Sexually Transmitted Diseases, Publish Ahead of Print
DOI: 10.1097/OLQ.0000000000000978

HIV incidence among women at high risk of HIV infection attending a dedicated clinic in Kampala, Uganda: 2008-2017

Ivan Kasamba MSc [1,2], Stephen Nash MSc [3],

Janet Seeley PhD [1,2], Helen A. Weiss PhD [3]

[1] Medical Research Council / Uganda Virus Research Institute and London School of Hygiene & Tropical Medicine(MRC/UVRI and LSHTM) Uganda Research Unit

[2] London School of Hygiene and Tropical Medicine (LSHTM), UK

[3] MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine (LSHTM), UK

Corresponding author:

Ivan Kasamba

MRC/UVRI and LSHTM Uganda Research Unit,

Plot 51-59 Nakiwogo Road, Entebbe Uganda.

Ivan.Kasamba@lshtm.ac.uk

Tel: +256 417 704000, Fax: +256 414 321137

Conflicts of Interest and Source of Funding:

This work was jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement (K012126/1). The clinic receives support from PEPFAR through the Centres for Disease Control and Prevention for the provision of clinical services. IK is jointly funded for a PhD by MRC Tropical Epidemiology Group, London School of Hygiene and Tropical Medicine (LSHTM), UK and UK Medical Research Council (MRC). The authors have no conflicts of interest.

Short Summary

In a cohort of HIV high-risk women, assigning random-point values as the seroconversion timing attenuated variation in incidence observed with the mid-point assignment during observation periods with longer seroconversion intervals.

Abstract

Background: High attrition and irregular testing for HIV in cohort studies for high-risk populations can bias incidence estimates. We compare incidence trends for high-risk women attending a dedicated HIV prevention and treatment clinic, using common methods for assigning when seroconversion occurs and whether seroconversion occurs among those with attrition.

Methods: Between April 2008-May 2009 women were enrolled into cohort-1 and from January 2013 into cohort-2, then scheduled for follow-up once every three months. Incidence trends based on assuming a mid-point in the seroconversion interval were compared to those of assigning a random-point. We also compared estimates based on the random-point with and without multiple imputation (MI) of sero-status for participants with attrition.

Results: By May 2017, 3084 HIV-negative women had been enrolled with 18,364 clinic visits. Before attrition, 27.6% (6,990/25,354) were missed visits. By August 2017, 65.8% (426/647) of those enrolled in cohort-1 and 49.0% (1194/2437) in cohort-2 were defined with attrition. Among women with ≥1 follow-up visit, 93/605 in cohort-1 and 77/1601 in cohort-2 seroconverted. Periods with longer seroconversion intervals appeared to have noticeable differences in incidences when comparing the mid-point and random-point values. MI for attrition is likely to have overestimated incidence following escalated attrition of participants. Based on random-point without MI for attrition, incidence at end of observation was 3.8/100 person-years in cohort-1 and 1.8/100 in cohort-2.

Conclusion: The random-point approach attenuated variation in incidence observed using mid-point. The high incidence after years of ongoing prevention efforts in this vulnerable population should be investigated to further reduce incidence.

Key words: HIV, incidence, high-risk, imputation, Uganda

Introduction

Monitoring HIV incidence rates over time facilitates informed public health decisions on HIV prevention and treatment. Longitudinal studies are commonly used to estimate incidence rates where test dates are scheduled at fixed time intervals of, say, every month or year[1]. The timing of a new infection can be assumed to be within the seroconversion interval of the last-negative and first-positive test dates, as either the mid-point or a random-point value. Recently, these two popular approaches for identifying the timing of seroconversion in estimating incidence rates have been discussed but little is still known about how these approaches influence interpretation of HIV incidence trends under different circumstances in cohorts for women at high risk of HIV[2].

Globally, HIV risk was 13 times higher among female sex workers (FSWs) than among other adult women[3]. HIV incidence in FSW studies in eastern and southern Africa is as high as 9.8/100 person-years[4]. In Uganda, where adult female HIV prevalence is 7.5%, estimated prevalence among FSWs varies between 32%-52%[5,6].

Using data on HIV-1-specific immunoglobulin G (IgG) levels that indicates if a patient was infected recently, assuming the mid-point value while estimating incidence was appropriate for cohort studies with regular testing intervals[7]. However, a recent simulation study estimating incidence trends demonstrated that, in the presence of missed scheduled testing, the mid-point approach leads to an artefactual clustering of seroconversion times in the middle of the observation period[2]. As a result, incidence rates are under-estimated at the beginning and at the

end of the observation period, and over-estimated in the middle. For example, incidence rates at the end of the observation period were in error of 27% for a simulated cohort with a 40-59.9% testing rate. An alternative approach that randomly assigns the timing of seroconversion within the seroconversion interval produced less biased incidence trend estimates[2].

Several factors can lead to varying visit attendance rates during observation. In follow-up studies of long durations, later observation periods usually have poorer attendance rates than earlier ones, partly due to participation fatigue[8]. Changes in certain cohort activities such as removing transport refund can also influence levels of attendance during those periods. In particular, cohorts for high-risk populations are vulnerable to missing visits and prematurely discontinuing participation (attrition) because of the high mobility of participants, and transport costs among other reasons[9,10]. The longer seroconversion intervals due to missed visits create larger uncertainty over when the event occurs, and attrition leads to uncertainty of whether the event occurs for participants defined with attrition. These uncertainties can compromise the interpretations of HIV incidence trends[8]. Participants with attrition are usually censored at their last observation, while assuming their risk of infection at censoring is the same as that in those under observation at that time-point. If this assumption is violated, for instance if women who discontinue are at a higher risk of HIV than those who remain in the study, then there is potential for bias in incidence estimates. Given the cohort experience of participants without attrition who have similar characteristics as those with attrition, using multiple imputation (MI) to predict unobserved HIV sero-status and timing of seroconversion for women with attrition could improve the accuracy of incidence estimates[11].

In this paper, we aim to compare HIV incidence trends based on three approaches: (1) estimating the date of seroconversion within the seroconversion interval using the mid-point; (2) estimating the date of seroconversion within the seroconversion interval using the random-point method; and (3) predicting the time-to-event data using MI for women with attrition. These comparisons are performed in the epidemiological context of a dedicated clinic which has been providing services to women at high risk of HIV in Kampala, Uganda since 2008.

Methods

Cohort description

The Good Health for Women Project (GHWP) is a stand-alone clinic offering free services to women at high risk of HIV-infection in Kampala, Uganda[12]. Briefly, peer educators recruited other women involved in commercial sex or employed in an entertainment facility within mapped sex work hotspots (e.g., clusters of bars, night clubs, lodges and guest houses providing rooms for sex work, street spots frequented by sex workers). Outreach workers visited these women at their workplace to assess eligibility. Women were eligible if aged ≥18 years and reported engaging in commercial sex (self-identified FSWs or received money, goods, or other favours in exchange for sex) or employed in an entertainment facility. Women <18 years were eligible if pregnant, had a child, or provided for their own livelihood. Those eligible were invited to attend the clinic for screening and enrolment.

Women were initially enrolled into cohort-1 between April 2008 and May 2009 and continually scheduled for follow-up once every three months. Previous publications on this cohort have examined prevalence of HIV and other STIs at baseline[12], HIV incidence and risk factors using data up to March 2011[13], and the epidemiology of alcohol use until September 2013[14]. Cohort-2 was started in January 2013 with continuous enrolment of participants and identical follow-up procedures as cohort-1. This paper considered follow-up visits up to 29th August 2017 for both cohorts.

Study intervention

The programme integrates sex worker friendly HIV and sexual and reproductive health services, including offering free general health care for participants and their children aged <5 years old. Figure 1 shows a summarised conceptual framework guiding the GHWP intervention derived from the standard framework for preventing HIV acquisition in FSWs developed by WHO. The intervention can be categorised into three broad components (behavioural, biomedical, structural) aiming at three broad categories of immediate outcomes: reducing unprotected sex; decreasing HIV transmission efficiency; and empowering FSWs[15,16]. The horizontal arrows depict the link from the intervention to the intermediate effects, which in turn determine the impact on HIV acquisition and transmission. The vertical arrows show the factors that can influence the immediate intervention effects, and the transition to HIV prevention indicators. To ensure high cohort retention, group meetings were held every 14 days with women who had scheduled visits within the following month. Participants with two or more missed visits were contacted by phone or visited and encouraged to attend.

Data

At every visit, HIV testing was performed using determine HIV-1/2 (Abbott Diagnostics, UK) for screening and a reactive result was confirmed by Stat-Pak dipstick HIV-1/2 (Chembio Diagnostics, US). The Uni-Gold HIV-1/2 (Trinity Biotech, Ireland) was used as a tie breaker if Determine and Stat-Pak were discordant. At every visit, trained nurse-counsellors administered structured questionnaires to collect data on socio-demographic characteristics, sexual behaviour, reproductive health, alcohol use, and illicit drug use.

Statistical analysis

Analyses were restricted to HIV-negative women enrolled before May 2017 to allow for follow-up visits for all participants. In estimating incidence, the endpoint for calculating person-years at risk was the earliest of i) estimated HIV seroconversion date for seroconverters or ii) last HIV-negative test date for non-seroconverters.

HIV seroconversion date was estimated using (a) mid-point between last-negative and first-positive HIV test dates; (b) random-point method. The random-point method assumed seroconversion to be a random date from a uniform distribution bounded by the last-negative and first-positive test dates. To account for the variability of using a random date, multiple imputations were made, and incidence rates over time were estimated each time and averaged at each time-point for all iterations.

To assess the potential consequence of attrition on incidence trends, we imputed the unobserved HIV sero-status then obtained timing for seroconversion for participants with attrition. Several methods for imputing unobserved time-to-event endpoints have been published[17,18]. Since our interest is incidence trends evaluated at time intervals of follow-up, we applied an MI approach that models the probability of seroconversion within each time interval given that the participant was HIV-negative up to the start of that interval. In this approach the follow-up period is divided into discrete intervals of time and logistic regression used to impute sero-status sequentially in each interval. Then, the imputed sero-statuses are mapped back to the time-scale by allocating a random time within the seroconversion interval, and censoring at each participant's expected complete follow-up time for imputed non-seroconverters. Estimates were based on 200 imputations. The log of incidence rates at each time-interval were estimated separately in each imputed dataset, then averaged using Rubin's rules[19]. Statistical tests assessing whether the age-adjusted incidence trends changed over time in each imputed dataset were combined after normalisation using Wilson-Hilferty transformation[20].

The imputation model included factors associated with both attrition and HIV incidence in this analysis. To improve imputation accuracy of sero-status, we included as a-priori the age, and time-varying characteristics of: frequency of alcohol consumption, reporting paid sex, condom use frequency with paid sex and number of partners in the last month in the imputation models. To identify factors associated with attrition, the outcome was time-to-attrition, and follow-up was right-censored at last observation. A participant had attrition if their last study attendance was 12 months or more prior to the administrative censoring date, consistent with what has been used in other studies. To investigate factors associated with attrition and then incidence, hazard

ratios (HRs), and associated 95% confidence intervals (CI) and p-values were estimated using Cox regression models. Age-adjusted models were fitted separately for each factor and independently associated sociodemographic factors (p-value≤0.15) were determined. These were then adjusted for behavioural and reproductive health factors that remained statistically significant at 15% level. Time-changing variables were fitted using the most recent reports at each visit.

Analyses were performed separately for each cohort because of the potential variation in exposures due to the five-year gap between the starts of the two cohorts. For each cohort, estimates of incidence trends from three models were compared: 1) Without MI using mid-point; (2) Without MI using random-point; (3) MI using random-point. Changes in incidence trends were assessed using Poisson regressions models by testing the inclusion of a linear and a quadratic trend with time in models adjusted for current age. Stata 14 (StataCorp, Tx) was used for all analyses.

Results

HIV prevalence at enrolment was 37.0% (380/1027) in cohort-1 and 35.4% (1335/3772) in cohort-2. Table 1 shows enrolment characteristics for the 3084 HIV-negative women. Overall, the mean age was 26.2 years (SD: 6.3) with over 80% having ≥1 child. Most (58.0%) reported their only source of income as sex work, and 37.2% reported having another job in addition to sex work.

Attrition and missed visits

The percentage of women with ≥1 follow-up visit was 93.5% (605/647) in cohort-1 and 65.7% (1601/2437) in cohort-2. A total of 32,762 visits (Cohort-1:17,970; Cohort-2:14,792) were expected while 18,364 (Cohort-1: 9,504; Cohort-2: 8,860) were made. Before attrition, 6,990 visits (Cohort-1:3,324; Cohort-2:3,666) were missed and 7,408 (Cohort-1:5142; Cohort-2:2266) after attrition. Table 1 SDC link http://links.lww.com/OLQ/A343 shows study participation and the supplementary Table 2 SDC link http://links.lww.com/OLQ/A344 shows the factors associated with attrition. Attrition was independently associated with follow-up characteristics of having no child, being pregnant, recruiting clients on phone, and reporting no illicit drug use in the last 3 months in cohort-1 and younger age, not using oral contraceptives, and less frequent paid sex in the last year in cohort-2. The percentage of participants attending each visit with ≥1 missed visits since the previous attendance in cohort-1 was stable at 10% until year four, when this proportion increased to an average of 30% for all the visits after year four (Figure 2). In cohort-2, this proportion was stable at 30% for all the visits following enrolment.

HIV incidence

Among 2206 HIV-negative participants with ≥1 follow-up visit, 170 were observed to seroconvert within 5540 person-years. Without MI, the mid-point and random-point incidence trends were similar in the first four years of enrolment and at years 7 and 8 of cohort-1 (Figure 3). At the introduction of cohort-2 in January 2013, corresponding to years 5 and 6, the random-point method attenuated the variation observed using mid-point estimation (Figure 3a). The

median length of the seroconversion interval was 3.4 months (IQR: 3.0-7.1) for cohort-1 participants whose last-negative test date occurred before January 2013, and 10.1 months (IQR: 5.8-18.3) between January 2013 and December 2014, and 3.9 months (IQR: 2.5-5.5) after January 2015. In cohort-2 the median was 14.9 months (IQR: 4.7-26.9) between January 2013 and December 2014, and 5.6 months (IQR: 3.2-7.7) after January 2015. In cohort-2 without MI, incidence based on mid-point increased between the first and second intervals peaked at year two then decreased steeply at the end of the observation period, while that based on random-point dropped from its peak at the first interval and decreased further in the last interval (Figure 3b).

In cohort-1, incidences based on MI using random-point estimation for attrition were higher than without MI (using either method), particularly starting from year five. In cohort-2, the overlapping 95% CIs for incidences are consistent with no differences between MI and without MI approaches.

Based on random-point without MI for attrition, there was a sharp fall in incidence from 6.1/100 person-years within the first six months to 2.5/100 in the following six months, and then 2.0/100 at year three in cohort-1 (age-adjusted trend p-value<0.001). After the fourth year, incidence increased to 3.8/100 person-years at the end of the observation period but there was weak evidence to suggest a changing trend (age-adjusted p-value=0.15). In cohort-2, incidence declined from 3.8/100 person-years in the first 6 months to 1.8/100 in year 3 (age-adjusted p-value=0.04).

Discussion

Our findings suggest that the variations in incidence patterns observed with assigning mid-points within seroconversion intervals were attenuated if random-points were assigned instead. With long seroconversion intervals spanning the observation period, incidences based on mid-point were lower at the beginning and at the end of the observation than in the middle where incidence peaked (cohort-2). Where long seroconversion intervals were concentrated in the middle of observation periods with short seroconversion intervals, there were larger fluctuations in incidences over time during that middle period using the mid-point than the random-point approach (cohort-1). MI for unobserved HIV sero-status and time-to-seroconversion for women defined with attrition showed noticeably higher incidences than without MI following an escalated attrition of participants (year 5 in cohort-1 and year 1 in cohort-2).

An incidence peak in the middle of the observation period and lower incidence at the beginning and at the end using the mid-point approach is consistent with results from a recent simulation study where incidence was underestimated at the beginning and at the end of the observation and overestimated in the middle[2]. This mid-point behaviour is a result of clustering of the seroconversion timing in the middle of the observation period. The authors observed this pattern with a large extent of missed scheduled test dates irrespective of the true incidence trend, cohort type and number of scheduled test dates. However, they described incidence patterns where irregular HIV testing spans the entire observation period, similar to what we observe in cohort-2. Conversely, in their study the random-point assignment of seroconversion date which was less restrictive than the mid-point showed less biased incidence trends.

Cohort-1 had much longer seroconversion intervals in the period spanning year 5 and 6 than during the first four years and years 7 and 8. This resulted in similar incidences for the random-point and mid-point approaches in the period with shorter seroconversion intervals and noticeable differences in the period with longer seroconversion intervals. In 2013 and 2014, peer-educators who were instrumental in study recruitment involving high-risk populations were not paid allowances. This, together with not providing transport compensation for participants, had a negative effect on visit attendance leading to longer seroconversion intervals and a subsequent clustering of seroconversion timing at year 5 but offset at the following year.

Although participants in years 7 and 8 had irregular visit attendance similar to that in years 5 and 6, they had shorter seroconversion intervals suggesting that the length of the seroconversion intervals rather than irregular visit attendance determine the extent of differences between the two approaches. If for some reason there is irregular visit attendance in general but the seroconversion intervals are shorter, then there would be little differences between the approaches.

In both cohorts, the higher MI-based than non-MI incidences following escalated attrition of participants suggests attrition of more-risky participants, given previously reported characteristics. However, it is likely that the estimated MI-based incidences following marked attrition may be overestimated if some participants dropped-out because they perceived themselves to be low-risk, for example, exited sex work but reported high-risk behaviours at the prior visit. For cohort-1, escalated attrition occurred at a time-point corresponding to the estimated mean duration of risk behaviours for FSWs in Africa of 5.5 years, as reported in a

review of duration of risk behaviours in key populations[21]. Overall, these MI-based estimates suggest that incidences obtained without MI may have been underestimated following escalated attrition.

A key limitation in this study is the use of self-reported data in predicting sero-status for women with attrition, which data is subject to social desirability and can ultimately affect MI-based incidence estimates. Also, some women defined with attrition in our study may be accessing services elsewhere including private services, may have exited sex work, or are living less-risky lifestyles yet their un-observed predicted sero-status is based on previously reported high-risk behaviours. Therefore, combined with the lack of proper data on the reasons for exit and data on mortality, the MI-based incidences may have been over-estimated. In addition, missing data in some women, which was used to perform MI, could potentially skew conclusions. However, we assumed these data were missing completely at random because most of the missingness was due to the electronic data capture system failing to capture data at roll-out during cohort-2. Nonetheless, it is unlikely that these observations could have substantially changed our conclusions.

In conclusion, the random-point approach reduced the improbable fluctuations in incidence and incidence patterns observed with mid-point in the presence of long seroconversion intervals. Based on random-point without MI, the HIV incidence declines in the three years following enrolment of 67% in cohort-1 and 53% in cohort-2 were substantial but incidence remains high. The declines in HIV incidence over time are likely due to the ongoing interventions but the high incidence rates after years of potential exposure to the intervention suggests a need for further

strengthening the intervention efforts in this cohort. In particular, the factors influencing the slight rise in incidence after four years among cohort-1 participants are not well understood, but these changes coincide with the period with substantial proportions of missed visits. We are currently examining some of these factors in more detail. Despite the importance of tracking HIV incidence trends in this vulnerable population, there was paucity of recent incidence data with which to compare our findings[10,21,22]. In choosing how to assign the seroconversion timing and interpreting the resulting incidence patterns, it is important to consider the changes in the extent of missed visits and length of seroconversion intervals over time.

Ethical issues

This study was approved by the Research Ethics Committee of the Ugandan Virus Research Institute, Uganda National Committee for Science and Technology, and Ethics Committee of the London School of Hygiene and Tropical Medicine. Each participant gave written informed consent and confidentiality was kept by using numerical identifiers only.

Acknowledgements

We thank the cohort participants and the many staff who have contributed to the success of the Good Health for Women Project over the last 10 years. Particularly, Yunia Mayanja for her comments to the final draft, Daniel Bagiire and Onesmus Kamacooko for their endless response to the many data requests. We acknowledge Jim Todd, Kathy Baisley and Maryam Shahmanesh for the initial input regarding the overall direction of the work of which this paper is part.

References

1. Van Beneden C, Olsen S, Skoff T, Lynfield R. Active, population-based surveillance for infectious diseases. *Infectious Disease Surveillance Blackwell.* 2008:32-43.
2. Vandormael A, Dobra A, Bärnighausen T, de Oliveira T, Tanser F. Incidence rate estimation, periodic testing and the limitations of the mid-point imputation approach. *International Journal of Epidemiology.* 2017.
3. HIV/AIDS JUNPo. *Global AIDS Update 2018.* 2018.
4. Hargreaves JR, Mtetwa S, Davey C, et al. Cohort analysis of programme data to estimate HIV incidence and uptake of HIV-related services among female sex workers in Zimbabwe, 2009-14. *Journal of acquired immune deficiency syndromes (1999).* 2015.
5. Muldoon KA. A systematic review of the clinical and social epidemiological research among sex workers in Uganda. *BMC public health.* 2015;15(1):1.
6. 2016 PPU. Uganda Population-Based HIV Impact Assessment: 2016–2017. Columbia University; 2016.
7. Skar H, Albert J, Leitner T. Towards estimation of HIV-1 date of infection: a time-continuous IgG-model shows that seroconversion does not occur at the midpoint between negative and positive tests. *PLoS One.* 2013;8(4):e60906.
8. Fewtrell MS, Kennedy K, Singhal A, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Archives of disease in childhood.* 2008;93(6):458-461.
9. Graham SM, Raboud J, McClelland RS, et al. Loss to follow-up as a competing risk in an observational study of HIV-1 incidence. *PloS one.* 2013;8(3):e59480.

10. Van Loggerenberg F, Mlisana K, Williamson C, et al. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PloS one.* 2008;3(4):e1954.

11. Young R, Johnson DR. Handling missing values in longitudinal panel data with multiple imputation. *Journal of Marriage and Family.* 2015;77(1):277-294.

12. Vandepitte J, Bukenya J, Weiss HA, et al. HIV and other sexually transmitted infections in a cohort of women involved in high-risk sexual behavior in Kampala, Uganda. *Sexually transmitted diseases.* 2011;38(4):316-323.

13. Vandepitte J, Weiss HA, Bukenya J, et al. Alcohol use, Mycoplasma genitalium and other STIs associated with HIV incidence among women at high risk in Kampala, Uganda. *Journal of acquired immune deficiency syndromes (1999).* 2013;62(1):119.

14. Weiss HA, Vandepitte J, Bukenya JN, et al. High levels of persistent problem drinking in women at high risk for HIV in Kampala, Uganda: A prospective cohort study. *International journal of environmental research and public health.* 2016;13(2):153.

15. WHO. Toolkit for targeted HIV/AIDS prevention and care in sex work settings. Geneva, WHO, 2005. Available at: http://whqlibdoc.who.int/publications/2005/9241592966.pdf (accessed on 13 September 2018).

16. Bekker L-G, Johnson L, Cowan F, et al. Combination HIV prevention for female sex workers: what is the evidence? *The Lancet.* 2015;385(9962):72-87.

17. Lipkovich I, Ratitch B, O'Kelly M. Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. *Pharmaceutical statistics.* 2016;15(3):216-229.

18. Carpenter J, Kenward M. *Multiple imputation and its application.* John Wiley & Sons; 2012.

19. Rubin DB. *Multiple imputation for nonresponse in surveys.* Vol 81: John Wiley & Sons; 2004.

20. Wilson EB, Hilferty MM. The distribution of chi-square. *Proceedings of the National Academy of Sciences.* 1931;17(12):684-688.

21. Fazito E, Cuchi P, Mahy M, Brown T. Analysis of duration of risk behaviour for key populations: a literature review. *Sex Transm Infect.* 2012;88(Suppl 2):i24-i32.

22. Hoffmann O, Zaba B, Wolff B, et al. Methodological lessons from a cohort study of high risk women in Tanzania. *Sexually Transmitted Infections.* 2004;80(suppl 2):ii69-ii73.

Figure 2: Distribution for the number of missed visits between consecutive visits attended following enrolment by cohort of enrolment.

Figure 3. Compares HIV incidence trends based on three methods, by cohort: 1) Without multiple imputation (Ml) for attrition using the mid-point method; (2) Without MI using random-point; (3) With MI for attrition using random-point. 95% Confidence intervals are shown only for the preferred method of random-point.

Notes: MI was based on 30 imputed datasets and combined rates using Rubin's rules. The imputation model for MI included age, number of partners in the last month, reporting paid sex, frequency of condom use in the last month with paid sex, frequency of alcohol consumption, source of income, number of children and marital status. Random-point estimation of the seroconversion date was based on 200 imputations, more than 3 times imputations required for convergence. 95% confidence intervals are shown for only random-point estimation of HIV seroconversion with and without MI. Missing data in some of the variables in the imputation model were mainly because some data was not captured as electronic data capture was being rolled-out in January 2015. Therefore, we assumed that this data was missing completely at random (MCAR), and multiple imputation that involved these variables was not conducted for these women.

Figure 1

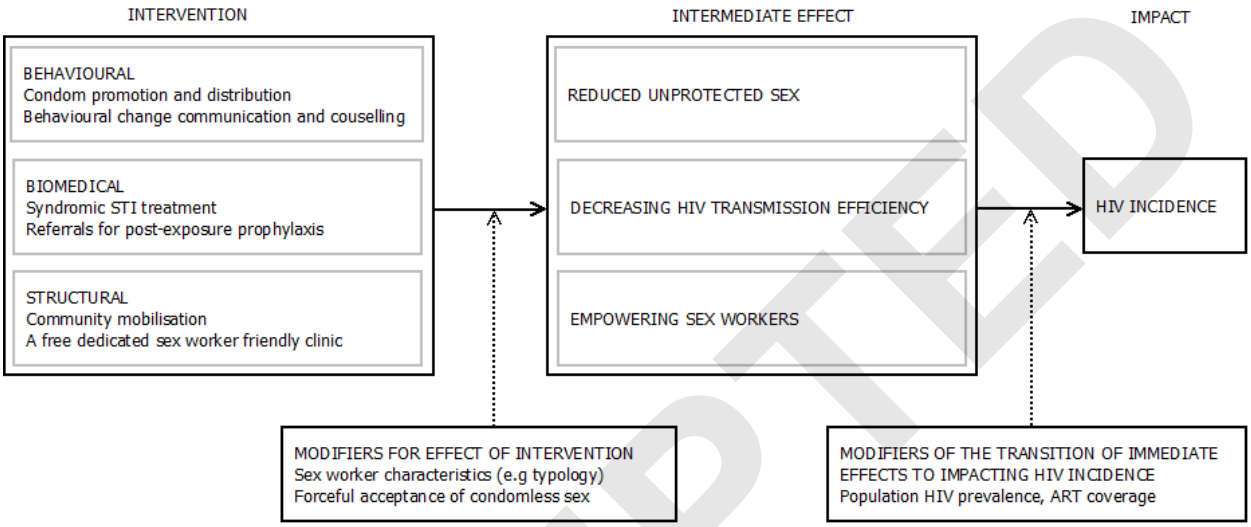


Figure 2

Figure 2: Distribution for the number of missed visits between consecutive visits attended following enrolment by cohort of enrolment.

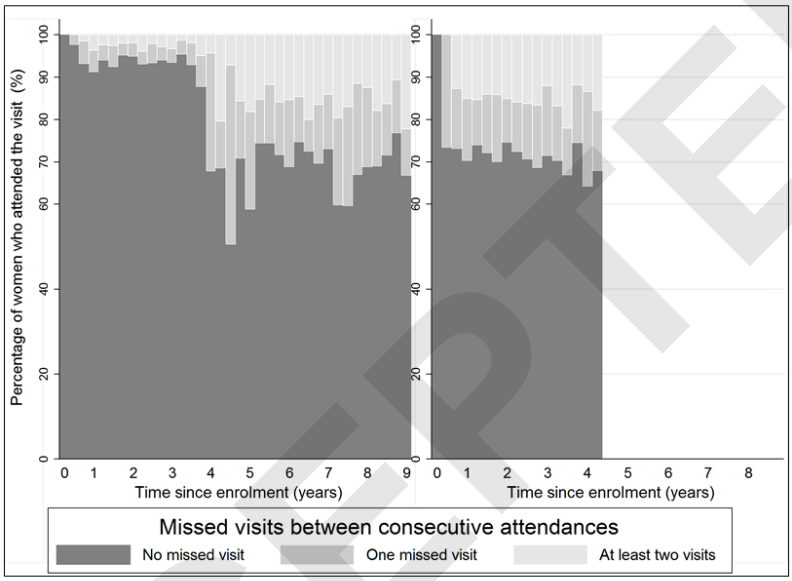
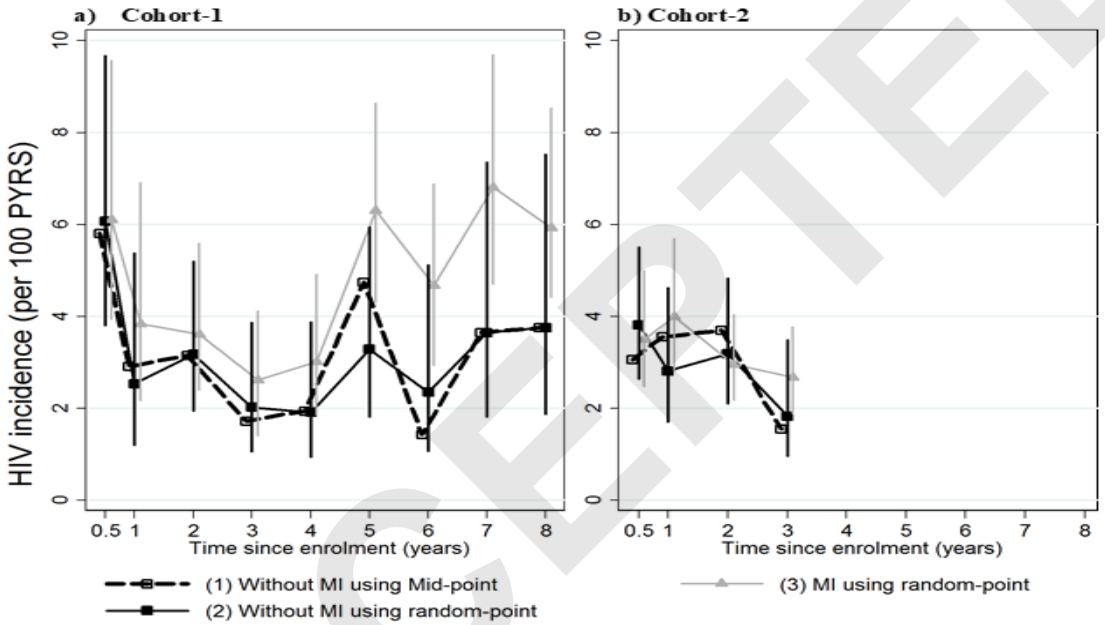


Figure 3

Figure 3. Compares HIV incidence trends based on three methods, by cohort: 1) Without multiple imputation (MI) for attrition using the mid-point method; (2) Without MI using random-point; (3) With MI for attrition using random-point. 95% Confidence intervals are shown only for the preferred method of random-point.



Notes: MI was based on 30 imputed datasets and combined rates using Rubin's rules. The imputation model for MI included age, number of partners in the last month, reporting paid sex, frequency of condom use in the last month with paid sex, frequency of alcohol consumption, source of income, number of children and marital status. Random-point estimation of the seroconversion date was based on 200 imputations, more than 3 times imputations required for convergence. 95% confidence intervals are shown for only random-point estimation of HIV seroconversion with and without MI. Missing data in some of the variables in the imputation model were mainly because some data was not captured as electronic data capture was being rolled-out in January 2015. Therefore, we assumed that this data was missing completely at random (MCAR), and multiple imputation that involved these variables was not conducted for these women.

Table 1: Characteristics at enrolment for HIV-negative women enrolled into the GHWP between 31st March 2008 and 1st May 2017, and a comparison of women who remained in each cohort with those defined with attrition.

Characteristic at enrolment			Cohort-1			Cohort-2	
		All	Successfully followed, n (%)	Loss to follow-up, n (%)	All[1]	Successfully followed, n (%)	Loss to follow-up, n (%)
Total		**647**	**221**	**426**	**2437**	**1243**	**1194**
Age in years	<24	291 (45.0)	92 (41.6)	199 (46.7)	1059 (44.6)	528 (43.5)	531 (45.6)
	25-34	298 (46.1)	107 (48.4)	191 (44.8)	1008 (42.4)	511 (42.1)	497 (42.7)
	35+	58 (9.0)	22 (10.0)	36 (8.5)	310 (13.0)	174 (14.3)	136 (11.7)
					60		
Highest education level	Less than primary	284 (43.9)	99 (44.8)	185 (43.4)	907 (40.8)	458 (41.8)	449 (39.8)
	Completed Primary	281 (43.4)	96 (43.4)	185 (43.4)	940 (42.2)	443 (40.4)	497 (44.1)
	Completed Secondary Ordinary level	82 (12.7)	26 (11.8)	56 (13.1)	378 (17.0)	196 (17.9)	182 (16.1)
					212		
Marital status	Widowed/divorced	417 (64.5)	147 (66.5)	270 (63.4)	1447 (60.7)	723 (59.3)	724 (62.3)
	Currently married	52 (8.0)	19 (8.6)	33 (7.7)	163 (6.8)	86 (7.1)	77 (6.6)
	Never married	178 (27.5)	55 (24.9)	123 (28.9)	772 (32.4)	410 (33.6)	362 (31.1)
					55		
Number of children	None	85 (13.1)	16 (7.2)	69 (16.2)	399 (17.0)	211 (17.6)	188 (16.3)
	One	169 (26.1)	63 (28.5)	106 (24.9)	612 (26.1)	296 (24.7)	316 (27.5)
	At least 2	393 (60.7)	142 (64.3)	251 (58.9)	1337 (56.9)	690 (57.6)	647 (56.2)
					89		
Source of income	Sex work alone	379 (58.6)	133 (60.2)	246 (57.7)	1371 (57.9)	732 (60.6)	639 (55.1)
	Sex work and other job	242 (37.4)	80 (36.2)	162 (38.0)	876 (37.0)	430 (35.6)	446 (38.5)
	No sex work	26 (4.0)	8 (3.6)	18 (4.2)	119 (5.0)	45 (3.7)	74 (6.4)
					71		
Where paying clients are recruited from	Bar, club or restaurant	334 (52.7)	117 (54.2)	217 (51.9)	1062 (48.1)	564 (49.0)	498 (47.1)
	Street	82 (12.9)	32 (14.8)	50 (12.0)	693 (31.4)	381 (33.1)	312 (29.5)
	Several avenues	218 (34.4)	67 (31.0)	151 (36.1)	452 (20.5)	205 (17.8)	247 (23.4)

[1] Most missing data were due to programming errors in the electronic data capture system as the system was rolled-out starting in January 2015.

	missing	13	5	8	230		
Pregnancy status	Not pregnant	585 (90.4)	197 (89.1)	388 (91.1)	2273 (96.8)	1161 (98.1)	1112 (95.5)
	Pregnant	62 (9.6)	24 (10.9)	38 (8.9)	76 (3.2)	23 (1.9)	53 (4.5)
					88		
Current contraceptive use	none or other	364 (56.3)	124 (56.1)	240 (56.3)	1449 (65.7)	724 (66.1)	725 (65.4)
	oral cc	63 (9.7)	20 (9.0)	43 (10.1)	142 (6.4)	79 (7.2)	63 (5.7)
	inject	158 (24.4)	53 (24.0)	105 (24.6)	538 (24.4)	270 (24.6)	268 (24.2)
	Pregnant	62 (9.6)	24 (10.9)	38 (8.9)	76 (3.4)	23 (2.1)	53 (4.8)
					232		
Frequency of paid sex in the last 12 months at enrolment	Less than once a week/None	92 (14.2)	29 (13.1)	63 (14.8)	308 (12.8)	127 (10.4)	181 (15.4)
	At least once a week	256 (39.6)	89 (40.3)	167 (39.2)	689 (28.7)	353 (28.9)	336 (28.5)
	Daily	299 (46.2)	103 (46.6)	196 (46.0)	1404 (58.5)	743 (60.8)	661 (56.1)
					36		
Number of sexual partners in the last month	<5	208 (32.1)	67 (30.3)	141 (33.1)	545 (24.6)	241 (22.0)	304 (27.0)
	5-19	200 (30.9)	74 (33.5)	126 (29.6)	471 (21.2)	246 (22.5)	225 (20.0)
	At least 20	239 (36.9)	80 (36.2)	159 (37.3)	1203 (54.2)	606 (55.4)	597 (53.0)
					218		
Number of paying partners in the last month	<5	227 (35.1)	80 (36.2)	147 (34.5)	562 (25.4)	249 (22.8)	313 (27.9)
	5-19	193 (29.8)	64 (29.0)	129 (30.3)	457 (20.7)	237 (21.7)	220 (19.6)
	At least 20 or cannot remember	227 (35.1)	77 (34.8)	150 (35.2)	1193 (53.9)	604 (55.4)	589 (52.5)
					225		
Condom use frequency with paid sex in the last month	Inconsistent	214 (33.1)	75 (33.9)	139 (32.6)	969 (43.4)	496 (45.0)	473 (41.9)
	Consistent (always)	359 (55.5)	127 (57.5)	232 (54.5)	1007 (45.1)	487 (44.2)	520 (46.1)
	No paid sex	74 (11.4)	19 (8.6)	55 (12.9)	255 (11.4)	120 (10.9)	135 (12.0)
					206		
Alcohol consumption frequency	Non-drinker	160 (24.7)	51 (23.1)	109 (25.6)	531 (23.0)	280 (23.9)	251 (22.0)
	Non-daily drinker	332 (51.3)	115 (52.0)	217 (50.9)	886 (38.4)	427 (36.5)	459 (40.3)
	Daily drinker	155 (24.0)	55 (24.9)	100 (23.5)	892 (38.6)	463 (39.6)	429 (37.7)
					128		
Binge drinking in last 3 months	Non drinker	160 (24.7)	51 (23.1)	109 (25.6)	530 (22.6)	280 (23.6)	250 (21.6)
	No binging	325 (50.2)	114 (51.6)	211 (49.5)	396 (16.9)	179 (15.1)	217 (18.8)
	Binged	162 (25.0)	56 (25.3)	106 (24.9)	1414 (60.4)	725 (61.2)	689 (59.6)
					97		

Illicit drug use in last 3 months	No	529 (81.8)	170 (76.9)	359 (84.3)	1548 (69.6)	782 (71.3)	766 (68.0)
	Yes	118 (18.2)	51 (23.1)	67 (15.7)	676 (30.4)	315 (28.7)	361 (32.0)
					213		
When the participant last tested prior to enrolment	>1 year ago or never	408 (63.1)	138 (62.4)	270 (63.4)	582 (24.8)	268 (22.3)	314 (27.4)
	7-12 months ago	110 (17.0)	36 (16.3)	74 (17.4)	294 (12.5)	151 (12.6)	143 (12.5)
	<6 months ago	129 (19.9)	47 (21.3)	82 (19.2)	1470 (62.7)	783 (65.1)	687 (60.1)
					91		
Ever experienced gender based violence by type of partner	Never				1308 (59.3)	606 (55.7)	702 (62.8)
	Marital or non-paying partners				458 (20.8)	240 (22.1)	218 (19.5)
	Paying clients				397 (18.0)	218 (20.0)	179 (16.0)
	Several of above/DK				43 (1.9)	24 (2.2)	19 (1.7)
	missing				231		

Table 2: Study participation and follow-up status at each follow-up time interval by enrolment cohort

Time interval (years)	Cohort-1 (closed cohort)							Cohort-2 (open cohort)						
	Retention		Status at the end follow-up					Retention		Status at the end follow-up				
	Women expected at the start of each interval	Cohort retention at end of interval†, n (%)	Women without a single follow-up visit	HIV-negative followed-up to end of interval	Attrition by end of interval ‡	Administratively censored	HIV incident cases (Sero-converters)	Women expected at the start of each interval*	Cohort retention at end of interval†, n (%)	Women without a single follow-up visit	HIV-negative followed-up to end of interval	Attrition by end of interval ‡	Administratively censored	HIV incident cases (Sero-converters)
Enrolled	647	647 (100.0)	42	-	-	-	-	2,437	2437 (100.0)	836	-	-	-	-
0 - 0.5	647	582 (90.0)	-	570	23	-	12	2,338	1692 (72.4)	-	1,332	134	120	15
0.5 - 1	647	563 (87.0)	-	540	19	-	11	2,065	1197 (58.0)	-	997	150	171	14
1 - 2	647	524 (81.0)		485	39	-	16	1,731	650 (45.4)		488	211	272	26
2 - 3	647	491 (75.9)	-	444	32	-	8	1,026	450 (43.9)	-	254	77	142	15
3 - 4	647	429 (66.3)	-	375	63	-	7	676	300 (44.4)	-	41	11	236	7
4 - 5 ‡	647	375 (58.0)	-	313	54	-	8							
5 - 6	647	331 (51.2)	-	264	44	-	4							
6 - 7	647	267 (41.3)	-	189	64	-	12							
7 - 8	647	203 (31.4)	-	116	46	18	9							
End of follow-up			-	-		110	6							

‡For cohort-2, the last interval of follow-up ends at time interval of 3-4 years, †cohort retention defined as the number of women followed-up to each time interval including those who had seroconverted among those who should have been followed by that time point, ‡ A participant is classified with attrition (drop-out) if their last attended visit was 12 months or more prior to 29th August, 2017. Participants were administratively censored if their last attended visit was within the last 12 months from the administrative date of 29th August 2017. *Participants expected at the start of each interval for cohort-2 vary because of continuous enrolment, meaning that to be included in an interval the participant should have been enrolled for that duration.

Table 3: HIV incidence rates and associated 95% confidence intervals (CI) for each of the three analytic methods, by time since enrolment for each cohort.

	Cohort-1			Cohort-2		
Years since enrolment	Mid-point without multiple imputation [per 100 pyr (95% CI)]	Random-point without multiple imputation [per 100 pyr (95% CI)]	Random-point with multiple imputation [per 100 pyr (95% CI)]	Mid-point without multiple imputation [per 100 pyr (95% CI)]	Random-point without multiple imputation [per 100 pyr (95% CI)]	Random-point with multiple imputation [per 100 pyr (95% CI)]
Combined	3.1 (2.5-3.8)	3.1 (2.5-3.8)	4.8 (4.2-5.5)	3.0 (2.4-3.8)	3.0 (2.4-3.8)	3.2 (2.7-3.8)
0.5	5.8 (3.6 - 9.3)	6.1 (3.8 - 9.7)	6.1 (3.9 - 9.6)	3.1 (2.0 - 4.6)	3.8 (2.6 - 5.5)	3.5 (2.5 - 5.0)
1	2.9 (1.5 - 5.8)	2.5 (1.2 - 5.4)	3.8 (2.2 - 6.9)	3.5 (2.3 - 5.5)	2.8 (1.7 - 4.6)	4.0 (2.8 - 5.7)
2	3.1 (1.9 - 5.1)	3.2 (2.0 - 5.2)	3.6 (2.4 - 5.6)	3.7 (2.5 - 5.4)	3.2 (2.1 - 4.8)	3.0 (2.2 - 4.0)
3	1.7 (0.9 - 3.4)	2.0 (1.1 - 3.9)	2.6 (1.4 - 4.1)	1.5 (0.8 - 3.1)	1.8 (1.0 - 3.5)	2.7 (1.9 - 3.8)
4	1.9 (1 - 3.9)	1.9 (0.9 - 3.9)	3.0 (1.9 - 4.9)			
5	4.7 (2.9 - 7.7)	3.3 (1.8 - 5.9)	6.3 (4.3 - 8.6)			
6	1.4 (0.5 - 3.8)	2.3 (1.1 - 5.1)	4.7 (2.9 - 6.9)			
7	3.7 (1.8 - 7.3)	3.6 (1.8 - 7.3)	6.8 (4.7 - 9.7)			
8	3.8 (1.9 - 7.5)	3.8 (1.9 - 7.5)	5.9 (4.4 - 8.5)			

Pyr – person-years of follow-up, 95% CI – 95% confidence intervals