Title:

Demographic characteristics of sources of HIV-1 transmission in the era of test and treat

One sentence summary:

Men aged 25-40 are the source of 43% of HIV transmissions in Zambia and HIV prevention activities should be more inclusive of this demographic group

Authors:

Matthew Hall[1], Tanya Golubchik[1], David Bonsall[1,2], Lucie Abeler-Dörner[1], Mohammed Limbada[3,4], Barry Kosloff[3,4], Ab Schaap[3], Mariateresa de Cesare[1,2], George MacIntyre-Cockett[1,2], William Probert[1], Oliver Ratmann[5], Ana Bulas Cruz[1], Estelle Piwowar-Manning[6], David N Burns[7], Myron S Cohen[8], Deborah J Donnell[9], Susan H Eshleman[6], Musonda Simwinga[3], Richard Hayes[4], Sarah Fidler[10], Helen Ayles[3,4], Christophe Fraser[1] on behalf of the HPTN 071 (PopART) Phylogenetics protocol team and the PANGEA consortium.

Affiliations:

[1] Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[2] Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[3] Zambart, University of Zambia, Lusaka, Zambia

[4] London School of Hygiene and Tropical Medicine, London, United Kingdom

[5] Department of Mathematics, Imperial College, London, UK

[6] Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[7] Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, United States of America.

[8] Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA

[9] Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[10] Department of Infectious Disease Epidemiology, Imperial College, London, UK

Abstract:

In recent years, universally available antiretroviral therapy has reduced HIV infection rates in sub-Saharan Africa. We used phylogenetics to understand the demographic characteristics of transmission in the HPTN 071-02 (PopART) study in Zambia. After estimating that 59.4% of transmissions were from a male source to a female recipient, we further found that 43.1% (95% CI: 36.6%-49.5%) were from men aged 25-40. Overall, men infected 2.09-fold more women per capita than vice versa, a ratio peaking at 5.98 (3.14-13.6) amongst those aged 35-39. 25.6% (19.9%-31.3%) of sources were recently infected, 16.8% (12%-21.7%) transmitted drug-resistant virus and 13% lived in a different community than the recipient; these characteristics played no outsize role in transmission. Men aged 25-40 should be a focus of treatment for prevention interventions.

Main Text:

The last decade has seen a global transformation in HIV care, with the near universal availability of affordable and effective combination antiretroviral therapy (ART), that durably suppresses viral replication, prevents and even reverses progression to AIDS, and prevents onward transmission of the virus. The discovery that combination ART blocks transmission of the virus led to the concept that annual

testing followed by initiation of ART for infected individuals may be an effective form of HIV prevention (*1*).

The HPTN 071 (PopART) cluster-randomised trial evaluated whether a combination prevention package including universal testing and treatment would reduce HIV incidence in 21 communities in Zambia and South Africa (*2*). In the home-based package, community workers conducted HIV testing and supported linkage to HIV care and ART adherence. The trial reported a 20% reduction of HIV incidence in study arms with the combination prevention package, but residual transmission remained. All reductions in HIV infection rates are welcome, yet reaching ambitious UNAIDS goals for reduced infection rates will require more rapid progress. The aim of this study is to quantify the sources of residual transmission, such that future prevention efforts can address gaps in prevention.

To do this, the HPTN 071-02 (PopART) Phylogenetics Study was set up as an ancillary study to HPTN 071 (PopART) in the Zambian communities. The aims of the study were to use phylogenetics to find the sources of ongoing transmission and characterise them, help to assess the effectiveness of the intervention had it been rolled out nation-wide, and to identify the most promising policies for prevention in the future. We report here on the primary findings of the study.

HIV phylogenies reflect the demographic and spatial history of transmission in a specific setting (*3*, *4*). In population-level analyses, HIV phylogenetic inferences have been used to describe the origin and global spread of HIV (*5*), characterize transmission dynamics at the population level in concentrated epidemics (*6*), and, building on studies that examined individual transmission events (*7*), to analyse transmission linkages in HIV prevention trials (*1*, *8*, *9*). More recently, phylogenetic analysis has been applied to understand epidemics in sub-Saharan African settings, looking at patterns of clustering (*10*) and spatial spread (*11–13*).

To understand the risk factors for onward HIV transmission, it is necessary to conduct large phylogenetic studies with high sampling density, such that populations of de-identified probable source-recipient pairs can be analysed. Two prior studies used this approach to identify the most likely transmitters and risk-factors for transmission in well-characterized, concentrated European HIV networks consisting mostly of men who have sex with men (*14*, *15*). Here, we report a study of this kind in a large generalised epidemic in sub-Saharan Africa. A similar study is being conducted for the Ya Tsie Trial in Botswana (*16*). We characterised the sources of transmission by age and gender, and determined if they were infected less than one year before the transmission, whether the sources were from a different community than the recipient, and if their dominant viral strain was resistant to first-line ART. Future work will address the question of whether phylogenetic analysis reveals significant differences between the arms of the trial.

Results

Trial participants and sequences

The HPTN 071 (PopART) cluster-randomised trial was conducted in Zambia and South Africa between 2013 and 2018. HPTN 071-02 (PopART) Phylogenetics was conducted in nine of the twelve Zambian PopART communities and included participants from the PopART Population Cohort (PC) and additional participants enrolled at health care facilities (HCF) in the same communities. For this study, we developed a custom high-throughput affordable virus sequencing method that could reliably infer whole genomes from samples of blood which had been previously used to determine a participant's CD4 count, thus avoiding unnecessary blood draws, addressing a key concern of participants (*17*). We validated the method by comparing sequences to a clinically accredited drug resistance assay, showing high sensitivity and reliability above 10,000 copies per mL (*18*). Full or near-full sequence coverage was obtained for 5,613 participants (Figure 1, see detailed table in the supplement). Next generation sequencing (NGS) files are available via the PANGEA consortium (www.pangea-hiv.org)(*19*) .

Estimation of time since infection

Our pipeline included the use of a machine learning approach (*20*) to estimate the date of infection of each individual from within-host phylogenetic data and their sampling date. The model used mean genomic diversity (minor allele frequency) and mean genomic divergence (largest subgraph root-to-tip distance) across all windows in the *gag*, *pol* and *env* genes to estimate how long the virus had been evolving in each host. This allowed us to estimate how long ago each individual had been infected, when the transmission event took place and for how long the source had been infected at transmission. Combined with the recorded date of birth for each participant, this information allowed us to estimate the age of each individual at any time point, for example the time they were infected or, potentially, the time they were estimated to transmit the virus to another individual.

Identification of transmission pairs

The phylogenetic analysis aimed to find phylogenetically probable source-recipient transmission pairs by analysing within-host genetic diversity, using our *phyloscanner* software package (*21*), which we further refined to work with large datasets. The phyloscanner approach has been validated in a previous trial by the HPTN network (*22*), and in population studies (*23*). This represents the largest prospective study using a methodology of this kind to investigate HIV transmission to date. Phylogenetic analysis of all NGS reads identified 468 probable opposite-sex transmission pairs. A model of sampling of transmission chains estimates that each of these pairs has a probability of 90.4% of representing direct transmission (see supplement). Next to the 468 opposite-sex pairs, the analysis yielded 336 same-sex pairs, which were excluded from further analysis as they are likely to represent unsampled intermediaries (see supplement). The distribution of mixed-sex and same-sex pairs was not consistent with a significant contribution of cryptic transmission amongst men who have sex with men (see supplement). The likely direction of transmission in the pairs was assessed by two independent methods, firstly the phylogenetic topology and secondly comparison of estimated infection dates. Phylogenetic topology identified 268 directed pairs, and the date estimate 230. 143 pairs were called by both methods, while 24 pairs were excluded due to

conflicting results. The combined analysis yielded 355 probable transmission pairs with good confidence in the direction of transmission. For 300 of these, transmission was estimated to have occurred during the trial period (between November 2013 and December 2019) (Figure 1). An example of how a transmission pair is identified via topology is depicted in Figure 2. The number of transmission pairs identified was consistent with a power calculation conducted before the study to determine the number of transmission pairs required to characterise the sources of infection (see supplement), which was reassuring both in terms of aims of the study, and the validity of the method to determine transmission.

Characteristics of sources

We characterised the source partners in the transmission pairs by gender, age, recency of infection, resistance to first-line ART, and whether they were registered in a different community than the recipients. A key step in the analysis is adjusting for sampling bias. To do this, we made the observation that a source attribution study can be seen as a survey in which the recipients of HIV infection provide information about the transmissions that resulted in their infections, in this case the characteristics of their sources. Thus, we used iterative proportional fitting to weight the population of recipients so they were demographically representative of the population of all recipients who acquired HIV infection during the trial with regard to four variables: sex, birth cohort, community, and marital status. Unless otherwise stated, proportions and medians given in this section are all weighted to adjust for sampling biases with respect to these four variables.

The uptake of prevention and treatment varies by age and sex, and different strategies are needed for care to reach different groups where there are gaps. A previous phylogenetic study has suggested that the epidemic is maintained through a renewal process, where young women are infected by older men, who themselves are infected by older women (*24*). Policies aimed at breaking this cycle have been promoted by UNAIDS and PEPFAR, in particular with the DREAMS programme. Conversely, much prevention effort has focused on intergenerational relationships, older men and the ‘sugar daddy’ phenomenon. We

find little support for either of these views. The age distributions for both males and females showed the highest transmission from middle-aged individuals: 72.4% (95% CI: 65.6%-80.3%) of male sources were between 25 and 40 years of age at the time of transmission, representing 43.1% (36.6%-49.5%) of all transmissions, while 72.8% (64.3%-81.3%) of female sources were between 20 and 35 years of age at the time of transmission, representing 29.5% (23.6%-35.5%) of all transmissions (Figure 3A). The median age at time of transmission was 32 years (range 20-64) for male sources and 25 (17-49) for female sources (Figure 3A). The distributions are wide and consistent with the age and sex of typical reported sexual partnerships, indicating a generalised epidemic in which risk is distributed widely and many individuals transmit to their regular partners.

Pre-weighting, 56.7% of transmissions were from male to female participants; weighting brought this to 59.4% to match the estimated proportion of women in the overall HIV-positive population (Figure 3A). We calculated the relative transmission rates amongst demographic groups, defined as the ratio of the proportion of sources identified from a group to the group's proportion of the overall HIV-positive population. For men, this relative transmission rate was 1.44 (bootstrapped 95% CI: 1.43-1.5) while for women it was 0.69 (0.65-0.69). The ratio of these two numbers is 2.09 (2.04-2.29), indicating that HIV-positive males were contributing more than twice as many new infections as HIV-positive females. The ratio was 1.79 (1.78-1.95) if the group's proportion of the HIV-positive population not on ART was used as a denominator.

Figures 3B and 3C show these statistics for both denominators stratified by age group. The peak relative rate amongst females was observed for the 20-24 year age group (2.03 compared to the HIV+ population, 95% CI 1.44-2.55) while for males it was in the 30-34 year age group (2.71; 2.03-3.33). The male-to-female ratio was 4.23 (2.57-7.8) in the 30-34 year age group and 5.98 (3.14-13.6) in the 35-39 year age group. This ratio was above 1 in all age groups except 20-24 (we saw no male sources in the 13-20 age

group and thus could not construct a bootstrapped confidence interval), but the confidence interval included parity for the 20-24, 25-29 and 45-49 year age groups for both denominators.

A second area of considerable uncertainty is the role of acute and early HIV infection in transmission (*25*). During early HIV infection, viral loads are elevated, and the virus itself may be adapted to transmission. Frequent transmission from early infection would compromise test and treat strategies for HIV prevention, since individuals would rarely start antiretroviral therapy early enough to prevent transmission. Estimates range from ~70% (*26*) to <5% (*27*). In the set of transmission pairs, 72 recipients were infected by sources who themselves were estimated to have been infected within the 12 months before the transmission occurred (Figure 4A). This corresponds to a weighted proportion of 25.6% (19.9%-31.3%). This estimate lies between published extremes, in line for example with (*28*), and indicates that transmission from acute and early infection is not the main barrier to treatment as prevention.

A third way in which treatment as prevention (and treatment for individual benefit) could be compromised is through the spread of viral lineages carrying drug resistance mutations. Drug resistance has been documented as increasing in many Southern African settings (*29*). For our data, drug resistance to first-line adult ART was predicted based on detection of mutations in the Illumina reads generated using the Stanford HIV Drug Resistance Database scoring system (HIVdb version 8.9.1) (*30*) with the following cut-offs: scores 0-14 - WT/susceptible, scores 15-29 - low-level resistance, scores 30 and above - high-level resistance. After omitting eleven individuals whose viruses had an unknown drug resistance profile, a large majority of sources (236/289) transmitted drug-sensitive viruses, while 20 participants transmitted HIV with low-level resistance to first-line ART and 33 HIV with high-level resistance to first-line ART (Figure 4B). The weighted proportions of individuals transmitting sensitive, low-level resistant and high-level resistant HIV were 82.5% (77.5%-87.5%), 6.7% (3.4%-10%) and 10.8% (6.7%-14.9%), respectively. This is one of the highest levels of transmitted drug resistance recorded. At the present time,

the World Health Organisation has changed the recommended first-line drug resistance such that this problem is mitigated, but this finding does question the sustainability of providing treatment without adequate clinical or virological follow up.

Finally, we determined the fraction of transmission pairs for which the source was living in a different community than the recipient (Figure 4C). Though movement does not necessarily compromise the effect of treatment as prevention, it does compromise the ability of the cluster randomised trials such as HPTN 071 (PopART) to accurately measure the effect of the intervention. Sources were classified as coming from outside the community if they were sampled in a different study community than the recipient. 263 of 300 transmissions occurred within the same community, a weighted proportion of 87% (83.6%-91.4%). For the interpretation of the HPTN 071 (PopART) trial, 13% cross-community transmission means that the effectiveness of the trial was likely underestimated in the original analysis, since it did not account for movement of virus between intervention and control communities. This finding indicates that the PopART intervention package would have been more effective if it had been implemented country-wide.

Combined characteristics of sources

Our findings so far rule out previously hypothesised drivers in the Zambian epidemic: large contributions from young women, early infection, or mobile people moving into the communities, and indicate that infections are predominantly local and disproportionately from men. We next sought to understand the extent to which characteristics of sources might occur disproportionately often in combinations. For example, do men aged 25-40, with drug resistance, and from another community, appear more or less often as sources of infection than would be expected from each of those characteristics independently. To do this, we ranked each combination according to the fraction of transmission in which they played a role (Figure 5A). The highest fraction of transmissions originated from males aged 25-40 who were infected more than a year prior to transmission, and lived in the same community. They were the sources for twice

as many transmissions as the female risk group aged 20-35 with the same characteristics and almost twice as many transmissions as all sources outside these two age groups. These groups were followed by the same groups who were infected less than a year before transmission, were registered in a different community, and whose dominant viral population was resistant to first-line ART. There was no group that disproportionately combined several of the risk factors, as real values were generally very close to values expected if risk factors present in the population of sources had been randomly assigned to this population (Figure 5B).

The analysis also shows which transmissions could have been averted by the PopART intervention package and which would have been missed. Testing in the trial was approximately yearly which means that transmissions caused by groups which have a black dot in the third row might have been missed. Transmissions from a different community would likely also have been missed if they originated in one of the control communities in which the prevention package was not offered. The analysis offers a convenient way to assess different prevention policies and the maximum impact on incidence they are likely to have.

Discussion

This study identified likely transmission pairs and the characteristics of sources of HIV transmission in Zambia between 2014 and 2019. The aim of the study is to provide policy makers with data on where transmission is still occurring in a generalised epidemic in sub-Saharan Africa despite delivery of a universal testing and treatment intervention and advise where prevention efforts should be focused.

We identified men aged 25-40 as the group most commonly transmitting HIV. This group was the source of 43.2% of transmissions (figure 5C), despite males in the 25-39 age group representing only 19.4% of

prevalent cases in 2017 in the PopART model. Women aged 20-35 were responsible for 30.3% of transmissions. Age distributions of sources were wide, suggesting that risk of infection was common across a large fraction of society. We did not find evidence for a renewal cycle involving chronically infected older women, as suggested previously (*24*). Each HIV positive man infected on average 2.09 times as many people as an HIV positive woman, and when stratified by age this number rose to 4.23 for men and women in the 30-34 year age group and to 5.98 in the 35-39 year age group. Men in this age group were less likely than other demographic groups to be linked to care and on ART (*31*). These results highlight the need for more intensive efforts in UTT programmes to achieve high coverage of men aged 25-40 and suggest that pre-exposure prophylaxis services should also be considered, particularly for young women (*32*, *33*).

A quarter of sources in this study were estimated to have been infected for less than a year before passing on the infection (figure 5C). Previous estimates have ranged between 5-10% and 60-80%, reviewed in (*25*, *34*). Our results suggest that both recently infected individuals and individuals with chronically unsuppressed HIV require attention: more frequent testing to avert transmission from recent infections and efforts to re-engage and retain people in care to avert transmission from chronic infection.

In this study, 10.4% and 6.4% of sources' viral population were dominated by variants with high-level and low-level resistance against first-line ART, respectively. This mirrors the high percentage of first-line ART resistance in Sub-Saharan Africa reported by other studies in the past few years (*29*, *35*, *36*). The rollout of dolutegravir is urgently needed and will hopefully reverse this trend. Resistance against first-line ART including dolutegravir should be monitored by local surveillance teams as availability of HIV drug resistance testing and NGS sequencing in sub-Saharan Africa increases.

While a large majority of transmissions occurred between partners living in the same community, 13% did not (figure 5C). As the primary analysis of the HPTN 071 (PopART) trial (*2*) took no account of

between-community transmission, our findings indicate that HIV incidence would have been further reduced if the PopART prevention package had been implemented country-wide. The complex statistical corrections required to assess the impact of the intervention taking into account between-community transmission are outside the scope of this study, but further analysis of the phylogenetics data is likely to lead to a higher estimate of the effectiveness of the intervention.

This study has limitations. While sampling fractions have been high and reached between 4.9% and 9.5% of people living with HIV in the individual study communities, the sample was still not fully representative of the HIV-positive population as a whole. In particular, recruitment of one person per household in the PC limited our ability to identify within-household transmission pairs. This is partially mitigated by the collection of samples in the HCFs which included additional members of the same households. The sequencing approach yielded sequences for participants with more than 5000 viral copies/mL. This corresponds to 92.8% of participants with viral loads above 1500 copies/mL who might have been involved in a transmission event (*37*).

HPTN 071-02 (PopART) Phylogenetics is the largest HIV phylogenetic study conducted to date, the first large transmission study to be based on an a priori power calculation, and the most comprehensive study of characteristics of sources of HIV-1 infection in sub-Saharan Africa. The study highlights that residual transmission occurred across a wide spectrum of common sexual interactions and was not limited to small risk groups. Men aged 25-40 however were, as a group, responsible for a large share of transmissions and should be prioritised in prevention efforts, even if linking them to care requires more effort.

Ethical considerations

Participants gave written consent specifically developed for the phylogenetics ancillary study (see protocol). Pseudonymised data were stored in a GDPR-compliant manner (*38*). Ethical approvals for the

HPTN 071 (PopART) trial and the HPTN 071-02 (PopART) Phylogenetics Study was granted by ethics committees at the London School of Hygiene and Tropical Medicine, University of Zambia, and Stellenbosch University, South Africa. The design of the study was discussed with community advisory boards. Results of the study were shared and discussed with participants in several local workshops.

References and notes:

References

1. M. S. Cohen, Y. Q. Chen, M. McCauley, T. Gamble, M. C. Hosseinipour, N. Kumarasamy, J. G. Hakim, J. Kumwenda, B. Grinsztejn, J. H. S. Pilotto, S. V. Godbole, S. Mehendale, S. Chariyalertsak, B. R. Santos, K. H. Mayer, I. F. Hoffman, S. H. Eshleman, E. Piwowar-Manning, L. Wang, J. Makhema, L. A. Mills, G. de Bruyn, I. Sanne, J. Eron, J. Gallant, D. Havlir, S. Swindells, H. Ribaudo, V. Elharrar, D. Burns, T. E. Taha, K. Nielsen-Saines, D. Celentano, M. Essex, T. R. Fleming, HPTN 052 Study Team, Prevention of HIV-1 infection with early antiretroviral therapy. *N. Engl. J. Med.* **365**, 493–505 (2011).

2. R. J. Hayes, D. Donnell, S. Floyd, N. Mandla, J. Bwalya, K. Sabapathy, B. Yang, M. Phiri, A. Schaap, S. H. Eshleman, E. Piwowar-Manning, B. Kosloff, A. James, T. Skalland, E. Wilson, L. Emel, D. Macleod, R. Dunbar, M. Simwinga, N. Makola, V. Bond, G. Hoddinott, A. Moore, S. Griffith, N. Deshmane Sista, S. H. Vermund, W. El-Sadr, D. N. Burns, J. R. Hargreaves, K. Hauck, C. Fraser, K. Shanaube, P. Bock, N. Beyers, H. Ayles, S. Fidler, HPTN 071 (PopART) Study Team, Effect of Universal Testing and Treatment on HIV Incidence - HPTN 071 (PopART). *N. Engl. J. Med.* **381**, 207–218 (2019).

3. A. Rambaut, D. L. Robertson, O. G. Pybus, M. Peeters, E. C. Holmes, Human immunodeficiency virus. Phylogeny and the origin of HIV-1. *Nature*. **410**, 1047–1048 (2001).

4. R. R. Gray, A. J. Tatem, S. Lamers, W. Hou, O. Laeyendecker, D. Serwadda, N. Sewankambo, R. H. Gray, M. Wawer, T. C. Quinn, M. M. Goodenow, M. Salemi, Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS*. **23**, F9–F17 (2009).

5. N. R. Faria, A. Rambaut, M. A. Suchard, G. Baele, T. Bedford, M. J. Ward, A. J. Tatem, J. D. Sousa, N. Arinaminpathy, J. Pépin, D. Posada, M. Peeters, O. G. Pybus, P. Lemey, HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science*. **346**, 56–61 (2014).

6. R. D. Kouyos, V. von Wyl, S. Yerly, J. Böni, P. Taffé, C. Shah, P. Bürgisser, T. Klimkait, R. Weber, B. Hirschel, M. Cavassini, H. Furrer, M. Battegay, P. L. Vernazza, E. Bernasconi, M. Rickenbach, B. Ledergerber, S. Bonhoeffer, H. F. Günthard, Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. *J. Infect. Dis.* **201**, 1488–1497 (2010).

7. T. Leitner, D. Escanilla, C. Franzén, M. Uhlén, J. Albert, Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10864–10869

(1996).

8. S. H. Eshleman, S. E. Hudelson, A. D. Redd, L. Wang, R. Debes, Y. Q. Chen, C. A. Martens, S. M. Ricklefs, E. J. Selig, S. F. Porcella, S. Munshaw, S. C. Ray, E. Piwowar-Manning, M. McCauley, M. C. Hosseinipour, J. Kumwenda, J. G. Hakim, S. Chariyalertsak, G. de Bruyn, B. Grinsztejn, N. Kumarasamy, J. Makhema, K. H. Mayer, J. Pilotto, B. R. Santos, T. C. Quinn, M. S. Cohen, J. P. Hughes, Analysis of genetic linkage of HIV from couples enrolled in the HIV Prevention Trials Network 052 trial. *J. Infect. Dis.* **204**, 1918–1926 (2011).

9. M. S. Campbell, J. I. Mullins, J. P. Hughes, C. Celum, K. G. Wong, D. N. Raugi, S. Sorensen, J. N. Stoddard, H. Zhao, W. Deng, E. Kahle, D. Panteleeff, J. M. Baeten, F. E. McCutchan, J. Albert, T. Leitner, A. Wald, L. Corey, J. R. Lingappa, Partners in Prevention HSV/HIV Transmission Study Team, Viral linkage in HIV-1 seroconverters and their partners in an HIV-1 prevention clinical trial. *PLoS One*. **6**, e16986 (2011).

10. V. Novitsky, H. Bussmann, A. Logan, S. Moyo, E. van Widenfelt, L. Okui, M. Mmalane, J. Baca, L. Buck, E. Phillips, D. Tim, M. F. McLane, Q. Lei, R. Wang, J. Makhema, S. Lockman, V. DeGruttola, M. Essex, Phylogenetic relatedness of circulating HIV-1C variants in Mochudi, Botswana. *PLoS One*. **8**, e80589 (2013).

11. V. Novitsky, M. Zahralban-Steele, S. Moyo, T. Nkhisang, D. Maruapula, M. F. McLane, J. Leidner, K. Bennett, PANGEA Consortium, K. E. Wirth, T. Gaolathe, E. Kadima, U. Chakalisa, M. Pretorius Holme, S. Lockman, M. Mmalane, J. Makhema, S. Gaseitsiwe, V. DeGruttola, M. Essex, Mapping of HIV-1C Transmission Networks Reveals Extensive Spread of Viral Lineages Across Villages in Botswana Treatment-as-Prevention Trial. *J. Infect. Dis.* **222**, 1670–1680 (2020).

12. O. Ratmann, J. Kagaayi, M. Hall, T. Golubchick, G. Kigozi, X. Xi, C. Wymant, G. Nakigozi, L. Abeler-Dörner, D. Bonsall, A. Gall, A. Hoppe, P. Kellam, J. Bazaale, S. Kalibbala, O. Laeyendecker, J. Lessler, F. Nalugoda, L. W. Chang, T. de Oliveira, D. Pillay, T. C. Quinn, S. J. Reynolds, S. E. F. Spencer, R. Ssekubugu, D. Serwadda, M. J. Wawer, R. H. Gray, C. Fraser, M. K. Grabowski, Rakai Health Sciences Program and the Pangea HIV Consortium, Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: a population-based study in Rakai, Uganda. *Lancet HIV*. **7**, e173–e183 (2020).

13. N. Bbosa, D. Ssemwanga, A. Ssekagiri, X. Xi, Y. Mayanja, U. Bahemuka, J. Seeley, D. Pillay, L. Abeler-Dörner, T. Golubchik, C. Fraser, P. Kaleebu, O. Ratmann, Phylogenetic and Demographic Characterization of Directed HIV-1 Transmission Using Deep Sequences from High-Risk and General Population Cohorts/Groups in Uganda. *Viruses*. **12** (2020), doi:10.3390/v12030331.

14. M. Fisher, D. Pao, A. E. Brown, D. Sudarshi, O. N. Gill, P. Cane, A. J. Buckton, J. V. Parry, A. M. Johnson, C. Sabin, D. Pillay, Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. *AIDS*. **24**, 1739–1747 (2010).

15. O. Ratmann, A. van Sighem, D. Bezemer, A. Gavryushkina, S. Jurriaans, A. Wensing, F. de Wolf, P. Reiss, C. Fraser, ATHENA observational cohort, Sources of HIV infection among men having sex with men and implications for prevention. *Sci. Transl. Med.* **8**, 320ra2 (2016).

16. S. Lockman, M. P. Holme, J. Makhema, P. Bachanas, J. Moore, K. E. Wirth, R. Lebelonyane, M. Essex, Implementation of Universal HIV Testing and Treatment to Reduce HIV Incidence in Botswana: the Ya Tsie Study. *Curr. HIV/AIDS Rep.* **17**, 478–486 (2020).

17. D. Bonsall, T. Golubchik, M. de Cesare, M. Limbada, B. Kosloff, G. MacIntyre-Cockett, M. Hall, C.

Wymant, M. A. Ansari, L. Abeler-Dörner, A. Schaap, A. Brown, E. Barnes, E. Piwowar-Manning, S. Eshleman, E. Wilson, L. Emel, R. Hayes, S. Fidler, H. Ayles, R. Bowden, C. Fraser, HPTN 071 (PopART) Team, A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *J. Clin. Microbiol.* **58** (2020), doi:10.1128/JCM.00382-20.

18. J. M. Fogel, D. Bonsall, V. Cummings, R. Bowden, T. Golubchik, M. de Cesare, E. A. Wilson, T. Gamble, C. Del Rio, D. S. Batey, K. H. Mayer, J. E. Farley, J. P. Hughes, R. H. Remien, C. Beyrer, C. Fraser, S. H. Eshleman, Performance of a high-throughput next-generation sequencing method for analysis of HIV drug resistance and viral load. *J. Antimicrob. Chemother.* **75**, 3510–3516 (2020).

19. L. Abeler-Dörner, M. K. Grabowski, A. Rambaut, D. Pillay, C. Fraser, PANGEA consortium, PANGEA-HIV 2: Phylogenetics And Networks for Generalised Epidemics in Africa. *Curr. Opin. HIV AIDS.* **14**, 173–180 (2019).

20. Golubchik T, Hall M, Bonsall D, Macintyre-Cockett G, WymantC , Abeler-Dörner L, Herbeck JT, Lingappa JR, Ssemwanga D, Kaleebu P, Grabowski MK, Gray R, Hayes R, Fidler F, Ayles H, Limbada M, Kosloff B, Pillay D, de Oliveira T, Kellam P, Cohen MS, Ratmann O, Laeyendecker O, Fraser C, PANGEA Consortium, PopART study, BEEHIVE study, HIV-phyloTSI: Accurate, subtype-independent estimation of time since HIV-1 infection using deep sequence data. *in prep.*

21. C. Wymant, M. Hall, O. Ratmann, D. Bonsall, T. Golubchik, M. de Cesare, A. Gall, M. Cornelissen, C. Fraser, STOP-HCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration, PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Mol. Biol. Evol.* **35**, 719–733 (2018).

22. Y. Zhang, C. Wymant, O. Laeyendecker, M. K. Grabowski, M. Hall, S. Hudelson, E. Piwowar-Manning, M. McCauley, T. Gamble, M. C. Hosseinipour, N. Kumarasamy, J. G. Hakim, J. Kumwenda, L. A. Mills, B. R. Santos, B. Grinsztejn, J. H. Pilotto, S. Chariyalertsak, J. Makhema, Y. Q. Chen, M. S. Cohen, C. Fraser, S. H. Eshleman, Evaluation of Phylogenetic Methods for Inferring the Direction of Human Immunodeficiency Virus (HIV) Transmission: HIV Prevention Trials Network (HPTN) 052. *Clin. Infect. Dis.* **72**, 30–37 (2021).

23. O. Ratmann, M. K. Grabowski, M. Hall, T. Golubchik, C. Wymant, L. Abeler-Dörner, D. Bonsall, A. Hoppe, A. L. Brown, T. de Oliveira, A. Gall, P. Kellam, D. Pillay, J. Kagaayi, G. Kigozi, T. C. Quinn, M. J. Wawer, O. Laeyendecker, D. Serwadda, R. H. Gray, C. Fraser, PANGEA Consortium and Rakai Health Sciences Program, Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis. *Nat. Commun.* **10**, 1411 (2019).

24. T. de Oliveira, A. B. M. Kharsany, T. Gräf, C. Cawood, D. Khanyile, A. Grobler, A. Puren, S. Madurai, C. Baxter, Q. A. Karim, S. S. A. Karim, Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. *Lancet HIV.* **4**, e41–e50 (2017).

25. M. S. Cohen, C. Dye, C. Fraser, W. C. Miller, K. A. Powers, B. G. Williams, HIV treatment as prevention: debate and commentary--will early infection compromise treatment-as-prevention strategies? *PLoS Med.* **9**, e1001232 (2012).

26. K. A. Powers, A. C. Ghani, W. C. Miller, I. F. Hoffman, A. E. Pettifor, G. Kamanga, F. E. Martinson, M. S. Cohen, The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. *Lancet.* **378**, 256–268 (2011).

27. B. G. Williams, How important is the acute phase in HIV epidemiology? *arXiv [q-bio.QM]* (2011), (available at http://arxiv.org/abs/1105.2767).

28. S. E. Bellan, J. Dushoff, A. P. Galvani, L. A. Meyers, Reassessment of HIV-1 acute phase infectivity: accounting for heterogeneity and study design with simulated cohorts. *PLoS Med*. **12**, e1001801 (2015).

29. R. K. Gupta, J. Gregson, N. Parkin, H. Haile-Selassie, A. Tanuri, L. Andrade Forero, P. Kaleebu, C. Watera, A. Aghokeng, N. Mutenda, J. Dzangare, S. Hone, Z. Z. Hang, J. Garcia, Z. Garcia, P. Marchorro, E. Beteta, A. Giron, R. Hamers, S. Inzaule, L. M. Frenkel, M. H. Chung, T. de Oliveira, D. Pillay, K. Naidoo, A. Kharsany, R. Kugathasan, T. Cutino, G. Hunt, S. Avila Rios, M. Doherty, M. R. Jordan, S. Bertagnolio, HIV-1 drug resistance before initiation or re-initiation of first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis. *Lancet Infect. Dis*. **18**, 346–355 (2018).

30. T. F. Liu, R. W. Shafer, Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin. Infect. Dis*. **42**, 1608–1618 (2006).

31. S. Floyd, K. Shanaube, B. Yang, A. Schaap, S. Griffith, M. Phiri, D. Macleod, R. Sloot, K. Sabapathy, V. Bond, P. Bock, H. Ayles, S. Fidler, R. Hayes, HPTN 071 (PopART) study team, HIV testing and treatment coverage achieved after 4 years across 14 urban and peri-urban communities in Zambia and South Africa: An analysis of findings from the HPTN 071 (PopART) trial. *PLoS Med*. **17**, e1003067 (2020).

32. E. M. Irungu, J. M. Baeten, PrEP rollout in Africa: status and opportunity. *Nat. Med*. **26**, 655–664 (2020).

33. H. Grant, G. B. Gomez, K. Kripke, R. V. Barnabas, C. Watts, G. F. Medley, Z. Mukandavire, Time to Scale Up Preexposure Prophylaxis Beyond the Highest-Risk Populations? Modeling Insights From High-Risk Women in Sub-Saharan Africa. *Sex. Transm. Dis*. **47**, 767–777 (2020).

34. S. E. Rutstein, J. Ananworanich, S. Fidler, C. Johnson, E. J. Sanders, O. Sued, A. Saez-Cirion, C. D. Pilcher, C. Fraser, M. S. Cohen, M. Vitoria, M. Doherty, J. D. Tucker, Clinical and public health implications of acute and early HIV detection and treatment: a scoping review. *J. Int. AIDS Soc*. **20**, 21579 (2017).

35. S.-Y. Rhee, J. L. Blanco, M. R. Jordan, J. Taylor, P. Lemey, V. Varghese, R. L. Hamers, S. Bertagnolio, T. F. Rinke de Wit, A. F. Aghokeng, J. Albert, R. Avi, S. Avila-Rios, P. O. Bessong, J. I. Brooks, C. A. B. Boucher, Z. L. Brumme, M. P. Busch, H. Bussmann, M.-L. Chaix, B. S. Chin, T. T. D'Aquin, C. F. De Gascun, A. Derache, D. Descamps, A. K. Deshpande, C. F. Djoko, S. H. Eshleman, H. Fleury, P. Frange, S. Fujisaki, P. R. Harrigan, J. Hattori, A. Holguin, G. M. Hunt, H. Ichimura, P. Kaleebu, D. Katzenstein, S. Kiertiburanakul, J. H. Kim, S. S. Kim, Y. Li, I. Lutsar, L. Morris, N. Ndembi, K. P. Ng, R. S. Paranjape, M. Peeters, M. Poljak, M. A. Price, M. L. Ragonnet-Cronin, G. Reyes-Terán, M. Rolland, S. Sirivichayakul, D. M. Smith, M. A. Soares, V. V. Soriano, D. Ssemwanga, M. Stanojevic, M. A. Stefani, W. Sugiura, S. Sungkanuparph, A. Tanuri, K. K. Tee, H.-H. M. Truong, D. A. M. C. van de Vijver, N. Vidal, C. Yang, R. Yang, G. Yebra, J. P. A. Ioannidis, A.-M. Vandamme, R. W. Shafer, Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med*. **12**, e1001810 (2015).

36. B. Chimukangara, R. J. Lessells, S.-Y. Rhee, J. Giandhari, A. B. M. Kharsany, K. Naidoo, L. Lewis, C. Cawood, D. Khanyile, K. A. Ayalew, K. Diallo, R. Samuel, G. Hunt, A. Vandormael, B. Stray-

Pedersen, M. Gordon, T. Makadzange, P. Kiepiela, G. Ramjee, J. Ledwaba, M. Kalimashe, L. Morris, U. M. Parikh, J. W. Mellors, R. W. Shafer, D. Katzenstein, P. Moodley, R. K. Gupta, D. Pillay, S. S. Abdool Karim, T. de Oliveira, Trends in Pretreatment HIV-1 Drug Resistance in Antiretroviral Therapy-naive Adults in South Africa, 2000-2016: A Pooled Sequence Analysis. *EClinicalMedicine*. **9**, 26–34 (2019).

37. T. C. Quinn, M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, R. H. Gray, Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. *N. Engl. J. Med.* **342**, 921–929 (2000).

38. C. F. Mondschein, C. Monda, in *Fundamentals of Clinical Data Science*, P. Kubben, M. Dumontier, A. Dekker, Eds. (Springer, Cham (CH), 2018; http://dx.doi.org/10.1007/978-3-319-99713-1).

39. The HIV Prevention Trials Network, (available at https://www.hptn.org/sites/default/files/inline-files/HPTN%20071-2%2C%20Version%202.0%20%2807-14-2017%29.pdf).

40. A. Gall, B. Ferns, C. Morris, S. Watson, M. Cotten, M. Robinson, N. Berry, D. Pillay, P. Kellam, Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes. *J. Clin. Microbiol.* **50**, 3838–3844 (2012).

41. D. E. Wood, S. L. Salzberg, Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

42. S. Nurk, A. Bankevich, D. Antipov, A. A. Gurevich, A. Korobeynikov, A. Lapidus, A. D. Prjibelski, A. Pyshkin, A. Sirotkin, Y. Sirotkin, R. Stepanauskas, S. R. Clingenpeel, T. Woyke, J. S. McLean, R. Lasken, G. Tesler, M. A. Alekseyev, P. A. Pevzner, Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J. Comput. Biol.* **20**, 714–737 (2013).

43. C. Kuiken, B. Korber, R. W. Shafer, HIV sequence databases. *AIDS Rev.* **5**, 52–61 (2003).

44. Picard, (available at https://broadinstitute.github.io/picard/).

45. *shiver: Sequences from HIV Easily Reconstructed* (Github; https://github.com/ChrisHIV/shiver).

46. S. L. Kosakovsky Pond, S. Weaver, A. J. Leigh Brown, J. O. Wertheim, HIV-TRACE (TRAnsmission Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens. *Mol. Biol. Evol.* **35**, 1812–1819 (2018).

47. C. Wymant, F. Blanquart, T. Golubchik, A. Gall, M. Bakker, D. Bezemer, N. J. Croucher, M. Hall, M. Hillebregt, S. H. Ong, O. Ratmann, J. Albert, N. Bannert, J. Fellay, K. Fransen, A. Gourlay, M. K. Grabowski, B. Gunsenheimer-Bartmeyer, H. F. Günthard, P. Kivelä, R. Kouyos, O. Laeyendecker, K. Liitsola, L. Meyer, K. Porter, M. Ristola, A. van Sighem, B. Berkhout, M. Cornelissen, P. Kellam, P. Reiss, C. Fraser, BEEHIVE Collaboration, Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. *Virus Evol.* **4**, vey007 (2018).

48. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

49. A. Rubinsteyn, *fancyimpute: Multivariate imputation and matrix completion algorithms implemented in Python* (Github; https://github.com/iskandr/fancyimpute).

50. scikit-learn: machine learning in Python — scikit-learn 0.16.1 documentation, (available at http://scikit-learn.org/).

51. M. Pickles, A. Cori, W. J. M. Probert, R. Sauter, R. Hinch, S. Fidler, H. Ayles, P. Bock, D. Donnell, E. Wilson, E. Piwowar-Manning, S. Floyd, R. J. Hayes, C. Fraser, HPTN 071 (PopART) Study Team, PopART-IBM, a highly efficient stochastic individual-based simulation model of generalised HIV epidemics developed in the context of the HPTN 071 (PopART) trial. *PLoS Comput. Biol.* **17**, e1009301 (2021).

52. Zambia Statistics Agency-ZSA, Ministry of Health-MOH, University Teaching Hospital Virology Laboratory-UTH-VL, ICF, Zambia Demographic and Health Survey 2018 (2020) (available at https://dhsprogram.com/publications/publication-fr361-dhs-final-reports.cfm).

53. Stan Development Team. 2021. Stan Modeling Language Users Guide and Reference Manual, 2.28, (available at https://mc-stan.org).

54. Pickles M, Cori A, Probert W, Sauter R, Fidler S, Ayles H, Bock P, Donnell D, Wilson E, Piwowar-Manning E, Floyd S, Hayes R, Fraser C on behalf of the HPTN 071 (PopART) Study Team., PopART-IBM, a highly efficient stochastic individual-based simulation model of generalised HIV epidemics developed in the context of the HPTN 071 (PopART) trial. *medRxiv*. **10.1101/2020.08.24.20181180** (2020), doi:10.1101/2020.08.24.20181180 .

55. M. Pickles, A. Cori, W. Probert, R. Sauter, S. Fidler, H. Ayles, P. Bock, D. Donnell, E. Wilson, E. Piwowar-Manning, S. Floyd, R. Hayes, C. Frase, PopART-IBM, a highly efficient stochastic individual-based simulation model of generalised HIV epidemics developed in the context of the HPTN 071 (PopART) trial. *bioRxiv* (2020), , doi:10.1101/2020.08.24.20181180.

Acknowledgements:

We thank all of the field staff and everybody working with the research team in Zambia for their contribution to this study.

Funding

HPTN 071 is sponsored by the National Institute of Allergy and Infectious Diseases (NIAID) under Cooperative Agreements UM1-AI068619, UM1-AI068617, and UM1-AI068613, with funding from the US President's Emergency Plan for AIDS Relief (PEPFAR). Additional

funding is provided by the International Initiative for Impact Evaluation (3ie) with support from the Bill & Melinda Gates Foundation, as well as by NIAID, the National Institute on Drug Abuse (NIDA), and the National Institute of Mental Health (NIMH), all part of NIH. HPTN 071-02 Phylogenetics is sponsored by NIAID, NIMH, and the Bill & Melinda Gates Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAID, NIMH, NIDA, PEPFAR, 3ie, or the Bill & Melinda Gates Foundation.

Author contributions

Conceptualization: CF, DBu, HA, MC, MS, OR, RH, SE, SF

Data curation: AS, CF, DD, TG

Formal Analysis: CF, DBo, MH, TG, WP

Funding acquisition: CF, DBu, HA, MC, OR, RH, SF

Investigation: DBo, GMC, MdC,

Methodology: Dbo, MH, TG

Project administration: ABC, BK, CF, DBo, EPM, LAD, ML, SE, SF

Software: DBo, MH, TG

Supervision: CF, HA

Dissemination of findings: HA, MS

Visualization: MH

Writing – original draft: CF, LAD, MH

Writing – review & editing: BK, DBo, DBu, MC, ML, RH, SE, SF, TG

Competing interests

WP currently works as a consultant for the WHO Global HIV, Hepatitis, and STI programme.

Data and materials availability

Code is available on request. HIV sequencing and epidemiological data are available via the PANGEA-HIV consortium data sharing procedures (www.pangea-hiv.org).

List of Supplementary Materials:

Materials and Methods

Supplementary Text

Tables S1 to S4

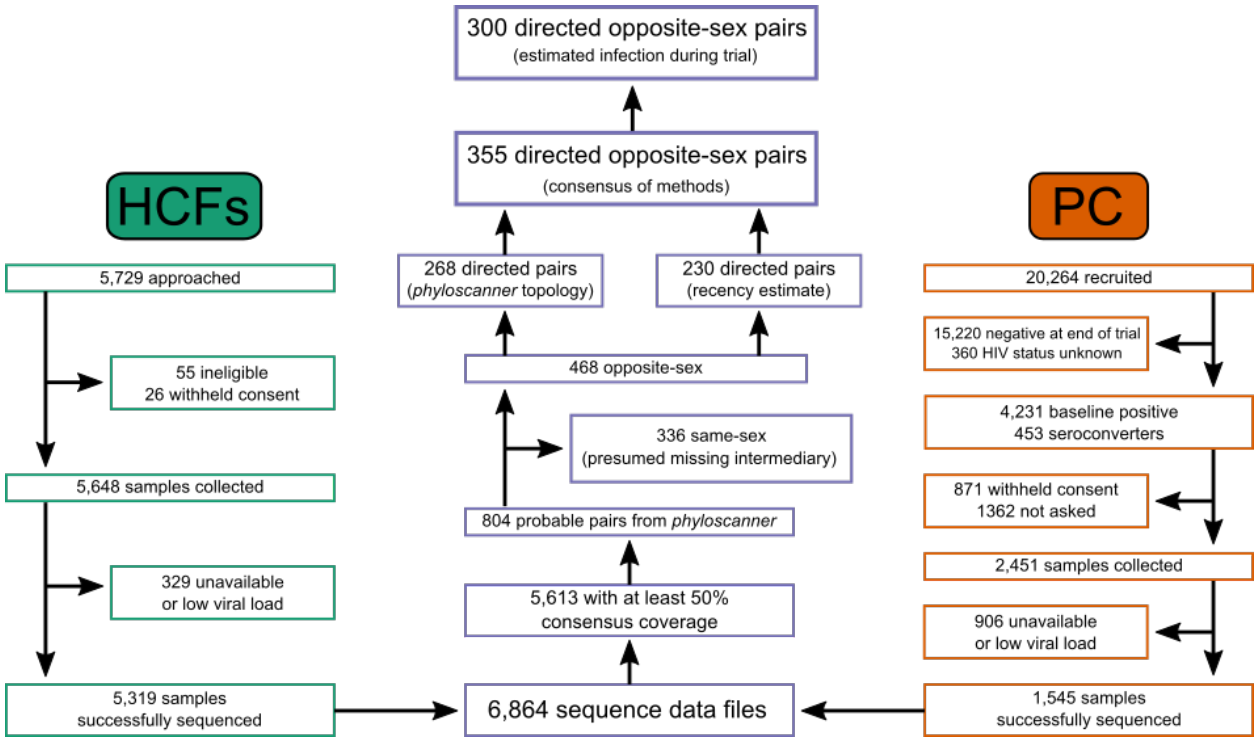


Fig. 1: *Flowchart depicting determination of transmission pairs from recruited participants. "PC" refers to the population cohort used to evaluate the trial's effect on HIV incidence, and "HCFs" to individuals recruited at health care facilities in the trial communities.*

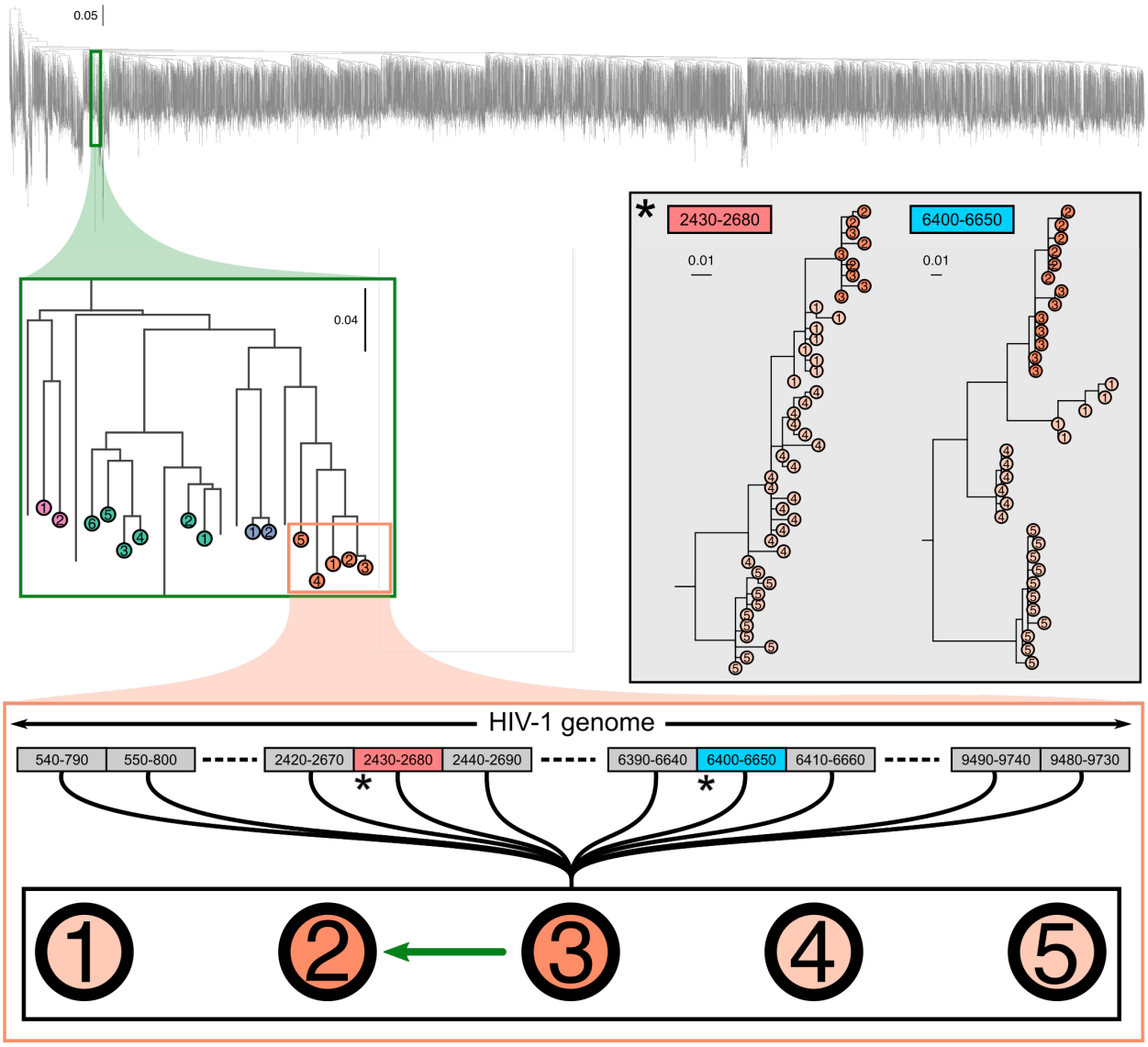


Figure 2: *Illustration of the selection of transmission pairs. The full phylogeny of the 5,613 consensus sequences with at least 50% coverage (top) has a subtree enlarged (green). Within that subtree, coloured tip shapes indicate members of four separate HIV-TRACE clusters. A single cluster (orange) is further enlarged. A separate phylogeny of within-host tips was constructed for each of 898 genomic windows; two examples for windows 2430-2680 and 6400-6650 are presented (with reference sequences pruned). In these windows (asterisk, inset), and when all windows were summarised (bottom), orange host 3 was*

reconstructed as the source of the infection in orange host 2. The other three members of the orange cluster (1,4,5) were not identified as belonging to any directed transmission pair. Branch lengths in all phylogenies are in substitutions per site.

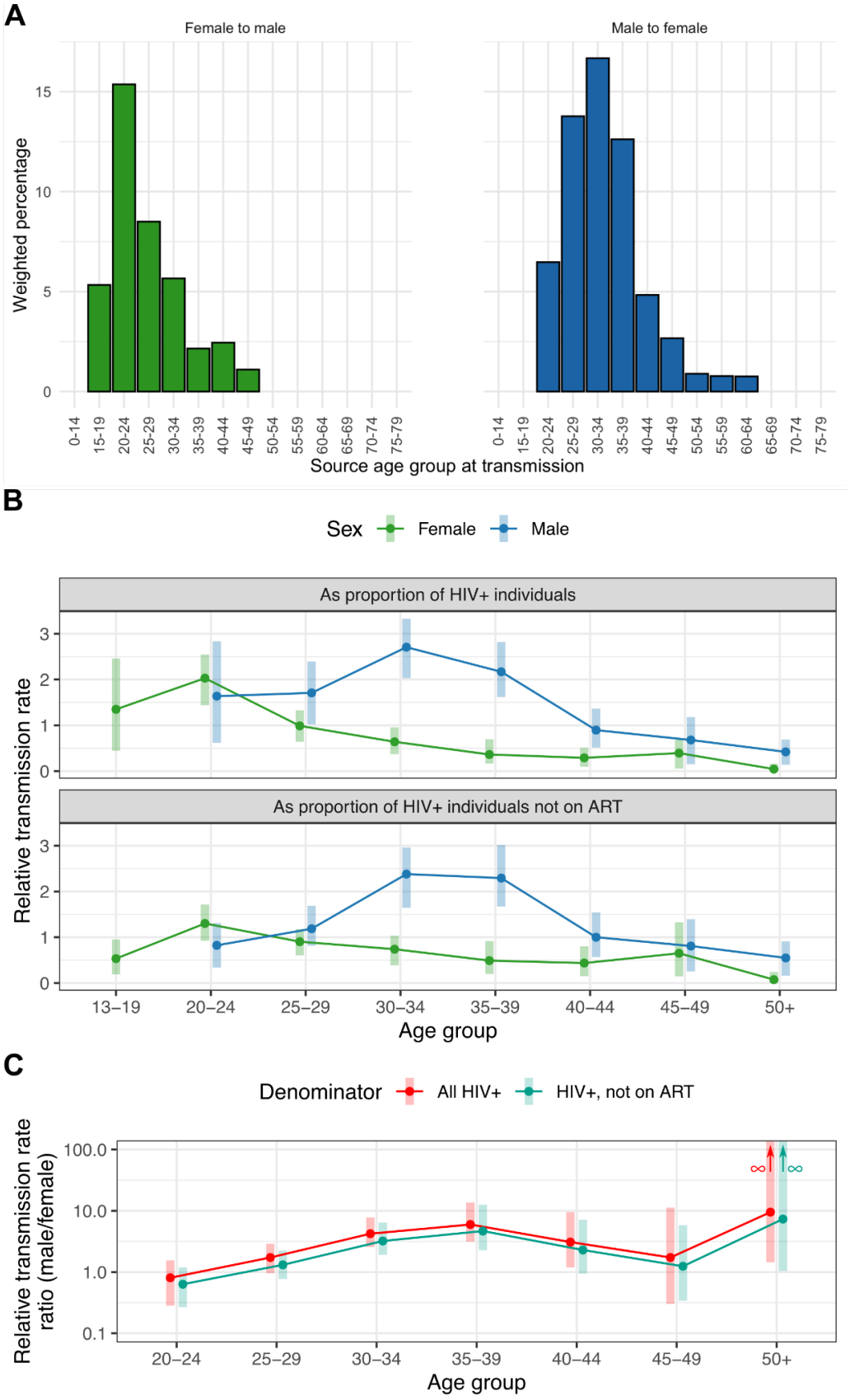

A
Female to male
Male to female
Weighted percentage
0
5
10
15
0-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
Source age group at transmission
B
Sex
Female
Male
As proportion of HIV+ individuals
As proportion of HIV+ individuals not on ART
Relative transmission rate
0
1
2
3
13–19
20–24
25–29
30–34
35–39
40–44
45–49
50+
Age group
C
Denominator
All HIV+
HIV+, not on ART
Relative transmission rate ratio (male/female)
0.1
1.0
10.0
100.0
∞
∞
20–24
25–29
30–34
35–39
40–44
45–49
50+
Age group

Figure 3: ***(A)*** *Age profile of sources (at the estimated time of transmission) of individuals estimated to have been infected during the trial period. Counts were weighted to make the recipient population representative of the population infected during the trial, using the PopART model for birth cohort, sex and community, and the 2018 Zambia demographic survey for marital status.* ***(B)*** *Contribution of sources of infection by age group and sex relative to the fraction of HIV+ individuals in the population in the same demographic group. Top: relative to all HIV+ individuals. Bottom: relative to HIV+ individuals not on ART. Ages are calculated as of July 2017. Shaded bars are 95% confidence intervals determined by non-parametric bootstrap. We identified no male sources in the 13-19 age group and thus this estimate is omitted.* ***(C)*** *Ratios of the relative contributions of male and female sources by age group (as calculated above). Shaded bars are 95% confidence intervals as before. The bootstrapped confidence intervals for the 50+ age group extend upwards to infinity.*

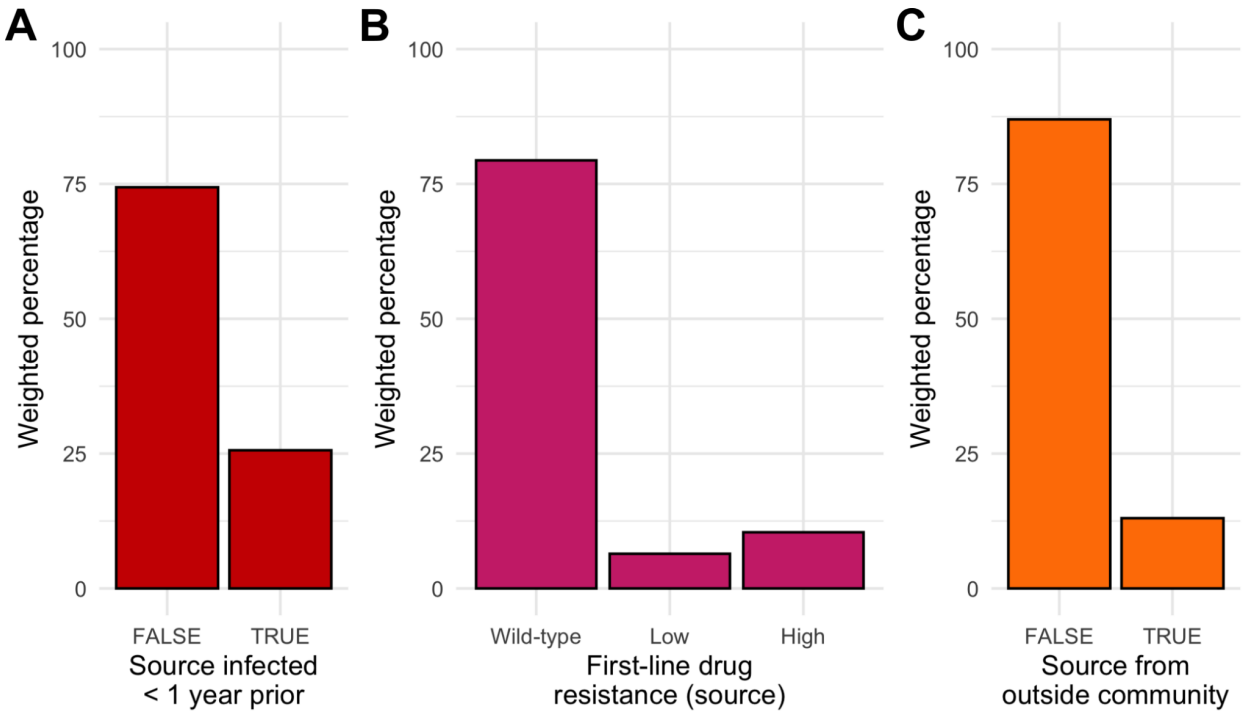
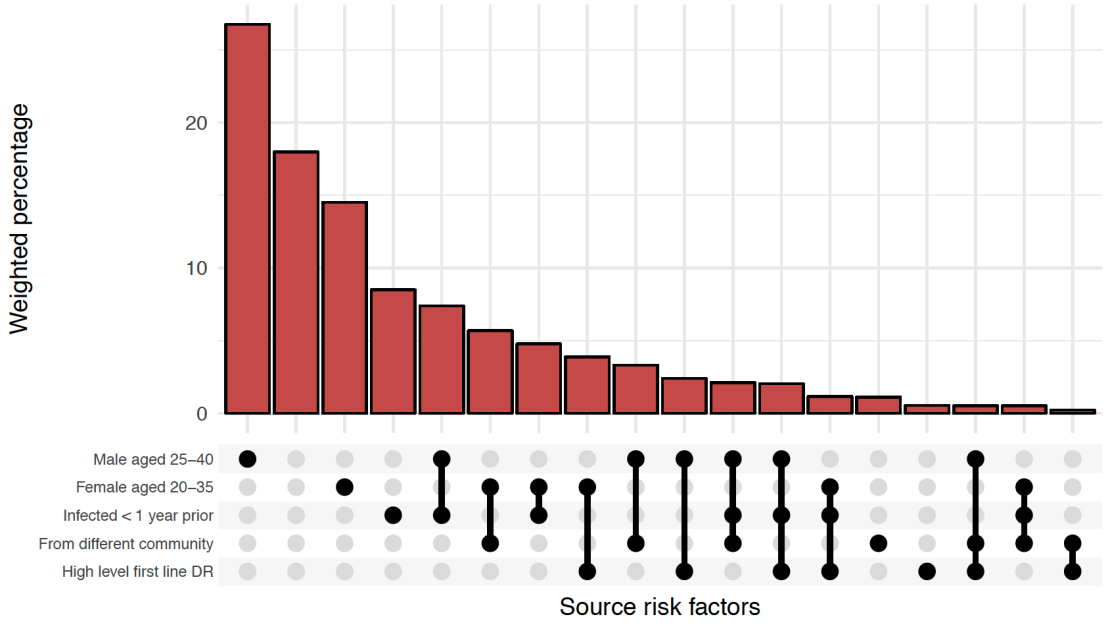
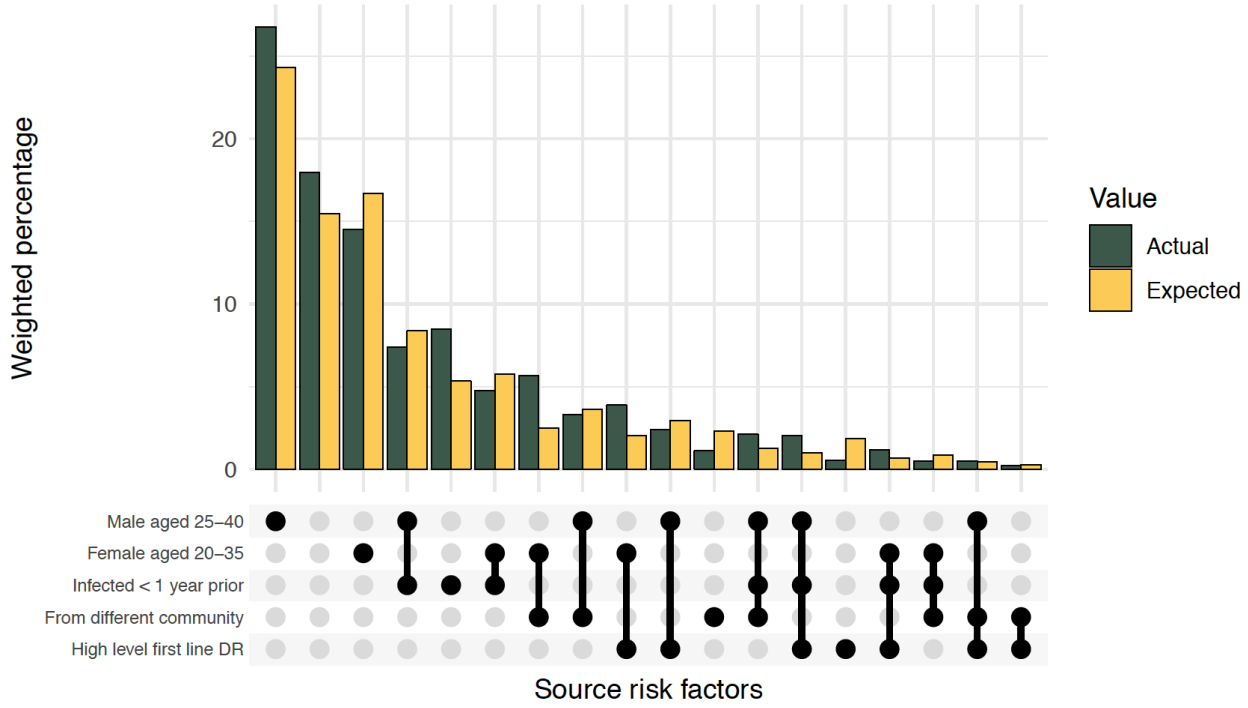


Figure 4: *Characteristics of source individuals in the set of directed opposite-sex pairs determined to have been infected during the trial period. Counts were weighted to make the recipient population representative of the population infected during the trial, using the PopART model for birth cohort, sex and community, and the 2018 Zambia demographic survey for marital status. (A) Recency of source infection (source estimated to be infected less than 1 year before recipient). (B) First-line drug resistance profiles. (C) Infections caused by sources residing in a different community from the recipient.*

A


Weighted percentage
20
10
0
Male aged 25–40
Female aged 20–35
Infected < 1 year prior
From different community
High level first line DR
Source risk factors

B


Weighted percentage
20
10
0
Value
Actual
Expected
Male aged 25–40
Female aged 20–35
Infected < 1 year prior
From different community
High level first line DR
Source risk factors

C

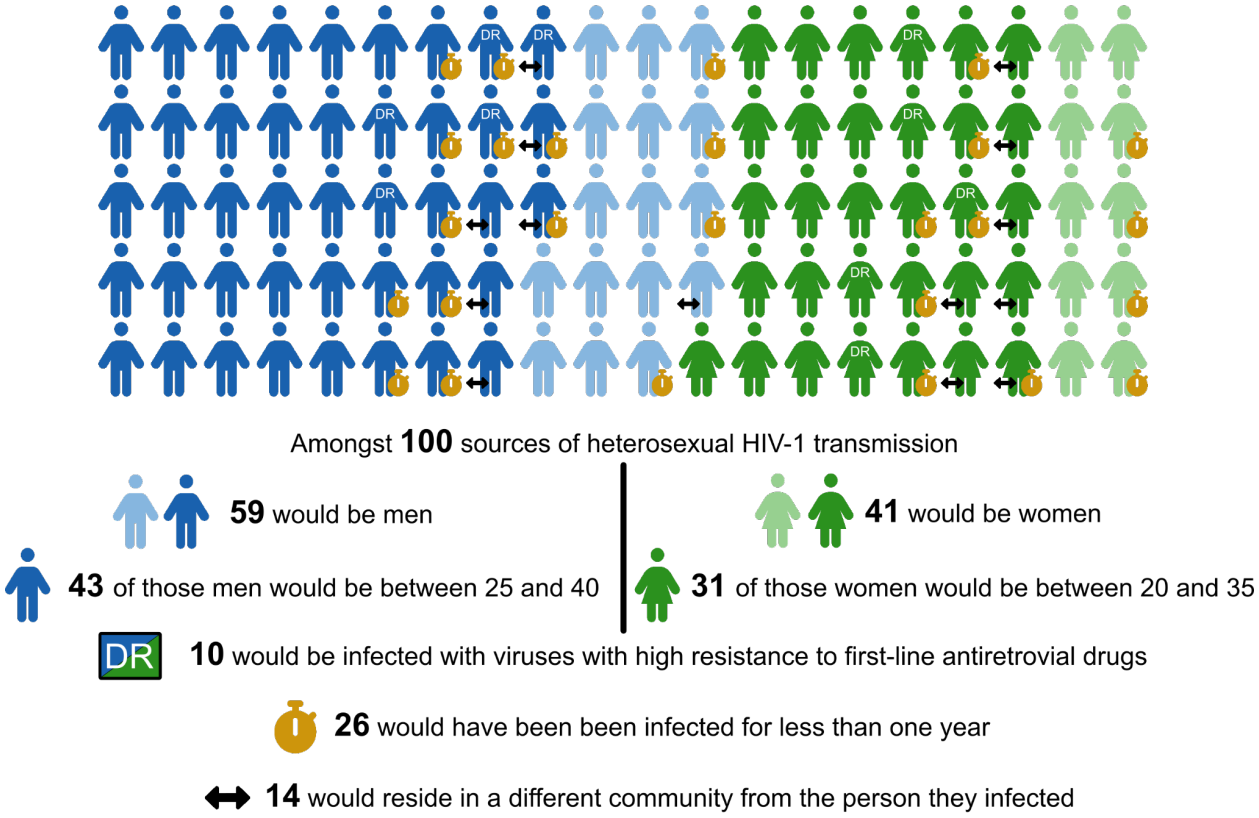


Figure 5: *(A) Distribution of all combinations of five key risk factors amongst the set of directed opposite-sex pairs determined to have been infected during the trial period. Each bar represents a group of sources whose characteristics are defined by the black dots in the bottom rows. Bars are weighted as in previous plots.* ***(B)*** *Actual distribution of risk factors compared to the distribution expected if they had been allocated randomly to sources based on risk factor percentages identified in the source population.* **(C)** *Pictorial representation of the expected characteristics of 100 sources of transmission. Note that as all transmission is assumed to be heterosexual, the 59 men and 41 women will have infected, respectively, 41 men and 59 women.*