# Supplementary Materials for

## High Coverage of ART Associated with Decline in Risk of HIV Acquisition in Rural KwaZulu-Natal, South Africa

Frank Tanser,* Till Bärnighausen, Erofili Grapsa, Jaffer Zaidi, Marie-Louise Newell

*To whom correspondence should be addressed. E-mail: tanserf@africacentre.ac.za

Published 22 February 2013, *Science* **339**, 966 (2013)
DOI: 10.1126/science.1228160

This PDF file includes:

Materials and Methods
Fig. S1 to S7
Tables S1 to S12
References

## Materials and Methods:

*Study Area*

Within the Hlabisa sub-district (part of Umkhanyakude District) is the Africa Centre demographic surveillance area (Figure S1,2) – comprising 60,000 resident individuals (*1*). The population is almost exclusively Zulu-speaking and the area is typical of many rural parts of South Africa in that while it is predominantly rural, it also contains an urban township and informal peri-urban settlements. Umkhanyakude is the poorest of the 11 districts in KwaZulu-Natal, and the second most deprived district in all of South Africa (*2*). The adult unemployment rate was 67% in the surveillance population in 2010 and Umkhanyakude had the lowest score of any district in the province on the United Nation's Human Development Index in 2007 (*3*) (HDI) – an HDI comparable to the country-level HDI for Swaziland, Malawi and Tanzania. The setting is characterised by high levels of circulatory migration (*4*).

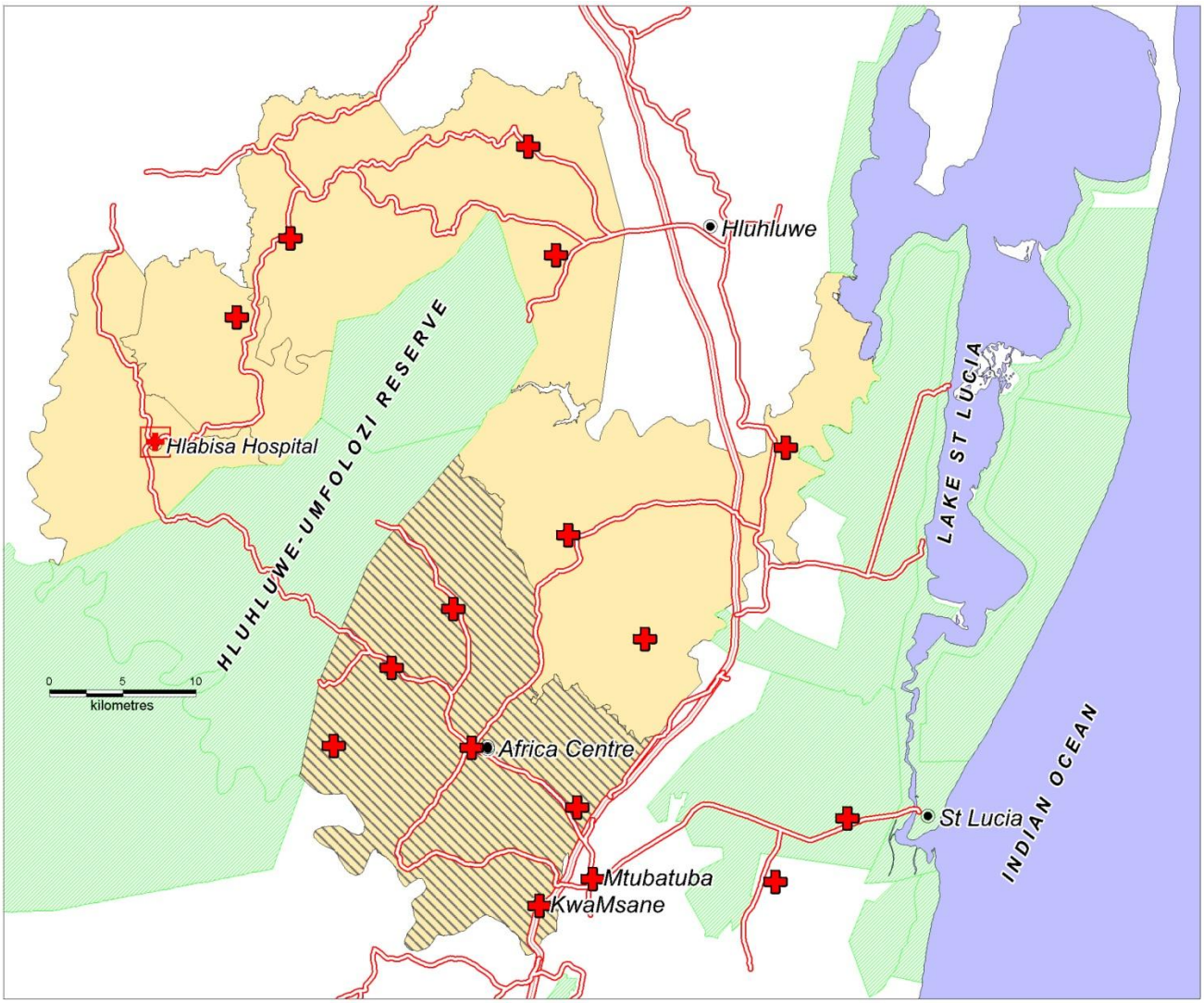


**Figure S1:** Map of the Hlabisa sub-district showing community-based clinics (red crosses) where ART is delivered by trained nurses and counselors. The striped area indicates the Africa Centre's demographic population-based surveillance area with 90,000 individuals under continual demographic surveillance – 60,000 of whom are resident in the study area at any given time. The National Road can be seen next to KwaMsane Township continuing along the eastern boundary of the surveillance area towards Mozambique.

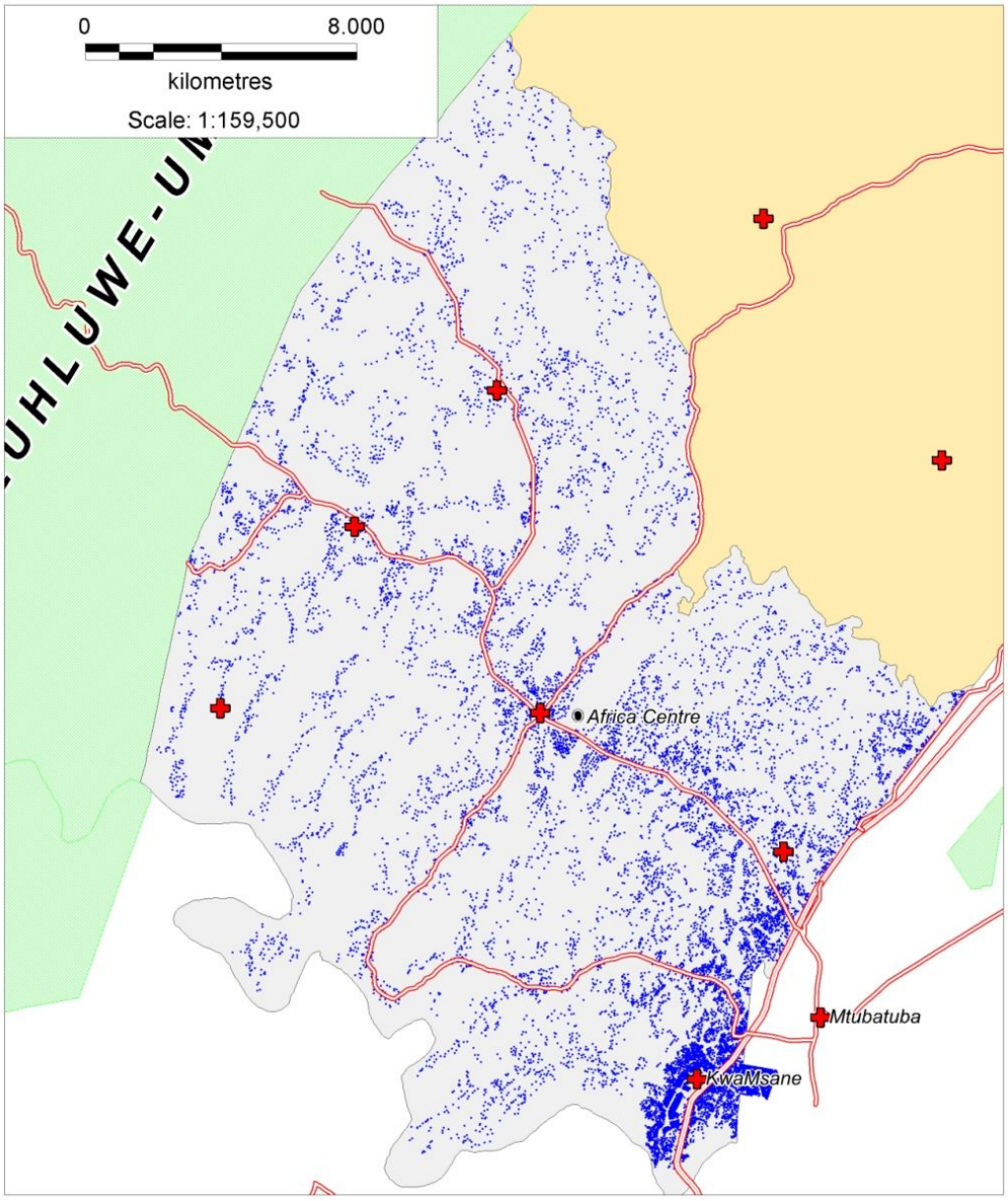


**Figure S2:** Map of the demographic surveillance area showing the geographical population distribution (population ≈60,000). Blue dots represent individual homesteads (N≈11,000) and ART clinics are shown as red crosses. Fieldworkers visit every homestead twice a year to collect socio-demographic data on all household members and visit every homestead once a year to conduct HIV testing in every consenting adult (≥15 years of age).

*The Africa Centre population-based HIV surveillance*

The Africa Centre conducts an annual population-based HIV survey among adults. Between 2004 and 2006 (3 rounds of data collection), all women aged 15-49 years and men aged 15-54 years who resided in the surveillance area were eligible for HIV testing. However, starting in 2007, eligibility was extended to cover all residents aged ≥15 years of age. Any individual who migrates into the area is immediately registered through the bi- annual demographic surveillance and becomes eligible for participation in the annual HIV surveillance. Trained fieldworkers obtain blood specimens by finger prick and prepare dried blood spots for HIV testing according to the Joint UN Programme on HIV/AIDS (UNAIDS) and WHO guidelines (*5*). Within a five-year period, about 80% of all individuals consent to HIV testing. However in a single year of testing the HIV consent rate is ~50% and has increased in the most recent three years. Every individual in the study area is geo-located to their homestead of residence (mapped to an accuracy of less than 2m) (*1*). The use of the Africa Centre demographic data on every individual in the population provides a comprehensive sampling frame for the HIV cohort and eliminates many of the problems commonly affecting surveys, e.g. errors with household listing and selection, and allows a quantification of the effects of non-participation on HIV prevalence estimates. All data collected by the HIV survey can be linked anonymously to other demographic, socio-economic, health and behavioural data collected by the demographic information system.

*Construction of community-level variables*

All patients actively on treatment as of June 31st in each year and all eligible adults in each of the annual population-based HIV surveillance rounds were plotted on maps at the precise location of their respective homesteads of residence (2004-2011). A standard Gaussian kernel (*6-8*) of search radius 3 kilometers (km) was then passed across each map (Figure S3). For every cell on a map, the kernel measured the proportion of the total HIV-infected population receiving ART (total numbers of adult patients [≥15 years of age] actively on ART divided by the total number of HIV-infected adults) within the 3 km search radius for each year. A median of 261 (interquartile range (IQR) 100-652) HIV individuals of known sero-status were included in each community-level calculation.

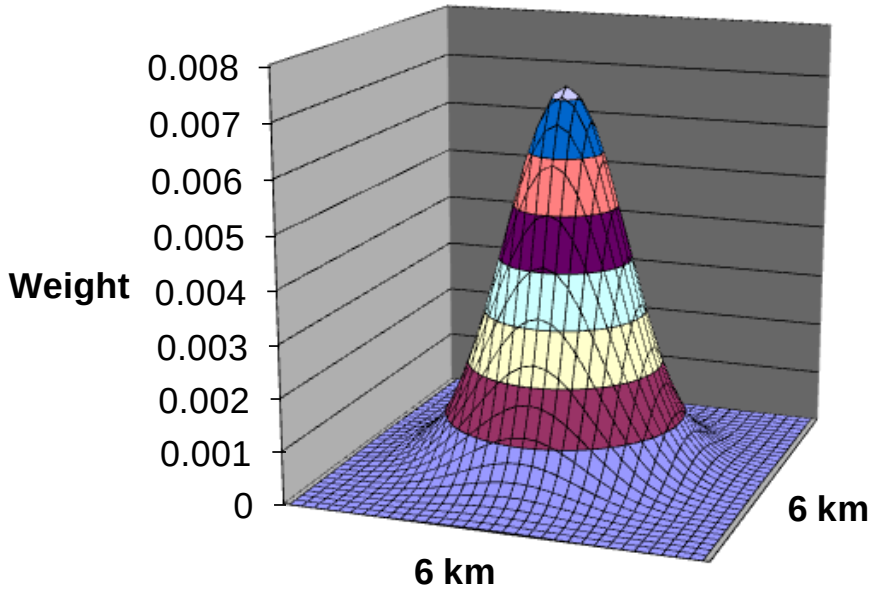


**Figure S3:** Two-dimensional standard Gaussian filter of search radius 3 km used to map geographical variations in HIV prevalence and the proportion of all HIV-infected individuals on ART across the surveillance area. The Z axis shows the weighting given to each cell. The greater the distance from the centre of the kernel the less the weighting assigned to that cell in the calculation of the community-level variables.

We also derived geographical variation in community-level HIV prevalence for each year (2004-2011) in the same fashion. We applied an adjustment factor to the resulting prevalence estimates for 2004-2006 to account for the change in age eligibility criteria in 2007. In view of the scattered distribution of the population and because the most appropriate size of the individual's community was not precisely known, we investigated the potential effect of the size and shape of the kernel, producing estimates with kernels of 2·5 km, 3·0 km, and 3·8 km radii, which evaluated areas between 19·6 $km^2$ and 45·4 $km^2$. We also checked the potential effect of the weighting in the kernel by deriving estimates from a Gaussian kernel of 3km search radius with standard deviation (SD) 2 (as compared to SD =1 in the base-case analysis) as well as a flat kernel of search radius 3km where every cell within a 3 km radius of a particular HIV-uninfected individual was weighted equally. The smaller the search radius used in the kernel and the smaller the standard deviation, the greater the range in community-level estimates.

*Statistical analysis*

All 16,667 participants from the population-based surveillance who lived in the study area and who tested HIV-negative at a first valid HIV test and had at least one subsequent valid HIV test result were included in the analysis. Participants were considered to be at risk of acquiring HIV after the date of their first HIV sero-negative test. The characteristics of the participants in the cohort are shown in Table S1. Each individual was geo-located to their homestead of residence and the ART coverage (%) and HIV prevalence (%) in the surrounding unique local community (as calculated with the Gaussian kernel) was extracted for the time points corresponding to the period of observation of a particular individual. In the base-case, Weibull survival analysis was used to model the time to HIV seroconversion and examine the effect of ART coverage on HIV acquisition risk, controlling for HIV prevalence and other well-established demographic, geographical, economic and behavioural determinants of HIV acquisition (*7, 9*). In addition to the Weibull survival analysis specification, we also repeated our analyses using the exponential specification.

The control variables included in the base-case model were age, sex, household wealth, marital status and number of sexual partners in the previous 12 months (derived by taking the highest number of partners reported in the previous 12 months by an individual over the period of observation) and HIV prevalence and ART coverage in the surrounding local community. Community-level ART coverage, community-level HIV prevalence, age strata and rural/peri-urban/urban homestead locale were included in the main model as time-varying covariates while all other variables were time-independent.

It is crucial to control for the effect of HIV prevalence in the surrounding local community when assessing the effect of ART coverage on HIV acquisition, because the probability that an HIV-uninfected person has unprotected sex with an HIV-infected community member increases with HIV prevalence in the community, at any given level of ART coverage. Community-level ART coverage and HIV prevalence (as measured by a standard Gaussian kernel of radius 3 km) around each individual were included as time-and-space varying exposures in the model. In the analysis, each HIV-uninfected individual in the population cohort is exposed to the ART coverage in the surrounding local community in a particular year from January 1$^{st}$ to December 31$^{st}$. In cases where a participant was resident at multiple homesteads in a year, the individual was assigned multiple exposure episodes in the model (corresponding to the ART coverage and HIV prevalence in the community surrounding each homestead whilst that individual was resident). By including both ART coverage and HIV prevalence in the model as space- and time-varying covariates, we account for the rapid roll-out in ART and associated increase in HIV prevalence in this population between 2004 and 2011 and the fact that each HIV-uninfected individual will have differing exposures over time and space.

We used exponential and Weibull distributions to model time to HIV sero-conversion. In comparison to the exponential specification, the Weibull specification offers more flexibility because it has an extra parameter and allows for time variation in the HIV acquisition hazard. The hypothesis that the shape parameter of the Weibull model is equal to zero was rejected (p-value=0.003), providing evidence in favour of the Weibull model.

In the base-case analysis, we thus assume that time to HIV seroconversion follows a Weibull survival distribution, $S(t)=\exp(-(\lambda t)^{\alpha})$, where $\lambda$ and $\alpha>0$ are the scale and shape parameter of the Weibull distribution. The hazard function is then defined:

$$h(\mathbf{t}|\mathbf{X},\mathbf{Z})=\alpha\mathbf{t}^{\alpha-1}\exp(\mathbf{X\beta}+\mathbf{Z\gamma})$$

where, $\mathbf{X}_{Nxk}$, $\mathbf{Z}_{Nxm}$ are the individual- and community-specific covariate matrices and $\boldsymbol{\beta}_{kx1}$,$\boldsymbol{\gamma}_{mx1}$ are the vectors of their respective coefficients. Community-level ART coverage, community-level HIV prevalence, rural/peri-urban/urban homestead locale and agestrata were allowed to vary over time in our analyses. The hazard ratio for covariate p is obtained through $HR_i=\exp(\beta_i)$ when the covariate is individual-specific or $HR_j=\exp(\gamma_j)$ when it is community-specific. We are interested in testing the hypothesis that $\beta_{ART}=0$ (there is no association between ART coverage and incidence of HIV).

Survey participants are not required to answer all questions, and thus missing data arise in marital status (3.7%) and number of sexual partners in the last twelve months (31.7%) variables. To account for this data missingness we used a multiple imputation procedure with five imputed datasets (*10*). All analyses were done with Stata (version 11.2). Since the exact date of seroconversion was not precisely known, we assigned a random date of seroconversion (drawn from a flat distribution) between the last HIV sero-negative and first sero-positive result.

**Table S1.** Descriptive characteristics of the population of HIV-negative individuals (N=16,667) followed over time (2004-2011).

| Covariate | Person-years | Sero-conversions (cases) | HIV incidence rate (per 100 person-years) |
|---|---|---|---|
| ***Community-level**** | | | |
| **ART coverage** | | | |
| <10% | 20385.5 | 615 | 3.02 (2.79-3.26) |
| 10-20% | 10877.7 | 312 | 2.87 (2.57-3.20) |
| 20-30% | 12987.8 | 326 | 2.51 (2.25-2.80) |
| 30-40% | 7100.5 | 128 | 1.80 (1.52-2.14) |
| >40% | 2253.8 | 32 | 1.42 (1.00-2.01) |
| **HIV prevalence** | | | |
| <10% | 2868.5 | 45 | 1.57 (1.17-2.10) |
| 10-15% | 10561.2 | 231 | 2.19 (1.92-2.49) |
| 15-20% | 17824.9 | 458 | 2.57 (2.34-2.82) |
| 20-25% | 14169.2 | 417 | 2.94 (2.67-3.24) |
| >25% | 8147.23 | 259 | 3.18 (2.81-3.60) |
| | | | |
| ***Individual-level*** | | | |
| **Year** | | | |
| 2004 | 5666.3 | 155 | 2.73 (2.33-3.20) |
| 2005 | 6907.1 | 216 | 3.12 (2.74-3.57) |
| 2006 | 7341.0 | 229 | 3.12 (2.74-3.55) |
| 2007 | 8012.5 | 235 | 2.93 (2.58-3.33) |
| 2008 | 8800.8 | 213 | 2.42 (2.12-2.77) |
| 2009 | 8092.0 | 172 | 2.12 (1.83-2.47) |
| 2010 | 6724.3 | 149 | 2.21 (1.88-2.60) |
| 2011 | 2027.1 | 41 | 2.20 (1.49-2.75) |
| **Sex-age** (years) | | | |
| Female <20 | 7035.8 | 312 | 4.43 (3.96-4.95) |
| Female 20-24 | 5775.9 | 375 | 6.49 (5.86-7.18) |
| Female 25-29 | 2375.7 | 131 | 5.51 (4.64-6.54) |
| Female 30-34 | 1758.4 | 62 | 3.52 (2.75-4.52) |
| Female 35-39 | 2254.78 | 51 | 2.26 (1.72-2.97) |
| Female 40-44 | 2868.6 | 64 | 2.23 (1.75-2.85) |
| Female >45 | 12242.2 | 81 | 0.66 (0.53-0.82) |
| Male <20 | 6458.4 | 48 | 0.74 (0.56-0.99) |
| Male 20-24 | 4309.9 | 109 | 2.53 (2.10-3.05) |
| Male 25-29 | 1556.5 | 69 | 4.43 (3.50-5.61) |
| Male 30-34 | 857.7 | 25 | 2.91 (1.97-4.31) |
| Male 35-39 | 853.2 | 27 | 3.16 (2.17-4.61) |
| Male 40-44 | 820.9 | 14 | 1.71 (1.01-2.88) |
| Male >45 | 4403.1 | 42 | 0.95 (0.70-1.29) |

| | | | |
|---|---|---|---|
| **Number of partners in the last 12 months** | | | |
| 0 | 3661.2 | 33 | 0.90 (0.64-1.27) |
| 1 | 32381.3 | 1107 | 3.42 (3.22-3.63) |
| >1 | 4115.6 | 162 | 3.94 (3.37-4.59) |
| **Marital status** | | | |
| Single | 41490.4 | 1269 | 3.06 (2.89-3.23) |
| Married, monogamous | 10040.3 | 105 | 1.05 (0.86-1.27) |
| Married, polygamous | 973.1 | 13 | 1.34 (0.78-2.30) |
| **Area of residence** | | | |
| Peri-urban | 15498.2 | 506 | 3.26 (2.99-3.56) |
| Rural | 36753.1 | 875 | 2.38 (2.23-2.54) |
| Urban | 1354.0 | 32 | 2.36 (1.67-3.34) |
| **Wealth quintile** | | | |
| Poorest | 10603.1 | 280 | 2.64 (2.35-2.97) |
| 2nd poorest | 11142.5 | 290 | 2.60 (2.32-2.92) |
| 3rd poorest | 11636.4 | 318 | 2.73 (2.45-3.05) |
| 4th poorest | 11776.4 | 317 | 2.69 (2.41-3.01) |
| Wealthiest | 8446.9 | 208 | 2.46 (2.15-2.82) |

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

**Table S2.** Full output from four nested multivariable survival models showing the effect of coverage of ART in the surrounding community on an HIV-uninfected individual's hazard of acquiring HIV infection (N=16,667). Model 1 shows the ART coverage hazard ratios adjusted for age and sex, while in model 2, other socio-economic and environmental covariates are added. Model 3 also includes community-level HIV prevalence and model 4 includes all covariates in model 3 with the addition of number of sexual partners in the last 12 months.

| | Univariable analysis | | Model 1 | | Model 2 | | Model 3 | | Model 4† | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Covariate** | **HR (95% CI)** | **p-value** | **aHR (95% CI)** | **p-value** | **aHR (95% CI)** | **p-value** | **aHR (95% CI)** | **p-value** | **aHR (95% CI)** | **p-value** |
| ***Community-level**** | | | | | | | | | | |
| **ART coverage‡** (vs <10%) | | | | | | | | | | |
| 10-20% | 0.85 (0.74-0.98) | 0.030 | 0.97 (0.84-1.12) | 0.642 | 0.96 (0.83-1.11) | 0.619 | 0.93 (0.80-1.08) | 0.321 | 0.93 (0.8-1.08) | 0.325 |
| 20-30% | 0.72 (0.62-0.83) | <0.0001 | 0.90 (0.78-1.05) | 0.182 | 0.87 (0.75-1.01) | 0.076 | 0.78 (0.76-0.92) | 0.002 | 0.79 (0.67-0.92) | 0.003 |
| 30-40% | 0.51 (0.42-0.63) | <0.0001 | 0.66 (0.54-0.81) | <0.0001 | 0.67 (0.55-0.82) | <0.0001 | 0.61 (0.50-0.75) | <0.0001 | 0.62 (0.50-0.76) | <0.0001 |
| >40% | 0.40 (0.28-0.58) | <0.0001 | 0.54 (0.38-0.78) | 0.001 | 0.59 (0.41-0.85) | 0.004 | 0.62 (0.43-0.89) | 0.010 | 0.63 (0.44-0.91) | 0.013 |
| **HIV prevalence‡** (vs <10%) | | | | | | | | | | |
| 10-15% | 1.40 (1.02-1.93) | 0.039 | - | - | - | - | 1.44 (1.04-1.99) | 0.030 | 1.42 (1.03-1.97) | 0.034 |
| 15-20% | 1.63 (1.20-2.22) | 0.002 | - | - | - | - | 1.70 (1.24-2.33) | 0.001 | 1.69 (1.23-2.31) | 0.001 |
| 20-25% | 1.86 (1.37-2.54) | <0.0001 | - | - | - | - | 2.01 (1.45-2.79) | <0.0001 | 2.00 (1.44-2.78) | <0.0001 |
| >25% | 1.99 (1.45-2.74) | <0.0001 | - | - | - | - | 2.24 (1.57-3.19) | <0.0001 | 2.20 (1.55-3.14) | <0.0001 |
| ***Individual-level*** | | | | | | | | | | |
| **Age-Sex strata‡** (vs male 15-19 years old) | | | | | | | | | | |
| Male 20-24 | 3.29 (2.34-4.64) | <0.0001 | 3.26 (2.32-4.59) | <0.0001 | 3.20 (2.27-4.50) | <0.0001 | 3.16 (2.24-4.45) | <0.0001 | 3.03 (2.15-4.27) | <0.0001 |
| Male 25-29 | 5.77 (3.98-8.36) | <0.0001 | 5.73 (3.95-8.3) | <0.0001 | 5.61 (3.86-8.13) | <0.0001 | 5.48 (3.78-7.95) | <0.0001 | 5.22 (3.6-7.58) | <0.0001 |
| Male 30-34 | 3.82 (2.35-6.2) | <0.0001 | 3.83 (2.36-6.23) | <0.0001 | 3.62 (2.20-5.96) | <0.0001 | 3.61 (2.19-5.94) | <0.0001 | 3.67 (2.23-6.04) | <0.0001 |
| Male 35-39 | 4.15 (2.58-6.65) | <0.0001 | 4.09 (2.55-6.56) | <0.0001 | 4.22 (2.61-6.82) | <0.0001 | 4.13 (2.56-6.68) | <0.0001 | 4.32 (2.67-6.99) | <0.0001 |
| Male 40-44 | 2.23 (1.23-4.05) | 0.008 | 2.20 (1.21-4.00) | 0.010 | 2.34 (1.28-4.25) | 0.006 | 2.29 (1.26-4.18) | 0.007 | 2.47 (1.36-4.5) | 0.003 |
| Male >=45 | 1.25 (0.82-1.89) | 0.294 | 1.29 (0.85-1.95) | 0.233 | 1.59 (1.03-2.46) | 0.037 | 1.55 (1.00-2.39) | 0.049 | 1.71 (1.1-2.64) | 0.017 |
| Female 15-19 | 5.97 (4.40-8.09) | <0.0001 | 5.98 (4.41-8.11) | <0.0001 | 5.97 (4.40-8.09) | <0.0001 | 5.97 (4.40-8.09) | <0.0001 | 6.36 (4.67-8.64) | <0.0001 |
| Female 20-24 | 8.45 (6.24-11.44) | <0.0001 | 8.40 (6.20-11.37) | <0.0001 | 8.36 (6.17-11.33) | <0.0001 | 8.36 (6.17-11.32) | <0.0001 | 8.92 (6.56-2.12) | <0.0001 |
| Female 25-29 | 7.16 (5.13-9.99) | <0.0001 | 7.21 (5.17-10.06) | <0.0001 | 7.28 (5.21-10.18) | <0.0001 | 7.34 (5.25-10.25) | <0.0001 | 7.74 (5.52-0.85) | <0.0001 |
| Female 30-34 | 4.62 (3.16-6.74) | <0.0001 | 4.58 (3.13-6.68) | <0.0001 | 4.93 (3.37-7.24) | <0.0001 | 4.92 (3.36-7.22) | <0.0001 | 5.38 (3.66-7.91) | <0.0001 |
| Female 35-39 | 2.96 (1.99-4.39) | <0.0001 | 2.93 (1.97-4.35) | <0.0001 | 3.35 (2.23-5.03) | <0.0001 | 3.42 (2.28-5.12) | <0.0001 | 3.71 (2.46-5.58) | <0.0001 |
| Female 40-44 | 2.91 (2.00-4.24) | <0.0001 | 2.86 (1.96-4.17) | <0.0001 | 3.25 (2.20-4.80) | <0.0001 | 3.21 (2.17-4.75) | <0.0001 | 3.61 (2.44-5.36) | <0.0001 |
| Female >=45 | 0.87 (0.6-1.24) | 0.429 | 0.91 (0.64-1.31) | 0.613 | 0.99 (0.68-1.43) | 0.939 | 0.96 (0.66-1.40) | 0.853 | 1.14 (0.78-1.67) | 0.494 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Area of residence‡** (vs rural) | | | | | | | | | | |
| Peri-urban | 1.37 (1.23-1.53) | <0.0001 | - | - | 1.37 (1.22-1.54) | <0.0001 | 1.12 (0.97-1.30) | 0.109 | 1.13 (0.98-1.3) | 0.105 |
| Urban | 1.00 (0.7-1.43) | 0.985 | - | - | 1.13 (0.79-1.63) | 0.037 | 0.92 (0.63-1.33) | 0.657 | 0.93 (0.64-1.35) | 0.703 |
| **Marital status** (vs single) | | | | | | | | | | |
| Married monogamous | 0.33 (0.27-0.41) | <0.0001 | - | - | 0.67 (0.53-0.84) | 0.001 | 0.69 (0.54-0.86) | 0.001 | 0.67 (0.53-0.84) | 0.001 |
| Married polygamous | 0.43 (0.25-0.75) | 0.003 | - | - | 1.14 (0.65-2.00) | 0.645 | 1.17 (0.67-2.05) | 0.585 | 1.14 (0.65-2.01) | 0.64 |
| **Number of partners in the last 12 months** (vs zero) | | | | | | | | | | |
| One | 3.22 (2.18-4.76) | <0.0001 | - | - | - | - | - | - | 1.77 (1.17-2.67) | 0.008 |
| More than one | 3.72 (2.47-5.61) | <0.0001 | - | - | - | - | - | - | 2.49 (1.61-3.86) | <0.0001 |
| **Wealth quintile** (vs poorest) | | | | | | | | | | |
| 2nd poorest | 0.99 (0.84-1.16) | 0.881 | - | - | 0.95 (0.80-1.12) | 0.520 | 0.96 (0.82-1.14) | 0.661 | 0.96 (0.82-1.14) | 0.658 |
| 3rd poorest | 1.03 (0.88-1.21) | 0.721 | - | - | 0.96 (0.81-1.13) | 0.621 | 0.97 (0.82-1.14) | 0.690 | 0.97 (0.82-1.14) | 0.676 |
| 4th poorest | 1.02 (0.87-1.20) | 0.813 | - | - | 0.92 (0.78-1.09) | 0.335 | 0.92 (0.78-1.08) | 0.297 | 0.92 (0.78-1.08) | 0.296 |
| Wealthiest | 0.94 (0.78-1.12) | 0.468 | - | - | 0.84 (0.70-1.02) | 0.072 | 0.82 (0.68-0.99) | 0.042 | 0.82 (0.68-1.00) | 0.044 |
| **Weibull shape parameter** | - | - | 2.95 (2.8-3.13) | 0.003 | 2.96 (2.80-3.14) | 0.002 | 2.97 (2.80-3.15) | 0.002 | 2.94 (2.80-3.13) | 0.003 |

HR = hazard ratio, aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

†Corresponding values for a model, in which community-level ART coverage and HIV prevalence are modeled as continuous covariates (%); ART coverage aHR = 0.986 (95% CI 0.981-0.991), p<0.0001; HIV prevalence aHR = 1.038 (95% CI 1.026 -1 .050), p<0.0001. ‡ Indicates covariates that vary with time.

**Sensitivity Analyses:**

*The potential influence of other prevention services that increased during the study period*

Although we control for the important determinants of acquisition of infection in this population (*7, 9*), we wanted to exclude virtually all other possible explanations for the relationship between ART coverage and risk of acquiring infection. To do this, we first investigated the extent to which other factors that could potentially be correlated with the space-time roll-out of ART could explain the observed association between ART coverage and HIV acquisition risk. In a recent paper we analysed changes in sexual behaviour indicators in this population (*11*). We investigated trends over time in the following indicators: % that had ever had sex; mean number of partners in the last year; % with multiple partners in the last year; point prevalence of those reporting concurrent partnerships (%); and age difference with regular partner (years). While these sexual partnership variables showed no significant change over the study period, condom use showed a significant increase. Condom use (both sexes) at the last sex act with a regular partner in the last year increased from 26.2% in 2005 to 54.3% in 2011. The proportion of men who reported being circumcised did not change significantly over the study period (5.8% in 2004 and 5.0% in 2011).

Next, we mapped condom use across the surveillance area over the study observation period (2006-2011) using the Gaussian kernel methodology (Figure S4). We then included this variable in the model as space- and time-varying covariate, as described in the paper, to investigate the effect on the ART coverage result (Table S3, Figure S5). Levels of condom use in the surrounding local community were not associated with a reduction in risk of infection and the ART coverage finding was

robust to controlling for this additional variable.

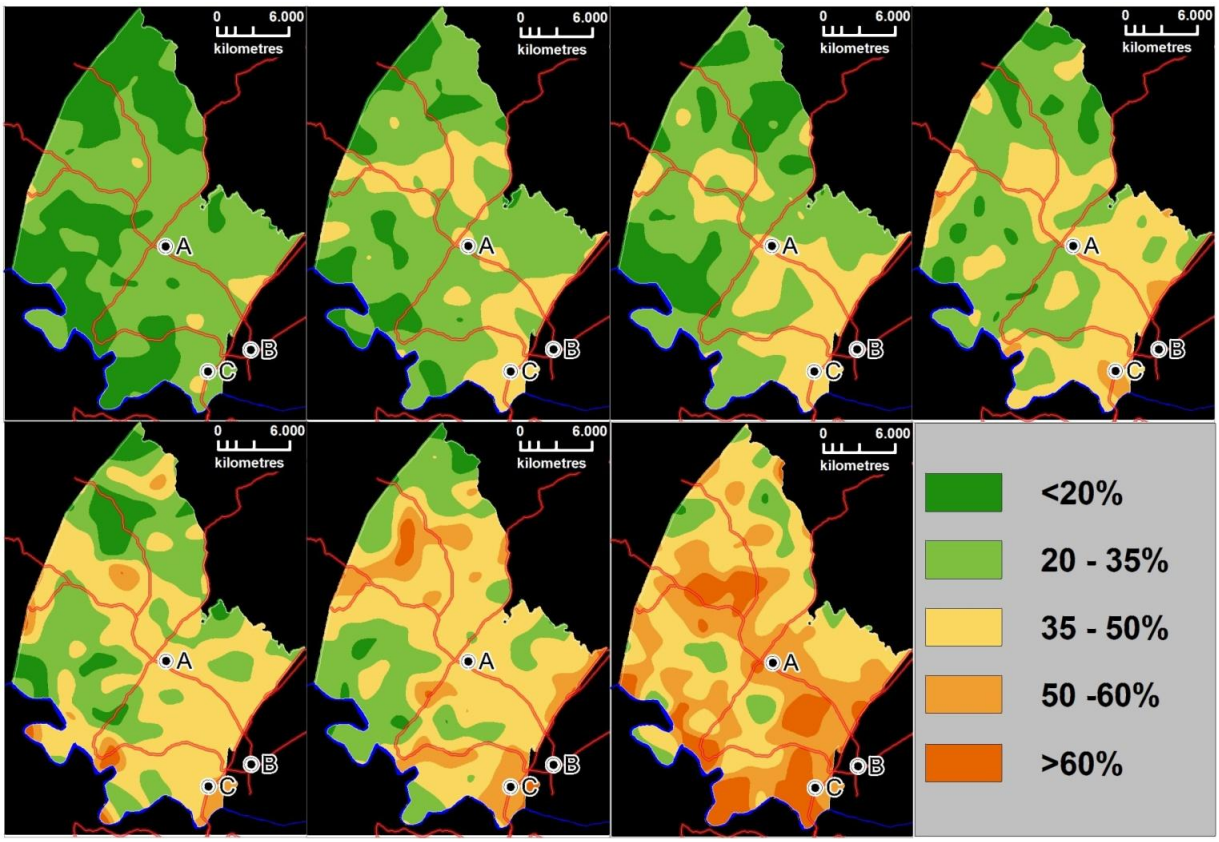


**Figure S4:** Time series of maps showing condom use (male and female) at last sex with regular partner (%) across the demographic surveillance area (2005-2008 from left to right in the top row, and 2009-2011 from left to in the right bottom row). A = Africa Centre, B = Mtubatuba Town, C = KwaMsane Township. Main roads are also superimposed for ease of reference. Maps are derived using a standard Gaussian kernel of radius 3 km. Condom use at last sex with regular partner increased from 26.2% in 2005 to 54.3% in 2011 ($p<0.001$).

**Table S3:** Final model with community-level condom use* (%) added – 2005 to 2011 (N=16,312).

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| **Covariate†** | **aHR (95% CI)** | **p-value** | **aHR (95% CI)** | **p-value** |
| ***Community-level*‡** | | | | |
| ART coverage (%) | 0.984 (0.978-0.990) | <0.0001 | 0.986 (0.979-0.992) | <0.0001 |
| HIV prevalence (%) | 1.034 (1.022-1.048) | <0.0001 | 1.037 (1.023-1.052) | <0.0001 |
| Condom use (%) | - | - | 0.996 (0.988-1.004) | 0.309 |
| ***Individual-level*** | | | | |
| **Age-sex groups** (vs male 15-19 years old) | | | | |
| Male 20-24 | 3.344 (2.144-5.215) | <0.0001 | 3.344 (2.144-5.215) | <0.0001 |
| Male 25-29 | 6.031 (3.771-9.646) | <0.0001 | 6.037 (3.775-9.655) | <0.0001 |
| Male 30-34 | 3.403 (1.779-6.510) | <0.0001 | 3.400 (1.778-6.504) | <0.0001 |
| Male 35-39 | 4.988 (2.761-9.010) | <0.0001 | 4.994 (2.764-9.021) | <0.0001 |
| Male 40-44 | 2.516 (1.177-5.377) | 0.017 | 2.512 (1.175-5.370) | 0.018 |
| Male >=45 | 1.818 (1.063-3.108) | 0.029 | 1.818 (1.064-3.108) | 0.029 |
| Female 15-19 | 7.648 (5.112-11.441) | <0.0001 | 7.630 (5.100-11.415) | <0.0001 |
| Female 20-24 | 10.607 (7.098-15.852) | <0.0001 | 10.575 (7.075-15.805) | <0.0001 |
| Female 25-29 | 8.325 (5.393-12.850) | <0.0001 | 8.296 (5.374-12.806) | <0.0001 |
| Female 30-34 | 6.664 (4.132-10.747) | <0.0001 | 6.649 (4.123-10.723) | <0.0001 |
| Female 35-39 | 3.432 (2.005-5.875) | <0.0001 | 3.424 (2.000-5.862) | <0.0001 |
| Female 40-44 | 4.062 (2.461-6.705) | <0.0001 | 4.051 (2.454-6.688) | <0.0001 |
| Female >=45 | 1.228 (0.769-1.961) | 0.39 | 1.228 (0.769-1.962) | 0.389 |
| **Area of residence** (vs rural) | | | | |
| Peri-urban | 0.993 (0.844-1.168) | 0.928 | 1.017 (0.858-1.204) | 0.848 |
| Urban | 0.751 (0.491-1.147) | 0.185 | 0.769 (0.502-1.179) | 0.229 |
| **Marital status** (vs single) | | | | |
| Married monogamous | 0.637 (0.480-0.846) | 0.002 | 0.635 (0.478-0.842) | 0.002 |
| Married polygamous | 1.301 (0.703-2.411) | 0.402 | 1.293 (0.698-2.396) | 0.413 |
| **Number of partners in the last 12 months** (vs 0) | | | | |
| 1 | 1.820 (1.155-2.867) | 0.011 | 1.817 (1.154-2.863) | 0.011 |
| >1 | 2.352 (1.443-3.831) | 0.001 | 2.350 (1.443-3.83) | 0.001 |
| **Wealth quintile** (vs poorest) | | | | |
| 2nd poorest | 0.923 (0.765-1.114) | 0.404 | 0.924 (0.766-1.114) | 0.407 |
| 3rd poorest | 0.863 (0.714-1.042) | 0.125 | 0.864 (0.716-1.044) | 0.130 |
| 4th poorest | 0.856 (0.708-1.034) | 0.107 | 0.862 (0.713-1.043) | 0.126 |
| Wealthiest | 0.801 (0.646-0.993) | 0.043 | 0.811 (0.654-1.007) | 0.058 |
| **Weibull shape parameter** | 3.040 (2.820-3.290) | 0.003 | 2.251 (1.922-2.737) | 0.003 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Condom-use at last sex (in the last year) with regular partner (%).
†Model 1 is the final model with continuous ART coverage and HIV prevalence. Community-level "Condom use" is added in Model 2.
‡Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

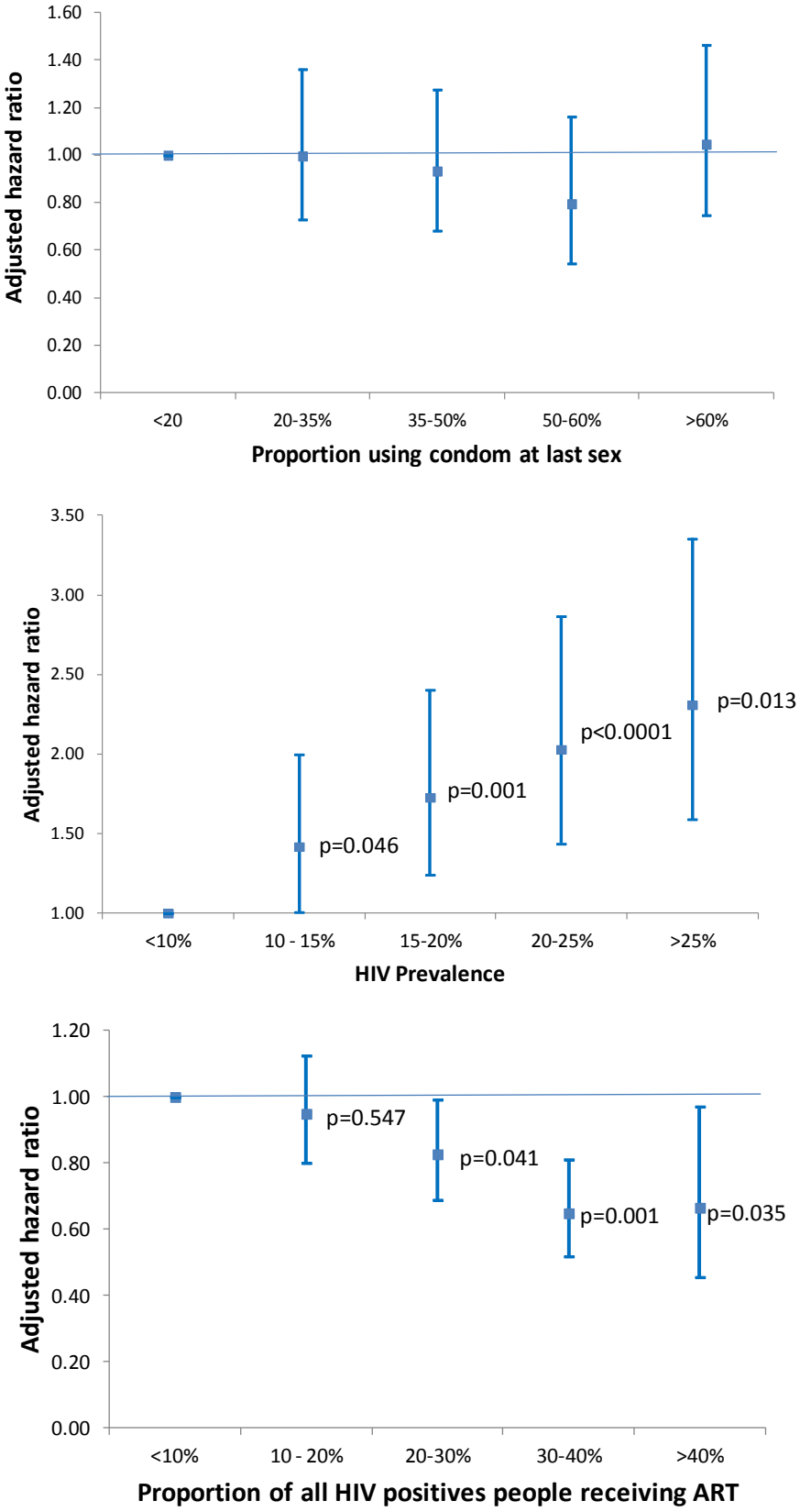


**Figure S5:** Graphs comparing the adjusted hazard of acquiring HIV across three community-level variables that increased significantly between 2005 and 2011 (top = condom use, middle = HIV prevalence, bottom = ART coverage). The hazard ratios are adjusted for all other variables in the final model (Table S3).

To further assess the potential influence of other prevention services that increased during the study period, we investigated whether the age-patterns of a reduction in risk of acquisition of HIV infection with increasing ART coverage were consistent with a 'treatment-as-prevention' effect. Since ART coverage is higher in the older age-groups (>35 years of age) (*12*) and individuals in this age-group largely choose partners of a similar age (*13*), we would expect to see larger declines in risk of acquisition of HIV infection in the older age-groups as a higher proportion of HIV-infected partners would be on ART. Consistent with our hypothesis, Figure S6 shows that declines in the risk of acquisition of infection with increasing coverage is this older segment of the population (>35 years) is more pronounced then for the population as a whole (Figure 3). These reductions occurred after adjustment for all of the factors in the final model, including HIV prevalence (Table 1).

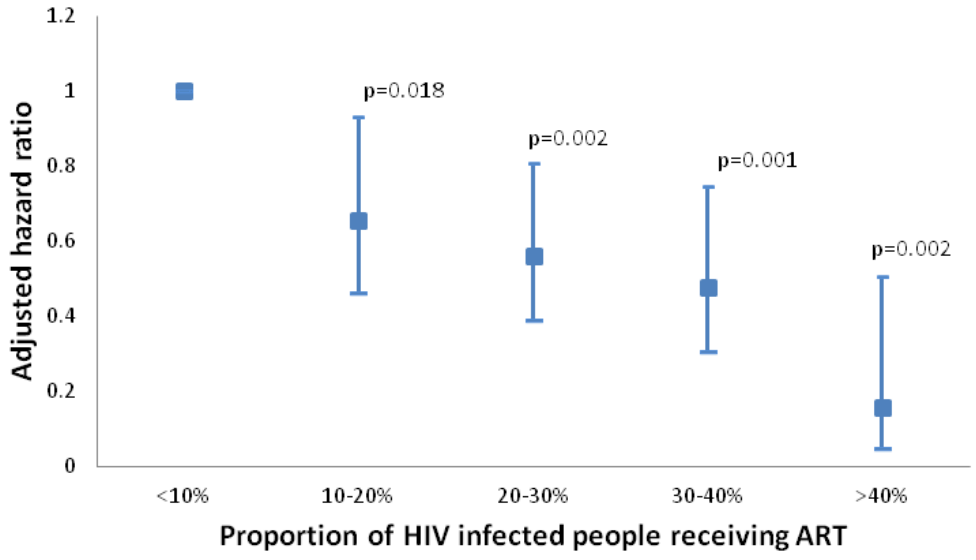


**Figure S6:** Graph showing the adjusted hazard of acquiring HIV for all HIV-uninfected individuals >35 years of age in the population cohort according to the proportion of the total HIV-infected population receiving ART in the surrounding local community.

*Influence of secular time-trends on the results*

Having directly controlled for the important determinants of acquisition of infection in this population, as well as the influence of additional prevention services that increased over the study period, we also wanted to establish whether other changes in the community over the time period (correlated with the space-time roll-out of ART but not directly measured) might explain our results. The adjusted hazard of acquiring HIV (Figure S7) shows a declining trend between 2004 and 2011. Over this period, there were several important programmatic changes in the Hlabisa Treatment and Care program (*14*) as well as national HIV testing and HIV prevention campaigns designed to get increased numbers of individuals to access testing. We investigated the influence of these effects through the introduction of time-period dummy variables in the model. In 2004, access to treatment was limited to district hospital; and in 2005 and 2006 the program started to devolve to all clinics in the surveillance area. The period 2007-2009 saw a massive scale-up of treatment delivery following a sharp increase in PEPFAR funding. In 2010, the CD4+ treatment eligibility threshold was increased to 350 cells/µl for TB patients and pregnant women and a national campaign was launched to encourage knowledge of HIV status. Finally, in 2011, the 350 cells/µl CD4+ treatment eligibility threshold was extended to all HIV-infected adults. After controlling for these key secular time effects in the model, we see that the ART coverage results remain large and significant, confirming that other changes that took place in the community over the time period are unlikely to explain the ART coverage results.

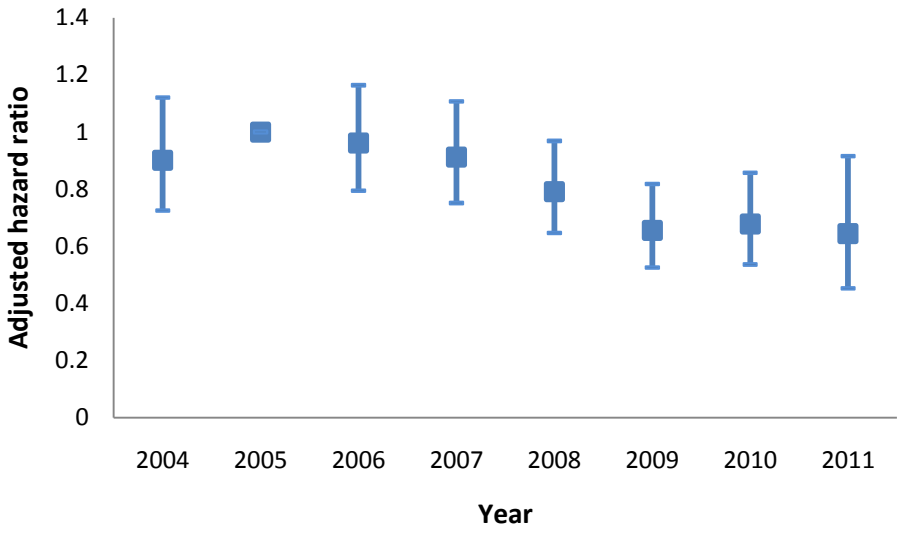


**Figure S7:** Graph showing the adjusted hazard of acquiring HIV by year (2004-2011) among the 16,667 HIV-uninfected participants. The hazard ratios are adjusted for age and sex, community-level HIV prevalence, rural/peri-urban/urban homestead locale, marital status, household-level wealth quintile, and number of partners in the last 12 months.

**Table S4:** Results of the final model (Table 1) with time-period dummy variables added that capture key changes in the scale-up of ART between 2004 and 2011 (N=16,667).

| Covariate | aHR (95% CI) | p-value |
|---|---|---|
| ***Community-level****† | | |
| **ART coverage (vs <10%)** | | |
| 10-20% | 0.94 (0.73-1.22) | 0.661 |
| 20-30% | 0.81 (0.61-1.08) | 0.148 |
| 30-40% | 0.63 (0.46-0.86) | 0.004 |
| >40% | 0.63 (0.40-1.00) | 0.049 |
| **HIV prevalence (vs <10%)** | | |
| 10-15% | 1.43 (1.03-1.99) | 0.031 |
| 15-20% | 1.70 (1.24-2.33) | 0.001 |
| 20-25% | 2.03 (1.46-2.82) | <0.0001 |
| >25% | 2.23 (1.56-3.19) | <0.0001 |
| | | |
| ***Individual-level*** | | |
| **Time period‡ (vs 2004)** | 1.09 (0.89-1.33) | 0.386 |
| 2005-06 | 1.05 (0.77-1.44) | 0.749 |
| 2007-09 | 1.01 (0.70-1.46) | 0.953 |
| 2010 | 1.14 (0.71-1.84) | 0.591 |
| 2011 | 1.09 (0.89-1.33) | 0.386 |
| **Age-Sex groups (vs male 15-19 years old)** | | |
| Male 20-24 | 3.04 (2.15-4.28) | <0.0001 |
| Male 25-29 | 5.24 (3.61-7.61) | <0.0001 |
| Male 30-34 | 3.68 (2.23-6.06) | <0.0001 |
| Male 35-39 | 4.33 (2.68-7.01) | <0.0001 |
| Male 40-44 | 2.48 (1.36-4.52) | 0.003 |
| Male >=45 | 1.71 (1.10-2.65) | 0.016 |
| Female 15-19 | 6.36 (4.68-8.64) | <0.0001 |
| Female 20-24 | 8.95 (6.58-12.16) | <0.0001 |
| Female 25-29 | 7.77 (5.54-10.9) | <0.0001 |
| Female 30-34 | 5.40 (3.67-7.95) | <0.0001 |
| Female 35-39 | 3.73 (2.48-5.61) | <0.0001 |
| Female 40-44 | 3.63 (2.45-5.38) | <0.0001 |
| Female >=45 | 1.14 (0.78-1.67) | 0.488 |
| **Area of residence** (vs rural) | | |
| Peri-urban | 1.12 (0.97-1.30) | 0.118 |
| Urban | 0.93 (0.64-1.34) | 0.688 |
| **Marital status** (vs single) | | |
| Married monogamous | 0.67 (0.53-0.84) | 0.001 |
| Married polygamous | 1.14 (0.65-2.01) | 0.638 |

| | | |
|---|---|---|
| **Number of partners in the last 12 months** (vs zero) | | |
| One | 1.77 (1.17-2.67) | 0.008 |
| More than one | 2.49 (1.61-3.86) | <0.0001 |
| **Wealth quintile** (vs poorest) | | |
| 2nd poorest | 0.96 (0.82-1.14) | 0.669 |
| 3rd poorest | 0.97 (0.82-1.14) | 0.683 |
| 4th poorest | 0.92 (0.78-1.08) | 0.298 |
| Wealthiest | 0.83 (0.68-1.00) | 0.045 |
| **Weibull shape parameter** | 2.93 (2.76-3.12) | 0.013 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

†Corresponding values for a model, in which community-level ART coverage and HIV prevalence are modeled as continuous covariates (%); ART coverage aHR = 0.985 (95% CI 0.977-0.993),p<0.0001; HIV prevalence aHR = 1.036 (95% CI 1.022 -1 .050), p<0.0001.
‡ 2004:ART only hospital based; 2005-2006:ART starts to devolve to clinics in the surveillance area; 2007-2009: rapid scale up of ART following increased PEPFAR funding; 2010: CD4+ treatment eligibility threshold increased to 350 cells/µl for TB patients and pregnant women and a national VCT campaign launched; 2011:350 cells/µl CD4+ treatment eligibility threshold was extended to all adults.

*Spatial clustering of residuals*

To further investigate the possibility that unmeasured characteristics of the local community that vary in space could partially explain the association, we investigated the spatial clustering of Martingale residuals in this population. The distribution of new infections in this population has been shown to exhibit strong spatial clustering patterns (*15*). If such spatial clustering were to persist after adjusting for the factors in our model it would indicate the influence of other factors not included in our model.

We calculated martingale residuals (one value per subject) using stata 11.2. We then used a Kulldorff spatial scan statistic (implemented within the SaTScan spatial cluster detection program (*16*) ) to analyse whether the residuals clustered in space. A spatial scan statistic is a cluster detection test able to both detect the location of clusters and evaluate their statistical significance without problems associated with multiple testing. This is done by gradually scanning a window across space. The scan statistic adjusts for the uneven geographical density of a background population and the analyses are conditioned on the total number of cases observed (*6, 17*). The scan statistic will then scan the entire study area for clusters using a normal model (*18*). The spatial scan statistic imposes a circular window on a map and it allows the centre of the circle to move across the study region constructing a series of circles around every one of the 12,000 homesteads in the surveillance area. The radius of the circle changes continuously so that it can take any value from zero up to a specified maximum value. For each circle *z*, a log likelihood ratio ($LLR(z)$) is calculated, and the test statistic is defined as the maximum *LLR* over all circles. The circle with the maximum likelihood is defined as the most likely cluster, implying that it is least likely to have occurred by chance. The maximum observed value of the test statistic for each possible cluster is then compared to the overall distribution of maximum values. The p-value of the statistic is obtained through

Monte Carlo hypothesis testing (9,999 iterations), where the null hypothesis of no clustering is rejected if the simulated p-value is <0.05.

After controlling for all variables in our model, no residual spatial clustering of new HIV infections was observed in our results. We therefore found no evidence to suggest that unmeasured characteristics of the local community were contributing significantly to the finding.

*The potential influence of uncertainty regarding date of sero-conversion*

We assigned a random date of sero-conversion (drawn from a flat distribution) between the last negative and the first positive HIV test in the base-case analysis. To examine whether the uncertainty of this date had any bearing on the result we performed 1000 repetitions of our analysis, choosing a different random date (between an individual's last HIV-negative and first HIV-positive test) for each repetition for every sero-converting individual. The results of these replications are summarized (Table S5); the ART coverage finding remained essentially unchanged. We further repeated our analysis assigning the midpoint between the last HIV-negative and the first HIV-positive results as the sero-conversion date (Table S6). Again, our findings remained essentially unchanged, when this approach to sero-conversion date assignment was used. Thus, these sensitivity analyses demonstrate that the ART coverage results are not sensitive changes in the approach used to determine sero-conversion dates.

**Table S5:** Results of multiple runs (N=1000) of the final model (Table 1) assigning a random date between last HIV-negative and first HIV-positive test.

| Covariate | aHR (95% CI)* † | p-value* |
|---|---|---|
| **ART coverage** (vs <10%) | - | - |
| 10-20% | 0.91 (0.82-1.02) | 0.088 |
| 20-30% | 0.77 (0.69-0.85) | <0.0001 |
| 30-40% | 0.74 (0.64-0.84) | <0.0001 |
| >40% | 0.56 (0.42-0.72) | <0.0001 |

aHR = adjusted hazard ratio, CI = confidence interval.

*The traditional definitions of confidence interval and p-value are not valid in this analysis, but we calculated the high-probability regions of the sampling distribution of the coefficients as the percentiles where $P(b \leq b_1)=0.025$ and $P(b \geq b_2)=0.975$ with $b_1$ and $b_2$ the lower and upper limit of the interval. Regarding the p-value, a definition commonly used in bootstrap analysis was applied; it was calculated as $2*P(b \leq 1)$, which is the probability of observing values equal or more extreme than 1 in the sample of coefficients coming from the simulated random dates. As the expected value of uniformly distributed dates is the mid-point between the last HIV-negative and first HIV-positive test result, the expected value of the coefficients under this simulation scheme would converge to the mid-point analysis coefficients, which is an unbiased estimate of the sero-conversion date when the sero-conversion density is constant over the interval. In our sample the average duration between tests is less than 2 years (652 days).

†Adjusted for age, sex, community-level HIV prevalence, area of residence, marital status, number of partners in the last 12 months, and household wealth.

**Table S6.** Results of the final model (Table 1) using a mid-point date of sero-conversion assumption.

| Covariate | aHR* (95% CI) | p-value |
|---|---|---|
| **ART coverage** (vs <10%) | - | - |
| 10-20% | 1.00 (0.86-1.16) | 0.987 |
| 20-30% | 0.71 (0.6-0.85) | <0.001 |
| 30-40% | 0.66 (0.55-0.8) | <0.001 |
| >40% | 0.45 (0.31-0.67) | <0.001 |

aHR = adjusted hazard ratio, CI = confidence interval.

*****Adjusted for age, sex, community-level HIV prevalence, area of residence, marital status, number of partners in the last 12 months, and household wealth.

*The potential role of the size and shape of the kernel on the results of the statistical analysis*

To assess the potential role of the size and shape of the "virtual community" on the results of the statistical analysis, we performed a set of parallel statistical analyses for kernels of different sizes and shapes (Table S7).

**Table S7.** Sensitivity analysis of the effect of size and shape of the kernel (used to derive community-level ART coverage and HIV prevalence) on an HIV-uninfected individual's hazard of acquisition of infection (N=16,667).

| **ART coverage** (vs <10%) | **aHR (95% CI)** | **p-value** |
|---|---|---|
| **Gaussian 3 km†** (SD1) | | |
| *Unadjusted* | | |
| 10-20% | 0.85 (0.74-0.98) | 0.030 |
| 20-30% | 0.72 (0.62-0.83) | <0.0001 |
| 30-40% | 0.51 (0.42-0.63) | <0.0001 |
| >40% | 0.40 (0.28-0.58) | <0.0001 |
| *Adjusted** | | |
| 10-20% | 0.93 (0.8-1.08) | 0.325 |
| 20-30% | 0.79 (0.67-0.92) | 0.003 |
| 30-40% | 0.62 (0.50-0.76) | <0.001 |
| >40% | 0.63 (0.44-0.91) | 0.013 |
| **Flat 3 km†** | | |
| *Unadjusted* | | |
| 10-20% | 0.88 (0.76-1.02) | 0.098 |
| 20-30% | 0.71 (0.61-0.82) | <0.0001 |
| 30-40% | 0.55 (0.46-0.66) | <0.0001 |
| >40% | 0.34 (0.19-0.62) | <0.0001 |
| *Adjusted* | | |
| 10-20% | 0.97 (0.83-1.13) | 0.716 |
| 20-30% | 0.82 (0.70-0.96) | 0.015 |
| 30-40% | 0.68 (0.56-0.84) | <0.0001 |
| >40% | 0.52 (0.28-0.94) | 0.032 |
| **Gaussian 2.5 km‡** (SD1) | | |
| *Unadjusted* | | |
| 10-20% | 0.82 (0.71-0.95) | 0.008 |
| 20-30% | 0.73 (0.63-0.84) | <0.0001 |
| 30-40% | 0.50 (0.41-0.61) | <0.0001 |
| >40% | 0.41 (0.30-0.55) | <0.0001 |
| *Adjusted* | | |
| 10-20% | 0.92 (0.79-1.07) | 0.281 |
| 20-30% | 0.81 (0.69-0.95) | 0.009 |
| 30-40% | 0.62 (0.50-0.77) | <0.0001 |
| >40% | 0.65 (0.48-0.88) | 0.006 |

| | | |
|---|---|---|
| **Gaussian 3 km†** (SD2) | | |
| *Unadjusted* | | |
| 10-20% | 0.90 (0.78-1.04) | 0.165 |
| 20-30% | 0.69 (0.60-0.79) | <0.0001 |
| 30-40% | 0.55 (0.45-0.67) | <0.0001 |
| >40% | 0.40 (0.24-0.68) | 0.001 |
| *Adjusted* | | |
| 10-20% | 0.99 (0.85-1.15) | 0.905 |
| 20-30% | 0.79 (0.67-0.92) | 0.003 |
| 30-40% | 0.68 (0.57-0.84) | <0.0001 |
| >40% | 0.61 (0.36-1.03) | 0.063 |
| **Gaussian 3.8 km§** (SD1) | | |
| *Unadjusted* | | |
| 10-20% | 0.91 (0.79-1.05) | 1.193 |
| 20-30% | 0.70 (0.61-0.81) | <0.0001 |
| 30-40% | 0.58 (0.48-0.70) | <0.0001 |
| >40% | 0.42 (0.30-0.61) | <0.0001 |
| *Adjusted* | | |
| 10-20% | 1.00 (0.86-1.16) | 0.964 |
| 20-30% | 0.78 (0.67-0.92) | 0.003 |
| 30-40% | 0.71 (0.58-0.86) | 0.001 |
| >40% | 0.65 (0.45-0.94) | 0.022 |

aHR = adjusted hazard ratio, CI = confidence interval. SD=standard deviation.

*Adjusted for age, sex, community-level HIV prevalence, area of residence, marital status, number of partners in the last 12 months, and household wealth (Table 1).

† A median of 261 (IQR 100-652) individuals of known HIV sero-status were evaluated in the community around each HIV-uninfected individual (2004-2011).

‡ A median of 186 (IQR 71-543) individuals of known HIV sero-status were evaluated in the community around each HIV-uninfected individual (2004-2011).

§ A median of 430 (IQR 164-789) individuals of known HIV sero-status were evaluated in the community around each HIV-uninfected individual (2004-2011).

*The potential influence of uncertainty regarding other survival model specifications*

The Weibull specification model is preferred for our primary survival analysis, because the hypothesis that the shape parameter $\alpha=0$ is rejected (p-value=0.003). However, estimates from parametric survival models can be sensitive to choice of the form of the baseline survival function. For comparative purposes the results of the exponential model are given (Table S8).

**Table S8:** Results of the final model (Table 1) using an exponential survival model (N=16,667).

| Covariate | aHR (95% CI) | p-value |
|---|---|---|
| ***Community-level**** | | |
| **ART coverage** (vs <10%) | - | - |
| 10-20% | 0.99 (0.86-1.14) | 0.926 |
| 20-30% | 0.86 (0.74-1.00) | 0.046 |
| 30-40% | 0.68 (0.56-0.82) | <0.0001 |
| >40% | 0.69 (0.48-0.99) | 0.045 |
| **HIV prevalence** (vs <10%) | - | - |
| 10-15% | 1.41 (1.02-1.96) | 0.039 |
| 15-20% | 1.68 (1.23-2.31) | 0.001 |
| 20-25% | 1.99 (1.44-2.77) | <0.0001 |
| >25% | 2.20 (1.54-3.13) | <0.0001 |
| | | |
| ***Individual-level*** | | |
| **Age-Sex groups** (vs male 15-19 years old) | - | - |
| Male 20-24 | 3.19 (2.26-4.48) | <0.0001 |
| Male 25-29 | 5.48 (3.78-7.94) | <0.0001 |
| Male 30-34 | 3.82 (2.32-6.29) | <0.0001 |
| Male 35-39 | 4.53 (2.80-7.32) | <0.0001 |
| Male 40-44 | 2.59 (1.42-4.72) | 0.002 |
| Male >=45 | 1.77 (1.14-2.73) | 0.011 |
| Female 15-19 | 6.34 (4.66-8.62) | <0.0001 |
| Female 20-24 | 9.40 (6.93-12.75) | <0.0001 |
| Female 25-29 | 8.16 (5.83-11.42) | <0.0001 |
| Female 30-34 | 5.64 (3.84-8.28) | <0.0001 |
| Female 35-39 | 3.89 (2.59-5.85) | <0.0001 |
| Female 40-44 | 3.82 (2.58-5.66) | <0.0001 |
| Female >=45 | 1.17 (0.80-1.71) | 0.410 |
| **Area of residence** (vs rural) | - | - |
| Peri-urban | 1.13 (0.97-1.3) | 0.108 |

| | | |
|---|---|---|
| Urban | 0.91 (0.63-1.32) | 0.628 |
| **Marital status** (vs single) | - | - |
| Married monogamous | 0.67 (0.53-0.84) | 0.001 |
| Married polygamous | 1.13 (0.65-1.99) | 0.661 |
| **Number of partners in the last 12 months(vs 0)** | - | - |
| 1 | 1.79 (1.18-2.7) | 0.007 |
| >1 | 2.52 (1.63-3.9) | <0.0001 |
| **Wealth quintile** (vs poorest) | - | - |
| 2nd poorest | 0.96 (0.82-1.14) | 0.670 |
| 3rd poorest | 0.97 (0.82-1.14) | 0.693 |
| 4th poorest | 0.92 (0.78-1.08) | 0.293 |
| Wealthiest | 0.83 (0.69-1.00) | 0.047 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

*Complete case analysis and restriction of analysis according to original (2004-2006) age eligibility criteria*

**Table S9:** Results of the final model (Table 1) restricting the analysis to include only complete cases (N=11,376).

| Covariate | aHR (95% CI) | p-value | Joint p-value* |
|---|---|---|---|
| ***Community-level**** | | | |
| **ART coverage** (vs <10%) | | | |
| 10-20% | 0.96 (0.74-1.26) | 0.792 | - |
| 20-30% | 0.83 (0.62-1.11) | 0.217 | - |
| 30-40% | 0.64 (0.46-0.89) | 0.008 | - |
| >40% | 0.64 (0.4-1.03) | 0.066 | 0.0074 |
| **HIV prevalence** (vs <10%) | | | |
| 10-15% | 1.31 (0.88-1.95) | 0.179 | - |
| 15-20% | 1.51 (1.03-2.20) | 0.035 | - |
| 20-25% | 1.76 (1.19-2.60) | 0.005 | - |
| >25% | 1.88 (1.23-2.89) | 0.004 | <0.0001 |
| ***Individual-level*** | | | |
| **Age-Sex groups** (vs male 15-19 years old) | | | |
| Male 20-24 | 2.74 (1.69-4.44) | <0.0001 | - |
| Male 25-29 | 4.93 (2.99-8.14) | <0.0001 | - |
| Male 30-34 | 2.87 (1.47-5.63) | 0.002 | - |
| Male 35-39 | 4.02 (2.15-7.52) | <0.0001 | - |
| Male 40-44 | 2.12 (0.97-4.64) | 0.059 | - |
| Male >=45 | 1.76 (0.97-3.18) | 0.062 | - |
| Female 15-19 | 7.79 (4.98-12.19) | <0.0001 | - |
| Female 20-24 | 9.1 (5.84-14.18) | <0.0001 | - |
| Female 25-29 | 7.1 (4.43-11.39) | <0.0001 | - |
| Female 30-34 | 5.59 (3.35-9.34) | <0.0001 | - |
| Female 35-39 | 2.86 (1.62-5.05) | <0.0001 | - |
| Female 40-44 | 3.41 (1.99-5.84) | <0.0001 | - |
| Female >=45 | 1.55 (0.91-2.65) | 0.107 | <0.0001 |
| **Area of residence** (vs rural) | | | |
| Peri-urban | 1.05 (0.89-1.25) | 0.557 | - |
| Urban | 1.00 (0.66-1.51) | 0.994 | 0.3759 |
| **Marital status** (vs single) | | | |
| Married monogamous | 0.61 (0.46-0.82) | 0.001 | - |
| Married polygamous | 1.13 (0.55-2.32) | 0.739 | 0.0015 |
| **Number of partners in the last 12 months** (vs 0) | | | |
| 1 | 2.22 (1.43-3.42) | <0.0001 | - |
| >1 | 3.32 (2.04-5.38) | <0.0001 | <0.0001 |

| | | | |
|---|---|---|---|
| **Wealth quintile** (vs poorest) | | | |
| 2nd poorest | 0.94 (0.77-1.14) | 0.501 | - |
| 3rd poorest | 0.89 (0.73-1.08) | 0.247 | - |
| 4th poorest | 0.89 (0.73-1.08) | 0.224 | - |
| Wealthiest | 0.83 (0.67-1.04) | 0.108 | 0.5033 |
| **Weibull shape parameter** | 1.06 (0.98-1.14) | 0.171 | |

aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

**Table S10:** Results of the final model (Table 1) restricting the analysis to include only those individuals eligible according to age eligibility criteria used in the HIV surveys between 2004 and 2006 (Females 15-49, males 15-54) - N=13,654.

| Covariate | aHR (95% CI) | p-value |
|---|---|---|
| ***Community-level***†* | | |
| **ART coverage** | | |
| (vs <10%) | | |
| 10-20% | 0.98 (0.84-1.13) | 0.742 |
| 20-30% | 0.84 (0.71-0.98) | 0.031 |
| 30-40% | 0.63 (0.51-0.78) | <0.0001 |
| >40% | 0.71 (0.49-1.03) | 0.075 |
| **HIV prevalence** | | |
| (vs <10%) | | |
| 10-15% | 1.51 (1.08-2.11) | 0.017 |
| 15-20% | 1.78 (1.29-2.47) | 0.001 |
| 20-25% | 2.14 (1.52-3) | <0.0001 |
| >25% | 2.36 (1.64-3.4) | <0.0001 |
| | | |
| ***Individual-level*** | | |
| **Age-Sex strata** | | |
| (vs male 15-19 years old) | | |
| Male 20-24 | 3.05 (2.16-4.3) | <0.0001 |
| Male 25-29 | 5.24 (3.61-7.62) | <0.0001 |
| Male 30-34 | 3.7 (2.24-6.09) | <0.0001 |
| Male 35-39 | 4.36 (2.69-7.05) | <0.0001 |
| Male 40-44 | 2.49 (1.37-4.53) | 0.003 |
| Male >=45 | 2.42 (1.52-3.85) | <0.0001 |
| Female 15-19 | 6.37 (4.69-8.67) | <0.0001 |
| Female 20-24 | 9.04 (6.65-12.29) | <0.0001 |
| Female 25-29 | 7.81 (5.57-10.96) | <0.0001 |
| Female 30-34 | 5.41 (3.68-7.96) | <0.0001 |
| Female 35-39 | 3.71 (2.46-5.6) | <0.0001 |
| Female 40-44 | 3.61 (2.43-5.37) | <0.0001 |
| Female >=45 | 1.98 (1.27-3.08) | 0.002 |
| **Area of residence** | | |
| (vs rural) | | |
| Peri-urban | 1.11 (0.96-1.29) | 0.166 |
| Urban | 0.93 (0.64-1.34) | 0.686 |
| **Marital status** | | |
| (vs single) | | |
| Married monogamous | 0.68 (0.53-0.87) | 0.002 |
| Married polygamous | 1.37 (0.74-2.53) | 0.311 |
| **Number of partners in the last 12 months** | | |
| (vs zero) | | |
| One | 1.72 (1.16-2.53) | 0.007 |
| More than one | 2.44 (1.61-3.7) | <0.0001 |
| **Wealth quintile** | | |
| (vs poorest) | | |
| 2nd poorest | 0.96 (0.81-1.13) | 0.612 |
| 3rd poorest | 0.95 (0.81-1.13) | 0.568 |
| 4th poorest | 0.93 (0.78-1.09) | 0.366 |
| Wealthiest | 0.85 (0.7-1.03) | 0.090 |
| **Weibull shape parameter** | 2.89 (2.73-3.07) | 0.033 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) around each HIV-uninfected individual in the population cohort.

†Corresponding values for a model, in which community-level ART coverage and HIV prevalence are modeled as continuous covariates (%); ART coverage aHR = 0.988 (95% CI 0.983-0.993), $p<0.0001$; HIV prevalence aHR = 1.038 (95% CI 1.027 -1 .051), $p<0.0001$.

*Isigodi fixed-effects analysis*

To ascertain whether an alternative (but less sensitive) approach would also demonstrate a robust relationship between ART coverage and risk of acquisition of infection we ran a fixed-effects analysis by aggregating HIV prevalence and ART coverage by the 23 traditional Zulu areas (*izigodi*) in the study area. We then reran our final model (Table 1) using the community-level variables created in this way and included them in the model as time varying covariates. This fixed-effects analysis controls for all observed and unobserved time-invariant *isigodi*-specific factors and solves the problem of correlation with other predictors as opposed to random effects (Table S11).

**Table S11:** *Isigodi* fixed-effects analysis with *isigodi*-specific ART coverage and HIV prevalence (N=16,667).

| Covariate | aHR† (95% CI) | p-value |
|---|---|---|
| ***Community-level*‡** | | |
| **ART coverage** (vs <10%) | - | - |
| 10-20% | 1.09 (0.92-1.28) | 0.329 |
| 20-30% | 0.88 (0.75-1.03) | 0.111 |
| >30% | 0.73 (0.61-0.88) | 0.001 |
| **HIV prevalence** (vs <15%) | - | - |
| 15-20% | 1.21 (0.94-1.55) | 0.136 |
| 20-25% | 1.13 (0.83-1.54) | 0.448 |
| >25% | 1.37 (0.95-1.99) | 0.096 |
| | | |
| ***Individual-level*** | | |
| **Age-Sex groups** (vs male 15-19 years old) | - | - |
| Male 20-24 | 3.07 (2.18-4.32) | <0.0001 |
| Male 25-29 | 5.32 (3.66-7.73) | <0.0001 |
| Male 30-34 | 3.65 (2.22-6.02) | <0.0001 |
| Male 35-39 | 4.34 (2.68-7.02) | <0.0001 |
| Male 40-44 | 2.48 (1.36-4.52) | 0.003 |
| Male >=45 | 1.72 (1.11-2.66) | 0.015 |
| Female 15-19 | 6.38 (4.69-8.67) | <0.0001 |
| Female 20-24 | 9.00 (6.62-12.23) | <0.0001 |
| Female 25-29 | 7.8 (5.56-10.93) | <0.0001 |
| Female 30-34 | 5.43 (3.69-7.99) | <0.0001 |
| Female 35-39 | 3.70 (2.46-5.57) | <0.0001 |

| | | |
|---|---|---|
| Female 40-44 | 3.63 (2.45-5.39) | <0.0001 |
| Female >=45 | 1.16 (0.79-1.69) | 0.454 |
| **Area of residence** (vs rural) | - | - |
| Peri-urban | 1.07 (0.88-1.31) | 0.475 |
| Urban | 0.32 (0.04-2.47) | 0.271 |
| **Marital status** (vs single) | - | - |
| Married monogamous | 0.67 (0.53-0.84) | 0.001 |
| Married polygamous | 1.13 (0.64-1.99) | 0.677 |
| **Number of partners in the last 12 months** (vs 0) | - | - |
| 1 | 1.78 (1.18-2.67) | 0.007 |
| >1 | 2.51 (1.62-3.87) | <0.0001 |
| **Wealth quintile** (vs poorest) | - | - |
| 2nd poorest | 0.95 (0.8-1.12) | 0.510 |
| 3rd poorest | 0.95 (0.81-1.12) | 0.566 |
| 4th poorest | 0.91 (0.77-1.07) | 0.262 |
| Wealthiest | 0.84 (0.7-1.02) | 0.082 |
| **Weibull shape parameter** | 2.94 (2.78-3.12) | 0.006 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Calculated by aggregating HIV prevalence and ART coverage by 23 *izigodi.* Aggregating by izgodi resulted in a smaller range in the resulting community-level variables in comparison to the Gaussian kernel approach used in the main analysis. Hence, different categories of community-level ART coverage and HIV prevalence were used in this analysis.

†HR adjusted for *isigodi* fixed-effects.

‡ Corresponding values for a model, in which community-level ART coverage and HIV prevalence are modeled as continuous covariates (%); ART coverage aHR = 0.990 (95% CI 0.985-0.996), p<0.0001; HIV prevalence aHR = 1.030 (95% CI 1.017 -1 .043), p<0.0001.

*The possibility of selection effects influencing the analysis*

We have recently shown that imputed HIV prevalence in the Africa Centre demographic surveillance area does not vary by more than 1 percentage point from the complete-case HIV prevalence results for any of the years (2004-2011), which we used in this study (*19*). However, to assess the possibility of selection effects influencing the analysis, we produced estimates of HIV status for all eligible individuals using multiple imputation with chained equations (*20, 21*). For each study year, non-response for HIV status was imputed using information on age, sex, household wealth, employment status, local area, highest educational attainment, and community-level HIV prevalence. The covariates used to impute HIV status had near-complete response. As HIV status is a binary outcome, we used a logistic regression specification to specify the relationship between HIV status and the covariates outlined above (*21*). Since the survey only started implementing HIV testing for women 50 years and above and men 55 years and above after 2006, the imputation procedure was carried out separately for these age strata. We produced 100 complete datasets through imputing for HIV status in each year of study. This analysis was conducted in R (version 2.14.2) with the MICE package (*20*). We then analysed spatial variation in the resulting imputed values using the Gaussian kernel to produce the community-level variables HIV prevalence and ART coverage. We then reran our main statistical analysis (as described in the statistical analysis section), using these results (Table S12).

**Table S12:** Results of the final model (Table 1) using HIV status imputed for every individual eligible for HIV testing in each survey round (2004-2011) to generate the community-level covariates.

| Covariate | aHR (95% CI) | p-value |
|---|---|---|
| ***Community-level**** | | |
| **ART coverage** (vs <10%) | - | - |
| 10-20% | 1.00 (0.86-1.17) | 0.959 |
| 20-30% | 0.85 (0.73-0.98) | 0.031 |
| 30-40% | 0.72 (0.59-0.88) | 0.001 |
| >40% | 0.74 (0.54-1.00) | 0.050 |
| **HIV prevalence** (vs <10%) | - | - |
| 10-15% | 1.06 (0.76-1.50) | 0.722 |
| 15-20% | 1.35 (0.97-1.88) | 0.073 |
| 20-25% | 1.48 (1.05-2.09) | 0.026 |
| >25% | 1.73 (1.21-2.50) | 0.003 |
| ***Individual-level*** | | |
| **Age-sex groups** (vs male 15-19 years old) | | |
| Male 20-24 | 3.02 (2.14-4.26) | <0.0001 |
| Male 25-29 | 5.24 (3.61-7.6-) | <0.0001 |
| Male 30-34 | 3.64 (2.21-6.00) | <0.0001 |
| Male 35-39 | 4.33 (2.68-7.00) | <0.0001 |
| Male 40-44 | 2.48 (1.36-4.51) | 0.003 |
| Male >=45 | 1.70 (1.10-2.64) | 0.017 |
| Female 15-19 | 6.35 (4.67-8.64) | <0.0001 |
| Female 20-24 | 8.91 (6.55-12.1) | <0.0001 |
| Female 25-29 | 7.75 (5.53-10.86) | <0.0001 |
| Female 30-34 | 5.39 (3.67-7.93) | <0.0001 |
| Female 35-39 | 3.71 (2.47-5.59) | <0.0001 |
| Female 40-44 | 3.61 (2.44-5.36) | <0.0001 |
| Female >=45 | 1.14 (0.78-1.67) | 0.502 |
| **Area of residence** (vs rural) | - | - |
| Peri-urban | 1.11 (0.95-1.3) | 0.195 |
| Urban | 0.86 (0.59-1.26) | 0.440 |
| **Maritial status** (vs single) | | |
| Married monogamous | 0.67 (0.53-0.84) | 0.001 |
| Married polygamous | 1.14 (0.65-2.01) | 0.645 |
| **Number of partners in the last 12 months** (vs 0) | - | - |
| 1 | 1.77 (1.17-2.67) | 0.008 |
| >1 | 2.49 (1.61-3.85) | <0.0001 |

| | | |
|---|---|---|
| **Wealth quintile** (vs poorest) | | |
| 2nd poorest | 0.96 (0.81-1.13) | 0.622 |
| 3rd poorest | 0.96 (0.82-1.13) | 0.642 |
| 4th poorest | 0.91 (0.77-1.08) | 0.276 |
| Wealthiest | 0.82 (0.68-0.99) | 0.037 |
| **Weibull shape parameter** | 2.94 (2.78-3.12) | 0.006 |

aHR = adjusted hazard ratio, CI = confidence interval.

*Derived using a standard Gaussian kernel (radius 3 km) using imputed HIV status for all eligible individuals.

## References


1. F. Tanser *et al.,* Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *Int J Epidemiol* **37**, 956 (2008).

2. Health Systems Trust, "South African Health Review" (Health Systems Trust, Durban, 2011).

3. Umkhanyakude District Municipality, "Umkhanyakude district municipality integrated development plan -2012/13-2016/17" (Mkuze, 2011).

4. M. N. Lurie *et al.,* The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners. *Sex Transm Dis* **30**, 149 (2003).

5. UNAIDS/WHO, "Guidelines for using HIV testing technologies in surveillance: selection, evaluation, and implementation" (UNAIDS/WHO, Geneva, 2001).

6. F. Tanser, T. Bärnighausen, G. S. Cooke, M. L. Newell, Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int J Epidemiol* **38**, 1008 (2009).

7. F. Tanser *et al.,* Effect of concurrent sexual partnerships on rate of new HIV infections in a high-prevalence, rural South African population: a cohort study. *Lancet* **378**, 247 (2011).

8. L. A. Waller, C. A. Gotway, *Applied Spatial Statistics for Public Health Data*. (Wiley-Interscience, Hoboken, N.J, 2004).

9. T. Bärnighausen *et al.*, High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. *AIDS* **22**, 139 (2008).

10. J. L. Schafer, Multiple imputation: a primer. *Stat Methods Med Res* **8**, 3 (1999).

11. N. E. McGrath, JW, F. Tanser, T. Barnighausen, M. Newell, *Sexual behaviour trends by gender in a rural South African population-based cohort during the era of scaled-up access to VCT and ART, 2005-2010* paper presented at the Presentation at XIX International AIDS Conference, Washington DC, USA, 22-27 July 2012.

12. T. Bärnighausen *et al.*, Structural barriers to ART access: a study of population-based CD4 count and linked programmatic data in rural KwaZulu-Natal. *Submitted*, (2012).

13. M. Q. Ott, T. Bärnighausen, F. Tanser, M. N. Lurie, M. L. Newell, Age-gaps in sexual partnerships: seeing beyond 'sugar daddies'. *AIDS* **25**, 861 (2011).

14. C. F. Houlihan *et al.*, Cohort Profile: Hlabisa HIV Treatment and Care Programme. *International Journal of Epidemiology* **40**, 318 (2011).

15. M. Kulldorff, L. Huang, K. Konty, A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics* **8**, 58 (2009).

16. L. Huang, M. Kulldorff, D. Gregorio, A spatial scan statistic for survival data. *Biometrics* **63**, 109 (2007).

17. M. Kulldorff, R. Heffernan, J. Hartman, R. Assuncao, F. Mostashari, A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* **2**, e59 (2005).

18. I. Jung, M. Kulldorff, O. Richard, A spatial scan statistic for multinomial data. *Manuscript*, (2008).

19. J. Zaidi, E. Grapsa, F. Tanser, M.-L. Newell, T. Bärnighausen, Dramatic increases in HIV prevalence after scale-up of antiretroviral treatment: a longitudinal population-based HIV surveillance study in rural KwaZulu-Natal. ***Submitted***, (2013).

20. S. Van Buuren, K. Groothuis-Oudshoorn, MICE: Multivariable imputation by chained equations in R. *Journal of Statistical Software*, 1 (2010).

21. D. B. Rubin, *Multiple imputation for nonresponse in surveys*. (Wiley Online Library, 1987), vol. 519.