Characteristics Associated With Human Immunodeficiency Virus Transmission Networks Involving Adolescent Girls and Young Women in Human Immunodeficiency Virus Prevention Trials Network 068 Study

Marie C.D. Stoner, PhD,† Ann M. Dennis, MD,‡ James P. Hughes, PhD,‡§ Susan H. Eshleman, MD, PhD,¶ Mariya V. Sivay, PhD,¶ Sarah E. Hudelson, BS,¶ M. Kate Grabowski, PhD,¶ F. Xavier Gómez-Olivé, PhD,¶¶ Catherine MacPhail, PhD,¶¶¶ Estelle Piwowar-Manning, MT (ASCP),¶ Kathleen Kahn, MD, PhD,¶¶¶¶ and Audrey Pettifor, PhD*¶¶¶¶

Abstract: We combined behavioral survey data from the human immunodeficiency virus (HIV) Prevention Trials Network 068 study with phylogenetic information to determine if cluster membership was associated with characteristics of young women and their partners. Clusters were more likely to involve young women from specific villages and schools, indicating some localized transmission.

Supplemental digital content is available in the text.

Successful prevention of human immunodeficiency virus (HIV) transmission requires in-depth knowledge of local HIV epidemics, including transmission patterns among high-risk groups. Even within generalized epidemics in Southern Africa, some subgroups are disproportionately affected by a high risk of HIV acquisition. Such is the case for young women in rural South Africa, where HIV prevalence is 3 to 4 times higher than similarly aged young men. In addition, migration is prevalent in Southern Africa, including in our study site in the rural northeast region near the Mozambique border. Characteristics, such as migration for work and contact with older partners, may affect HIV acquisition among young women, but the contribution of these factors to local HIV transmission networks is uncertain.

Phylogenetic analysis provides information about HIV genetic networks and can identify putative transmission chains involving more than one young woman through shared sexual partners. Such analyses can help assess if HIV transmission occurs in social circles, through age-disparate relationships, in certain geographic areas, or through extensive migration of male partners. We previously characterized viral clustering among young women in South Africa, finding high level of HIV diversity, suggesting that migration is an important contributor to HIV transmission in rural South Africa; however, behavioral factors were not assessed. In this report, we evaluated sociobehavioral associations with cluster membership including newly detected HIV infections from a follow-up visit. Socio-behavioral information included measures of migration, partner characteristics, and geographic residence.

METHODS

Study Population

We analyzed samples and data from the HIV Prevention Trials Network (HPTN) 068 study, a randomized trial where young women and their households were given a cash transfer to reduce HIV acquisition. The study enrolled 2533 young women aged 13 to 20 years who were unmarried, not pregnant, and attending high school grades 8 to 11 in the Bushbuckridge subdistrict of Mpumalanga province, South Africa. Young women enrolled in the study were seen annually from 2011 to 2015 until study
Phylogenetic Analysis

HIV pol sequences from study participants were aligned to HXB2 and were manually edited with stripping of gapped positions. Sequences with a high fraction of ambiguous nucleotides (>5%) were excluded. We assessed regional diversity by evaluating clades that involved sequences sampled outside of South Africa. We conducted a BLAST search to identify the 10 sequences in the Los Alamos National Laboratory HIV database (http://www.hiv.lanl.gov) that were most closely related to each study sequence; only sequences in Los Alamos National Laboratory with known locations were used in the phylogenetic analysis. A maximum-likelihood (ML) tree was constructed in PhyML3.0 with the general time-reversible model of nucleotide substitution to evaluate clades involving study sequences. Statistical support of clades was assessed with local support values (Shimodaira-Hasegawa-like test). We defined closely related clusters as clades involving 2 or more study sequences with pairwise divergence of 0.020 or less nucleotide substitutions to another study sequence, which is less than 0.3% quantile of all pairwise comparisons between sequences (Figures 2 and 3, Supplementary Content http://links.hww.com/OLQ/A329). Clusters were visualized in R using the igraph Package. Pairwise genetic distances were evaluated with the Ape Package using the Tamura-Nei 93-substitution model.

Statistical Analysis of Characteristics

We examined descriptive characteristics of HIV-infected young women in a cluster versus not and, among those in a cluster, characteristics of cluster membership. We used a χ² test to test for differences between groups. For continuous variables, we calculated the intraclass correlation statistic as a measure of cluster membership. For categorical measures, we estimated the probability of having the same value for a randomly chosen pair of girls in the same cluster, over the probability of having the same value for a randomly chosen pair of girls overall. In the case of no association with cluster membership, we expected this ratio to be 1; an association with cluster membership would lead to a number greater than 1. Confidence intervals (CIs) were calculated using the bootstrap standard deviation from 200 full samples (with replacement) from the observed data. Formula: Cluster membership statistic (CMS) = P(cluster)/P(overall) where

\[ P_{\text{cluster}} = \frac{\sum_{i=1}^{n_{\text{clust}}} \sum_{j>1}^{n_{\text{clust}}} I(Y_i = Y_j)}{\sum_{h=1}^{n_{\text{clust}}} m_h (m_h-1)/2} \]

\[ P_{\text{overall}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} I(Y_i = Y_j)}{(N(N-1))/2} \]

where N is the number of girls; nclust, number of clusters; mh, size of cluster h; Y, outcome (school/village) for girl i and I(Yi = Yj) is equal to 1 if Yi = Yj and 0 otherwise.

**FINDINGS**

Of 231 (80.2%) participants with HIV sequence data, 228 were included in the phylogenetic analysis. Three sequences had ambiguity fractions >5% and were excluded. Among study participants, completion or they graduated from high school. Annual study visits included an Audio Computer-Assisted Self-Interview (ACASI) with young women and HIV testing for the young women who were HIV-uninfected at the previous visit. More information on the study design and main trial results are reported elsewhere. A postintervention follow-up visit was also conducted 1 to 2 years after the young women exited the main trial (2015–2017).

During the study period, 288 participants acquired HIV infection. Genotyping was successful for 231 (80.2%) of the total 288 HIV cases. Our previous report included analysis of HIV sequences from 201 women (68 infected at enrollment, 92 infected in the main study, 41 infected in the follow-up study). This report includes sequences from 30 additional women who seroconverted in the follow-up study. The annual ACASI survey collected self-reported information on demographics, risk behaviors, and partner characteristics. As proxies for migration, we examined self-reported characteristics including young women's report of moving in the last 12 months or having slept away from home at least once a week, and having a partner that lives outside the village or province. For baseline and post-intervention visits, we used survey information from the same visit; for incident infections, we used survey information collected at the visit before HIV diagnosis. Among girls who were members of a cluster, all girls had complete school information but 8 of 51 were out of school at the time of infection.

**Figure 1.** HIV transmission clusters involving sequences from 51 young women sampled in the Bushbuckridge subdistrict of Mpumalanga province, South Africa. Clusters are shown grouped by region of village location (some clusters involve more than 1 region). Clusters are labeled by cluster id number. Node shape corresponds to individuals connected by village location; edges (lines) represent linkages ≤0.02 nucleotide substitutions per site based on Tamura-New 93 model. Bold outlines indicate that clusters include ≥2 individuals (sequences) from the same village. Nodes are color coded by (A) schools and (B) wealth quantile (3, lowest to 0, highest) that included 3 or more persons in the network (A-F). Schools that are not reported indicate girls out of school and other refers to other school that are not A-F.
sequences, nearly all (227 of 228) were subtype C; one was subtype A. There were 841 unique references identified by the BLAST search. In the ML tree, a high degree of viral diversity was noted, including within clades containing only South African sequences (Fig. 1, Supplementary Content http://links.lww.com/OLQ/A329). While the majority of sequences in all clades were from South Africa, several clades included a high proportion of sequences from Botswana. In total, 22% (51 of 228) of sequences were identified in 22 clusters based on ≤2% pairwise divergence from at least 1 other study sequence (Fig. 1). Nineteen clusters were in pairs (n = 2 sequences); 2 clusters included 3 sequences and 1 cluster included 7 sequences.

Among all, being in a cluster was associated with the village of residence (P = 0.0469) and grade (P = 0.0225), where young women in higher grades were more likely to be part of a cluster. Among girls that were part of a cluster, cluster membership was associated with village (P = 0.0131) but not with other characteristics.

Although over 27 schools were represented in the survey, 37% (n = 19) of clustered persons were at 4 schools and 22% of all HIV infections were at these schools; of 25 villages represented, 39% (n = 20) of clustered persons were in 3 villages, while 25% of all HIV infections were in these villages. We found multiple localized chains; School B was part of 5 clusters and schools A, C, and D were all part of 4 clusters (Fig. 1). Eight young women that were part of a cluster had infections out of school; the CMS was similar after excluding these girls. Among girls in a cluster, the highest percentage were from the lowest quartile (n = 16, 31%) which represented 24% of all HIV cases. In our calculated measure of cluster membership, we observed excess co-occupying by school (CMS, 1.73; 95% CI, 1.14–2.62), wealth quartile (CMS, 1.51; 95% CI, 1.09–2.07), and village (CMS, 2.13; 95% CI, 1.31–3.47) (Table 1). We did not find any clustering by other factors examined. However, the sample sizes for the clusters were small, which may have limited our ability to detect associations with these factors.

**DISCUSSION**

We identified that cluster membership was related to school, village, and wealth quartile. These associations involved multiple transmission chains and suggest localized sexual networks with shared partnerships. Similar geographic location, including village and school, may be a driver of HIV transmission for young women in this age group. Cluster membership was not related to age-disparate relationships, alcohol use or any measures of migration. Of note, the number of young women who were in the same cluster was very small. Lack of clustering may be an artifact of small sample sizes and low sampling fraction rather than a lack of association with these characteristics. Samples were not available from male partners of the young women in the study. Therefore, we are unable to link the characteristics of those men directly to the young women who were infected. Instead, our study relied on inferring that phylogenetically linked young women may have shared a common partner and, given the genetic distance threshold, indirect connections may be greater than 1 degree. Phylogenetic analyses confirming sexual transmission in HIV serodisordant couples indicates that this assumption is likely valid. Additionally, behavioral characteristics were self-reported and may be misreported. Audio Computer-Assisted Self-Interview was used to minimize social desirability bias in the reporting of sensitive behaviors. Lastly, many girls at the post-intervention visit had graduated from school (65%, n = 65 of 100) limiting cluster membership by school.

Clustering in this study was typically between pairs, although we did find 3 larger clusters and associations between cluster membership and geographical characteristics. Our results indicate that most transmissions were not related to each other and multiple viral introductions through migration or external sexual partnerships are common. Additionally, we show that the larger clusters included multiple girls from the same village or school, and that some rapid, localized transmission occurred in the area. Future studies could expand this analysis by collecting phylogenetic and behavioral information directly from the male partners of young women, and by using larger sample sizes.

**REFERENCES**


