The Wellcome Trust grant-GR078530MA, and Hewlett Foundation grant-2006-8376. We appreciate the contribution of our field and data management staff. Their dedication to work in a difficult environment at the height of political tensions was immensely appreciated.

References

1. Mishra V, Vaessen M, Boerma JT, Arnold F, Way A, Barrere B, Cross A, Hong R, Sangha J: **HIV testing in national population-based surveys: experience from the Demographic and Health Surveys.** *Bull World Health Organ* 2006, **84**:537-545.
2. Garcia-Calleja JM, Gouws E, Ghys PD: **National population based HIV prevalence surveys in sub-Saharan Africa: results and implications for HIV and AIDS estimates.** *Sex Transm Infect* 2006, **82 Suppl 3**:iii64-70.
3. Marston M, Harriss K, Slaymaker E: **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* 2008, **84 Suppl 1**:i71-i77.
4. Reniers G, Araya T, Berhane Y, Davey G, Sanders EJ: **Implications of the HIV testing protocol for refusal bias in seroprevalence surveys.** *BMC Public Health* 2009, **9**:163.
5. Reniers G, Eaton J: **Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys.** *Aids* 2009, **23**:621-629.
6. National Statistical Office (NSO) [Malawi], ORC Macro: **Malawi Demographic and Health Survey 2004.** Calverton, Maryland: NSO and ORC Macro; 2005.
7. Zou J, Yamanaka Y, Muze J, Watt M, Ostermann J, Thielman N: **Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes.** *BMC Public Health* 2009, **9**.
8. Simbayi LC, Kalichman S, Strebel A, Cloete A, Henda N, Mqeketo A: **Internalized stigma, discrimination, and depression among men and women living with HIV/AIDS in Cape Town, South Africa.** *Soc Sci Med* 2007, **64**:1823-1831.
9. Akwara P. A, Madise N.J, Hinde A: **Perception of risk of HIV/AIDS and sexual behaviour in Kenya.** *J biosoc Sci* 2003, **35**:385-411.
10. Moutsiakis DL, Chin PN: **Why blacks do not take part in HIV vaccine trials.** *J Natl Med Assoc* 2007, **99**:254-257.
11. Zuma K, Gouws E, Williams B, Lurie M: **Risk factors for HIV infection among women in Carletonville, South Africa: migration, demography and sexually transmitted diseases.** *Int J STD AIDS* 2003, **14**:814-817.
12. Lydie N, Robinson NJ, Ferry B, Akam E, De Loenzien M, Abega S: **Mobility, sexual behavior, and HIV infection in an urban population in Cameroon.** *J Acquir Immune Defic Syndr* 2004, **35**:67-74.

13. Kishamawe C, Vissers DC, Urassa M, Isingo R, Mwaluko G, Borsboom GJ, Voeten HA, Zaba B, Habbema JD, de Vlas SJ: **Mobility and HIV in Tanzanian couples: both mobile persons and their partners show increased risk.** *Aids* 2006, **20**:601-608.
14. Nunn AJ, Wagner HU, Kamali A, Kengeya-Kayondo JF, Mulder DW: **Migration and HIV-1 seroprevalence in a rural Ugandan population.** *Aids* 1995, **9**:503-506.
15. Matrix Development Consultants: **Nairobi's informal settlements: An inventory. A Report Prepared for USAID/REDSO/ ESA.** Nairobi: USAID; 1993.
16. Central Bureau of Statistics (Kenya), Ministry of Health (MOH-Kenya), Macro O: **Kenya Demographic and Health Surveys (2003).** Calverton, Maryland: Central Bureau of Statistics [Kenya], Ministry of Health (MOH) [Kenya], ORC Macro; 2004.
17. National AIDS and STI Control Programme Ministry of Health Kenya: **Kenya AIDS Indicator Survey 2007. Preliminary Report.** Nairobi, Kenya: NASCOP; 2008.
18. Zulu EM, Dodoo FN, Chika-Ezee A: **Sexual risk-taking in the slums of Nairobi, Kenya, 1993-8.** *Popul Stud (Camb)* 2002, **56**:311-323.
19. APHRC: **Population and health dynamics in Nairobi's informal settlements.** Nairobi: African Population and Health Research Center; 2002.
20. Kyobutungi C, Ziraba AK, Ezeh A, Ye Y: **The burden of disease profile of residents of Nairobi's slums: Results from a Demographic Surveillance System.** *Popul Health Metr* 2008, **6**:1.
21. APHRC: **The Economic, Health, and Social Context of HIV Infection in Informal Urban Settlements of Nairobi.** Nairobi: African Population and Health Research Center; 2008.
22. van Buuren S., Boshuizen H. C., Knook D. L.: **Multiple imputation of missing blood pressure covariates in survival analysis.** *Statistics in Medicine* 1999, **18**:681-694.
23. Royston P: **Multiple imputation of missing values: Update of ice.** *The Stata Journal* 2005, **5**:527-536.
24. Royston P: **Multiple imputation of missing values.** *The Stata Journal* 2004, **4**:227-227.
25. Clark SJ, Collinson MA, Kahn K, Drullinger K, Tollman SM: **Returning home to die: circular labour migration and mortality in South Africa.** *Scand J Public Health Suppl* 2007, **69**:35-44.

Additional files provided with this submission:

Additional file 1: Table 1.doc, 29K

<http://www.pophealthmetrics.com/imedia/2133207551368435/supp1.doc>

Additional file 2: Table 2.doc, 98K

<http://www.pophealthmetrics.com/imedia/2155287293684359/supp2.doc>

Additional file 3: Table 3.doc, 73K

<http://www.pophealthmetrics.com/imedia/1270983669368436/supp3.doc>

Additional file 4: Table 4.doc, 73K

<http://www.pophealthmetrics.com/imedia/1686554383684362/supp4.doc>

Additional file 5: Table 5.doc, 46K

<http://www.pophealthmetrics.com/imedia/1034009494368436/supp5.doc>