Effect of population viral load on prospective HIV incidence in a hyperendemic rural African community

Frank Tanser, Alain Vandoorne, Diego Cuadros, Andrew N. Phillips, Tulio de Oliveira, Andrew Tomita, Till Bärnighausen, Deenan Pillay

Monitoring HIV population viral load (PVL) has been advocated as an important means of inferring HIV transmission potential and predicting the future rate of new HIV infections (HIV incidence) in a particular community. However, the relationship between PVL measures and directly measured HIV incidence has not been quantified in any setting and, most importantly, in a hyperendemic sub-Saharan African setting. We assessed this relationship using one of Africa’s largest population-based prospective population cohorts in rural KwaZulu-Natal, South Africa in which we followed 8732 HIV-uninfected participants between 2011 and 2015. Despite clear evidence of spatial clustering of high viral loads in some communities, our results demonstrate that PVL metrics derived from aggregation of viral load data only from the HIV-positive members of a particular community did not predict HIV incidence in this typical hyperendemic, rural African population. Only once we used modified PVL measures, which combined viral load information with the underlying spatial variation in the proportion of the population infected (HIV prevalence), did we find a consistently strong relationship with future risk of HIV acquisition. For example, every 1% increase in the overall proportion of a population having detectable virus (PDV) was independently associated with a 6.3% increase in an individual’s risk of HIV acquisition (P = 0.001). In hyperendemic African populations, these modified PVL indices could play a key role in targeting and monitoring interventions in the most vulnerable communities where the future rate of new HIV infections is likely to be highest.

INTRODUCTION

The rapid scale-up of combination antiretroviral therapy (ART) to more than 19 million people with HIV (1) has resulted in substantial population-level reductions in HIV-related mortality and evidence of reductions in the rate of new HIV infections in some developed (2) and developing (3) country contexts. Notwithstanding this impact, the rate of new HIV infections remains unacceptably high, with 70% of new HIV infections continuing to occur in sub-Saharan Africa (4). Some mathematical models suggest that under certain conditions, the HIV epidemic could be reversed by 2050 if high levels of ART coverage are achieved (in combination with other effective interventions such as male circumcision) (5). Scaling up ART therefore remains a key global priority. In response, the Joint United Nations Programme on HIV/AIDS (UNAIDS) has set “the ambitious but achievable” targets of diagnosing 90% of all people living with HIV to initiating 90% of all diagnosed people on ART and achieving virologic suppression for 90% of people on ART by 2020, with the aim of increasing all of these targets to 95% by 2030 (6). As the world ramps up to meet and exceed these targets, it becomes critical to not only monitor progress but also provide empirical evidence for any resultant impact of expanding levels of viral suppression on life expectancy and the rate of new HIV infections at a population level.

One of the ways in which the impact of moving toward the UNAIDS 90-90-90 treatment targets could be monitored is through measuring the trends in HIV population viral load (PVL) (7, 8). The HIV viral load level in semen or blood is the single most important biological determinant of transmission between an HIV-positive and an HIV-negative individual (9, 10). It follows that an aggregation of individual HIV viral RNA concentrations for a particular geography or community over a given time period (viz, PVL) could constitute a sensitive biological index of treatment program success and potentially estimate the "transmission potential" of a particular geography. Population/community viral load has been endorsed as a key group of measures by the Centers for Disease Control and Prevention (CDC) (11) and has been used to infer both the effectiveness of a treatment program and a proxy for HIV incidence (2, 12–14).

Notwithstanding the above, the concept as a measure of both transmission potential and ART program effectiveness has been critically (15). One key issue raised by the authors of the article is the degree to which the in-care or facility-based viral load measures are reflective of the true underlying PVL. Other issues raised include the interpretation of various PVL metrics and their relationship to ongoing HIV transmissions, ecologic bias due to aggregation over large geographic regions or heterogeneous populations, and the failure of some PVL measures to account for the underlying HIV prevalence within the community (15). To date, no study has tested the relationship between these PVL measures and directly measured HIV incidence in a hyperendemic sub-Saharan African setting.

Here, we make use of one of Africa’s largest population-based prospective cohorts, which is located in the rural KwaZulu-Natal province of South Africa, to empirically quantify the relationship between the true PVL (among all HIV-infected participants irrespective of whether these participants have accessed care or know their HIV status) and the prospective risk of HIV acquisition for more than 8700 participants...
who were HIV-uninfected at baseline (HIV incidence). We use novel geospatial techniques to analyze the micro-level spatial variation in three key PVL measures—called the geometric mean viral load (MVL), prevalence of detectable viremia (PDV), and community transmission index (CTI)—derived from all HIV-positive participants. We then extend these three measures not only to consider HIV-positive participants but also to evaluate the entire population (irrespective of HIV status).

RESULTS

Population-based viral load survey

This study uses data from one of the most comprehensive demographic and HIV surveillance sites in Africa—the Africa Centre Demographic Information System (16). The site has collected sociodemographic information on a population of 87,000 participants within a circumscribed geographic area (438 km²) for more than a decade. Within the study site, ongoing population-based HIV surveillance and sexual behavior surveys take place annually among all adults ≥15 years of age. Of those contacted during the 2011 survey, 78.6% agreed to be tested for HIV in the survey and provide a dried blood spot (DBS) sample (3). We performed viral load measurements on the blood spots of all 2456 participants who tested HIV-positive in the population-based HIV surveillance round of 2011. Using the Generic HIV Viral Load kit (Biocentric), we successfully obtained 2420 (98%) viral load measurements after nucleic acid extraction. The extraction method has a lower detection limit of 1550 copies/ml and is described in greater detail elsewhere (17).

The geometric mean population-based viral load [95% confidence interval (CI)] among HIV-infected people was 8259 copies/ml (7507 to 9087 copies/ml) (Table 1). Of the 2420 HIV-positive participants in the population-based cohort, 30% (n = 726) had undetectable viral loads, and 26% (n = 629) had a viral load of >50,000 copies/ml. Males had overall higher geometric mean viral loads and a higher proportion of males were characterized by viral loads of >50,000 copies/ml (Fig. 1, A and B).

Table 1. Summary statistics: Viral load measurements by sex and age for the population-based survey data (2011). CI, confidence interval.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Samples (N)</th>
<th>Geometric mean</th>
<th>&gt;50,000 copies/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>2420</td>
<td>8,259 (7,507–9,087)</td>
<td>0.26 (0.24–0.27)</td>
</tr>
<tr>
<td>Male</td>
<td>506 (20.91)</td>
<td>12,802 (10,352–15,832)</td>
<td>0.34 (0.3–0.38)</td>
</tr>
<tr>
<td>Female</td>
<td>1,914 (79.09)</td>
<td>7,356 (6,614–8,181)</td>
<td>0.24 (0.22–0.26)</td>
</tr>
<tr>
<td>15−</td>
<td>98 (4.05)</td>
<td>25,139 (15,934–39,660)</td>
<td>0.46 (0.36–0.56)</td>
</tr>
<tr>
<td>20−</td>
<td>295 (12.19)</td>
<td>13,508 (10,376–17,586)</td>
<td>0.29 (0.24–0.35)</td>
</tr>
<tr>
<td>25−</td>
<td>405 (16.74)</td>
<td>10,815 (8,606–13,592)</td>
<td>0.28 (0.24–0.33)</td>
</tr>
<tr>
<td>30−</td>
<td>360 (14.88)</td>
<td>7,766 (6,026–10,008)</td>
<td>0.22 (0.22–0.31)</td>
</tr>
<tr>
<td>35−</td>
<td>353 (14.59)</td>
<td>8,460 (6,529–10,963)</td>
<td>0.28 (0.23–0.32)</td>
</tr>
<tr>
<td>40−</td>
<td>243 (10.04)</td>
<td>6,823 (5,026–9,262)</td>
<td>0.23 (0.18–0.29)</td>
</tr>
<tr>
<td>45+</td>
<td>666 (27.52)</td>
<td>5,238 (4,397–6,240)</td>
<td>0.19 (0.16–0.22)</td>
</tr>
</tbody>
</table>

Fig. 1. Age-sex differences in viral load patterns in the 2011 population-based survey. Geometric mean viral load (A) and proportion of viral loads (>50,000 copies/ml) (B). Estimates are shown with 95% confidence intervals.
Spatial variation in PVL indices

We used the viral load measurements of HIV-positive participants \( n = 2420 \) from the 2011 survey to construct the following PVL indices across the study area: (i) the MVL, which is the geometric mean of the viral loads among all HIV-positive participants in the population-based survey; (ii) the PDV, which represents the proportion of the HIV-positive population that has a detectable viral load \( (>1550 \text{ copies/ml}) \) (17); and (iii) the CTI, which is a potentially more sensitive biological measure than (i) and (ii) and represents the relation between the viral load and the risk of HIV transmission per unprotected sexual contact (17). To derive this measure, we used the result from Quinn et al. (9), who showed that each \( \log_{10} \) increment in viral load was associated with a 2.45-fold increase in the rate ratio for HIV transmission risk. Following the notation of Wilson et al. (18), the risk of HIV transmission per sexual act is therefore given by

\[
\beta_1 = 2.45^\log_{10}(V_1/V_0) \beta_0,
\]

where \( V_1 \) is the viral load level associated with the participant, \( V_0 = \log_{10}(150) \) is a baseline viral load level, and \( \beta_0 = 0.003 \) is the probability of HIV transmission from a person with the baseline viral load level \( (V_0) \) (18). The \( \beta_0 \) value is based on a prior research undertaken in low-income settings (19).

Using the standard binomial formula, we then calculated CTI = \( 1 - [1 - \beta_1]^{100} \times 100 \) to obtain the estimated number of transmission events that occur in 100 sexual contacts between persons at risk of infection and HIV-positive members of a particular community. We standardized all of the PVL measures against the age (10-year bands) and sex characteristics of the eligible population in 2011.

Next, we constructed population-based versions of the three PVL indices described above on the basis of the population in each community irrespective of HIV status. We use “\( P \)” (in subscript) to denote the fact that these modified versions of the PVL indices are based on information from the full population (irrespective of HIV status). In effect, all HIV-negative participants in a particular community are included by assigning a viral load value of zero (to construct the MVL\(_P\), PDV\(_P\), and CTI\(_P\) indices) or a transmission probability of zero (to construct the CTI\(_P\) index). Thus, for these analyses, we included 7919 individuals who tested HIV-negative in 2011 (total tested, 10,375). Hence, the denominator of these modified indices becomes the entire population rather than only those who have tested HIV-positive. By including information on the number of HIV-negative participants in a particular community, the modified PVL measures inherently account for the underlying spatial variation in HIV prevalence.

Marked spatial heterogeneity in all of the PVL measures (MVL, PDV, and CTI) was observed across the surveillance area. There was clear evidence of spatial clustering of participants with high geometric MVL (Fig. 2A) and proportion of participants with detectable viral loads (Fig. 2B) as well as the estimated number of transmission events per 100 sexual contacts between individuals at risk of infection and HIV-infected members of a particular community (Fig. 2C). Likewise, clear evidence for spatial clustering of the modified PVL measures (MVL\(_P\), PDV\(_P\), and CTI\(_P\)) was observed when we considered the entire adult population (Fig. 2, D to F) in the calculation. Remarkably, in some...
of the high-incidence communities, although 65% of HIV-positive individuals had unsuppressed viral loads (Fig. 2B), >20% of the entire population (irrespective of HIV status, that is, PDV$_p$) was viremic for HIV (Fig. 2E), underscoring the scale of the epidemic in these areas.

### Relationship between the PVL measures and the risk of HIV acquisition

We followed up all 8732 repeat-testers (males aged 15 to 54 and females aged 15 to 49) who were resident within the surveillance area between January 2011 and December 2015. During this period, we observed 859 seroconversions and 26,219 person-years of observation in the 8732 repeat-testers who were HIV-uninfected at baseline (crude incidence, 3.28 events per 100 person-years; Table 2). Our aim was to quantify the relationship between the PVL measures, which are derived from the HIV-positive population, and the prospective risk of HIV acquisition using a Cox proportional hazard model. We then compared these results against the set of PVL indices derived from the entire adult population (irrespective of HIV status).

At an ecological level, communities with the overall highest viral loads and highest levels of unsuppressed viral loads were not characterized by the highest prospective crude HIV incidence (Fig. 3, A to C). However, with respect to the measures constructed on the basis of those who were HIV-infected and those who were HIV-uninfected, there was a clear dose-response pattern evident with communities with the highest overall viral loads across the whole population (irrespective of HIV status) having the highest crude HIV incidence (Fig. 3, D to F).

Results for the individual-level Cox proportional hazard models demonstrated that none of the PVL measures derived from the HIV-positive population predicted future individual HIV acquisition risk, both before and after controlling for key socioeconomic and behavioral factors, as well as HIV prevalence in the surrounding local community [adjusted hazard ratio (aHR), 1.000 to 1.048; factors, as well as HIV prevalence in the surrounding local community both before and after controlling for key socioeconomic and behavioral factors (Table 3, model A1 and A2)]. In addition, these indices were robust to the addition of any of the modified PVL metrics that were calculated on the population (irrespective of HIV status) having the highest prospective crude HIV incidence (Fig. 3, A to C).

Results for the individual-level Cox proportional hazard models demonstrated that none of the PVL measures derived from the HIV-positive population predicted future individual HIV acquisition risk, both before and after controlling for key socioeconomic and behavioral factors (Table 3, models B1 and B2). In addition, these indices were robust to the addition of any of the modified PVL metrics that were calculated on the population (irrespective of HIV status) having the highest prospective crude HIV incidence (Fig. 3, A to C).
and living in a household with higher wealth were protective against HIV acquisition (3).

**Sex-specific PVL patterns and risk of acquisition of infection in the opposite sex**

As a further robustness check of our findings, we constructed sex-specific PVL indices in the same manner. We then conducted a set of parallel analyses to ascertain the degree to which a female’s HIV acquisition hazard was related to viral load patterns (and HIV prevalence) of men in the surrounding local community and vice versa. The ability to detect any PVL effect on HIV acquisition risk is markedly attenuated in these analyses (because of smaller overall observation time and numbers of events) and the fact that the PVL measures are subject to more random noise, given the smaller number of observations used in their construction. Nevertheless, this suite of analyses also confirmed the same pattern described in the mixed sex analyses above. In these models, the PVL measures of the HIV-positive population was similarly not associated with HIV acquisition risk in the opposite sex both before and

---

**Fig. 3.** Graphs showing the ecological relationship between HIV incidence (2011 to 2015) and PVL quartiles derived from the 2011 population-based viral load survey (quartile 1, communities with lowest PVL values). The PVL indices are derived from HIV-positive participants only (A to C) and HIV-positive and HIV-negative participants (D to F). Incidence estimates are shown with 95% confidence intervals.
after adjustment for key socioeconomic, behavioral risk factors, and HIV prevalence in the surrounding local community (P values ranged from 0.18 to 0.99; see Tables 4 and 5 and tables S3 and S4). By contrast, all of the modified PVL measures that included information on the HIV-negative population were strongly predictive of acquisition risk in the opposite sex (P values ranged from <0.001 to 0.048; see also tables S5 and S6).

PVL measures derived from routine data collected at health facilities

Because facility-based viral load data are typically used to derive community viral load metrics in some settings, we also wanted to establish the degree to which facility-based viral load measurements corresponded with their population-based counterparts and whether such data could be harnessed to infer transmission potential. We used the viral load measurements from routine clinical data collected at the facility level for 3196 patients living in the surveillance population who visited 1 of the 17 health care clinics in 2011 to compute the facility-level for 3196 patients living in the surveillance population who

Results show that the overall geometric mean (95% CI) viral load for the facility-based data was 819 copies/ml (771 to 870 copies/ml) (table S7). Of the 3196 HIV-positive participants in the facility-based cohort, 87% (n = 2789) had undetectable viral loads, and 4% (n = 133) had a viral load of >50,000 copies/ml. Males had overall higher geometric mean viral loads and a higher proportion of males were characterized by viral loads of >50,000 copies/ml (fig. S1).

Use of routine facility–based viral load data (20) to construct the viral load indices demonstrated that there was little or no correlation between these indices and those derived using the equivalent population-based viral load data (correlation coefficient, −0.089, −0.32, and −0.28 for MVL, PDV, and CTI, respectively). Empirically, there was no relationship found between any of the facility-based viral load measures and future risk of acquisition of HIV infection both before and after adjustment for key risk factors, as well as HIV prevalence in the surrounding local community (table S8).

DISCUSSION

Our study has empirically tested the relationship between various PVL metrics and the prospective risk of HIV acquisition in a typical hyperendemic rural African setting. Despite identifying remarkable spatial variation in viral load patterns across the study area, with clear evidence of clustering of high viral loads in some communities, we did not find a relationship between any of the PVL indices derived solely from HIV-infected participants and prospective HIV incidence. Our findings occurred in the context of using population-based viral load measurements that were largely free from the biases typically associated with facility–based viral load data. Only when we evaluated the viral load information as a function of the entire population (both HIV-positive and HIV-negative participants) was the resulting measures highly predictive of risk of new HIV infection. This pattern also held in a series of parallel analyses in which we tested the relationship between the sex-specific PVL patterns and the HIV acquisition risk in the opposite sex. Furthermore, our results demonstrate that in this rural African setting, facility-based viral load measures do not correlate well with their population-based counterparts and were not predictive of HIV incidence. On the basis of these findings, we would therefore caution against using such data to infer transmission potential in similar settings.

Another important finding to emerge from the analyses is the large differences in HIV viral load by sex at a population level. Overall, males had a substantially higher proportion with viral loads of >50,000 copies/ml (34% versus 24%) and higher geometric mean viral load (12,802 copies/ml versus 7,356 copies/ml) in comparison to females. This finding likely reflects the fact that men are less likely to be tested for HIV and successfully
Table 4. Results of the multivariable analysis (Cox proportional hazard model) to examine the relationship between the risk of HIV acquisition for females and the three male PVL measures. The PVL measures were derived from the HIV-positive males (models A1 to A3) and the HIV-positive and HIV-negative males (models B1 to B3) of a population-based survey. The full output is given in tables S3 and S5. Model 1 shows the unadjusted HRs for the PVL measures; model 2 shows these HRs after adjusting for age, sex, urban status, marital status, number of sexual partners in the last year, and household wealth; model 3 shows these HRs after adjusting for the model 2 covariates as well as HIV prevalence.

<table>
<thead>
<tr>
<th>HIV acquisition risk for females</th>
<th>Geometric mean viral load*</th>
<th>Prevalence detectable viremia†</th>
<th>Community transmission index‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
<td>HR (95% CI)</td>
</tr>
<tr>
<td>Population-based: HIV-positive males only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A1: Unadjusted HR</td>
<td>1.000 (1.000–1.000)</td>
<td>0.393</td>
<td>1.000 (0.994–1.006)</td>
</tr>
<tr>
<td>Model A2: Adjusted HR without HIV prevalence</td>
<td>1.000 (1.000–1.000)</td>
<td>0.836</td>
<td>1.001 (0.995–1.008)</td>
</tr>
<tr>
<td>Model A3: Adjusted HR with HIV prevalence</td>
<td>1.000 (1.000–1.000)</td>
<td>0.487</td>
<td>1.003 (0.996–1.011)</td>
</tr>
<tr>
<td>Population-based: HIV-positive and HIV-negative males</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model B1: Unadjusted HR</td>
<td>1.066 (1.021–1.113)</td>
<td>0.004</td>
<td>1.039 (1.016–1.063)</td>
</tr>
<tr>
<td>Model B2: Adjusted HR without HIV prevalence</td>
<td>1.056 (1.006–1.108)</td>
<td>0.027</td>
<td>1.039 (1.013–1.066)</td>
</tr>
<tr>
<td>Model B3: Adjusted HR with HIV prevalence</td>
<td>1.160 (1.058–1.271)</td>
<td>0.001</td>
<td>1.061 (1.020–1.104)</td>
</tr>
<tr>
<td>N</td>
<td>5,188</td>
<td>5,188</td>
<td>5,188</td>
</tr>
</tbody>
</table>

*For a 1% increase in geometric mean viral load. †For a 1% increase in the prevalence of detectable viremia. ‡For a predicted one transmission event increase per 100 sexual contacts.

link to HIV care, as well as being less likely to successfully adhere to treatment, supporting previous findings in this setting and others (21–23).

Community viral load has been used as a proxy for HIV transmission potential and/or program effectiveness in many developed (2, 12, 24, 25) and developing (14) country contexts. Previous work has been based on ecological associations between group-level variables and group-level outcomes and, hence, do not provide a strong basis for causal inference (26). Most recently, Solomon and colleagues conducted a cross-sectional study in India, which estimated the site-level correlations between four community viral load measures and an estimate of HIV incidence (derived from a multiassay algorithm) in people who inject drugs and men who have sex with men (20).

Nevertheless, our work has some limitations. Most notably, the link between the PVL in the surrounding local population and HIV acquisition risk is dependent on a reasonable proportion of the population preferentially selecting partners from the surrounding local community. However, if this was not the case, we would not expect any of the community-level PVL indices used in the analyses to significantly influence the risk of HIV acquisition. As indicated above, we find strong independent relationships between all of the modified PVL measures that analyze the entire population (irrespective of HIV status) and risk of HIV acquisition. Moreover, in our previous work, we reported that partner choice is strongly affected by geography with 61% of women reporting at least one partnership with a man in the same immediate local Zulu community over a 5-year period (27). A potential second limitation is that we treat the PVL exposure as being time-invariant. Although unlikely, it is theoretically possible that there could be large systematic geographic shifts in viral load patterns over time that could bias the result (particularly in a more sensitive PVL measure such as the CTI). However, such a “measurement error” would likely bias the finding toward the null hypothesis and would not account for the strong positive associations identified between all of the modified PVL measures and future risk of HIV acquisition. In addition, there are difficulties common to any PVL measure in extrapolating the biological differences seen within-host to the population in a meaningful way. Such difficulties are particularly
salient for sexually transmitted infections that are dependent on highly selective sexual partnership patterns including the phenomenon of HIV “sero-sorting” (28). Thus, large differences in viral load patterns between two populations would be relatively meaningless from a transmission potential perspective if there is a high degree of sero-sorting or if high viral loads occurred mainly in the segment of the population that was no longer sexually active as a result of, for example, HIV-related illness.

In our previous work, we demonstrated substantial space-time differences in ART coverage in this population, as the ART program had scaled up over 8 years (2004 to 2011). These large longitudinal differences in ART coverage independently predicted individual HIV acquisition risk (3). By 2011 (the year in which the viral load survey was conducted), the heterogeneity in ART coverage had decreased markedly (interquartile range, 31 to 41% in 2011 versus 0 to 27% for the whole study period) (34). With the reduced variation in ART coverage, we lack the power to see a significant effect between the population-based HIV testing of all consenting adults aged 15 years or older (16). After obtaining written informed consent, field workers collect blood by finger prick and prepare DBSs for HIV testing according to the UNAIDS and World Health Organization’s Guidelines for Using HIV Testing Technologies in Surveillance. We conducted a population-based viral load survey in 2011 (described below) in the surveillance area and followed up participants known to be HIV-negative on 1 January 2011 until December 2015. We then used this information to quantify the relationship between viral load in the surrounding local community and future risk of acquisition of HIV infection. The analysis is described in detail below.

**MATERIALS AND METHODS**

**Study design**

Since 2004, the Africa Health Research Institute has conducted annual population-based HIV testing of all consenting adults aged 15 years or older (16). A clear rationale for targeting areas of high transmission (30). There has also been a recent shift toward a more locally tailored epidemic response based on differences in both the underlying epidemiology and characteristics of the population at risk (31–33). In this regard, our findings show that even in a severely affected rural African population with a well-established HIV treatment program, a PVL measure, such as the proportion of the overall population having HIV viremia (PDV_v), could play a role in targeting and monitoring the effectiveness of interventions in the most vulnerable communities where future levels of HIV incidence are likely to be highest.

**Table 5. Results of the multivariable analysis (Cox proportional hazard model) to examine the relationship between the risk of HIV acquisition for males and the three female PVL measures.** The PVL measures were derived from the HIV-positive females (models A1 to A3) and the HIV-positive and HIV-negative females (models B1 to B3) of a population-based survey. The full output is given in tables S4 and S6. Model 1 shows the unadjusted HRs for the PVL measures; model 2 shows these HRs after adjusting for age, sex, urban status, marital status, number of sexual partners in the last year, and household wealth; model 3 shows these HRs after adjusting for the model 2 covariates as well as HIV prevalence.

| Setting | Study site has collected detailed sociodemographic and health-related information on a population of about 87,000 individuals within a circumscribed geographic area (438 km² in area) in rural KwaZulu-Natal, South Africa (16). One of the strengths of the comprehensive surveillance platform is its longitudinal integrity and ability to follow the entire population. | 8 of 10 |
to record exact periods of time spent living at multiple locations by each individual under surveillance. In the study population, 29% of the adult population is infected with HIV (34). The rate of new infections remains high and relatively constant over time at about three new infections per 100 person-years (3, 35). The incidence of new HIV infection peaks in women at about 7.5 per 100 person-years (at age 25) and in men at 5 per 100 person-years (at age 30). The population is characterized by low levels of marriage, with only 31% of women and 23% of men ever having been married and 14% of those marriages being polygamous for men (36).

### Construction of the PVL measures

We used the viral load measurements of the 2011 population-based survey (n = 2420) to obtain sensitive and realistic PVL measures in the unique virtual community surrounding each homestead in the study area. These PVL measures were computed by means of a moving two-dimensional Gaussian kernel of a 3-km search radius (37). The size of the kernel was determined from the results of a previous work (38). First, all participants were located to an exact homestead of residence, and the viral load measurements are superimposed on a geographic representation of the study area consisting of a grid of 30 × 30-m pixels. Next, the kernel moves systematically across the grid and calculates a Gaussian-weighted estimate of the PVL measure for the unique neighborhood around each and every pixel on the grid. The method is well suited to the scattered distribution of the population because it does not impose any static geographical boundaries on the data. Instead, it uses the precise location of each homestead to derive a PVL measure that is both responsive to local variations and robust to the effects of random noise.

We used exactly the same methods to quantify spatial variations in the viral load indices obtained from routine clinical data. These data were collected at the facility level for 3196 patients living in the surveillance population who visited 1 of the 17 health care clinics in the subdistrict in 2011. Patients were geolocated to their exact homestead of residence through linkage to the population database (20).

### Statistical analysis

We followed up all 8732 repeat-testers who were resident within the surveillance area between January 2011 and December 2015. A repeat-tester is a study participant who was known to be HIV-negative on 1 January 2011 and who was tested for HIV during the study period. Given the time to event structure of the data, we used Cox proportional hazard models to conduct the analyses. Because the data are interval-censored, we imputed a single random seroconversion date (using a uniform distribution) between the repeat-tester’s latest HIV-negative and earliest HIV-positive dates, as described in greater detail elsewhere (39). Participants seldom test every year, and the median interval of time between the last HIV-negative and first HIV-positive tests in this group of repeat-testers is 1.74 years.

We evaluated each of the PVL measures separately in the analysis. We adjusted for well-established determinants of HIV acquisition identified in our previous works (3, 35), that is, sex, age, area of residence (rural/urban), marital status, number of partners in the last 12 months, household socioeconomic status (based on household assets), and HIV prevalence in the unique community surrounding each HIV-negative individual. We computed community-level HIV prevalence using the Gaussian kernel methodology described above on the basis of the data from 10,375 participants who participated in population-based HIV testing in 2011 (3). We treated the PVL measures and the covariates as time-invariant. In other words, a repeat-tester was exposed to the PVL of his or her surrounding local community for the duration of the study (that is, until the right censoring date). As a further robustness check of our findings, we conducted a series of analyses to test whether the same patterns could be seen in the relationship between sex-specific PVL patterns and risk of acquisition of infection in the opposite sex. Last, we performed a further set of parallel analyses to ascertain whether there was any relationship between the viral load indices derived using the routine facility–based viral load data and future risk of HIV acquisition.

### REFERENCES AND NOTES


Effect of population viral load on prospective HIV incidence in a hyperendemic rural African community

Frank Tanser, Alain Vandormael, Diego Cuadros, Andrew N. Phillips, Tulio de Oliveira, Andrew Tomita, Till Bärnighausen and Deenan Pillay

Sci Transl Med 9, eaam8012.
DOI: 10.1126/scitranslmed.aam8012

Status is not everything

Many parameters are examined to try to understand HIV transmission in endemic areas. Tanser et al. use longitudinal population-based data from rural South Africa to show that population viral load indices incorporating geographical location and local HIV prevalence can be used to infer HIV transmission potential. Their data demonstrate that accounting for HIV-negative individuals in calculations and transmission models is important for appropriate interpretations. Their findings could be helpful in guiding prevention intervention strategies in hyperendemic settings.