The age and sex dynamics of heterosexual HIV transmission in Zambia: an HPTN 071 (PopART) phylogenetic and modelling study

M. Hall*[1,2], W. Probert*[2], L Abeler-Dörner[1,2], C. Wymant[1,2], F. Di Lauro[1,2], X. Xi[3], R. Sauter[2], T. Golubchik[2,4], D. Bonsall[1,2], M. Pickles[5,6], A. Cori[5], J. Bwalya[7], S. Floyd[8], N. Mandla[9], K. Shanaube[7], B. Yang[9], P. Bock[9], D. Donnell[10], M.K. Grabowski[11], D. Pillay[12], A. Rambaut[13], O. Ratmann[3], S. Fidler[14], H. Ayles[7,15], R. Hayes[8], C. Fraser[1,2], the PANGEA-HIV consortium and the HPTN 071 (PopART) study team

[1] Pandemic Sciences Institute, University of Oxford, Oxford, UK

[2] Big Data Institute, University of Oxford, Oxford, UK

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

[4] Sydney Infectious Diseases Institute, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, NSW, Australia

[5] MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College, London, UK

[6] HIV Prevention Trials Network Modelling Centre, Imperial College, London, UK

[7] Zambart, Lusaka, Zambia

[8] Department of Infectious Disease Epidemiology and International Health, London School of Hygiene and Tropical Medicine, London, UK

[9] Desmond Tutu TB Centre, Stellenbosch University, Cape Town, South Africa

[10] Fred Hutchinson Cancer Research Center, Seattle, US

[11] Bloomberg School of Public Health, Johns Hopkins University, Baltimore, US

[12] Division of Infection and Immunity, University College London, London, United Kingdom

[13] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

[14] Department of Infectious Disease, Faculty of Medicine, Imperial College, London, UK

[15] Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK

* These authors contributed equally

ABSTRACT

While much progress has been made in reducing the incidence of HIV-1 infection in sub-Saharan Africa in recent years, bringing the epidemic to an end will require identification of the demographic groups that continue to contribute to transmission. Pathogen phylogenetics and individual-based mathematical models (IBMs) of transmission are approaches that enable researchers to explore such questions. Here, we used both methods to characterise the ages and sexes of the individuals involved in heterosexual transmission in the context of the HPTN 071 (PopART) trial in Zambia. The results were concordant, and show that the male partner was on average older than the female by less than seven years, with larger age gaps in male-to-female than female-to-male transmissions. We found that the largest gaps for female recipients were amongst the youngest of those recipients. Conversely, the youngest male recipients saw the smallest gaps. We further used the IBM to

demonstrate that transmission to new age cohorts first entering into sexual activity is driven predominantly by male-to-female transmission. We also simulated the PopART universal testing and treatment intervention into the future to show that effective treatment of under-35-year-olds accounts for 93.8% of the reduction in incidence by 2039, while effective treatment of under-35-year-old men accounts for 62.1%. Finally, we simulated a one-year cessation of ART treatment for the whole population, which resulted in an immediate increase in the average age at transmission of both sources and recipients. With it becoming ever more expensive and difficult to find treatment-naive individuals and link them to care, targeted interventions for demographic groups such as under-35 men may be the key to finally ending HIV.

INTRODUCTION

The widespread rollout of HIV testing, antiretroviral treatment (ART), and voluntary medical male circumcision (VMMC) in combination prevention programmes has resulted in considerable declines in both AIDS deaths and new HIV infections (UNAIDS 2016). ART is highly effective in preventing HIV transmission as well as clinical illness (Cohen et al. 2016), and recent dolutegravir-containing regimens have reset the clock with regards to antiretroviral resistance, whose prevalence had recently been increasing (Fogel et al. 2022). However, questions remain concerning in which demographic groups transmissions are still occurring and how prevention interventions can best reach the groups that contribute most to new infections; and also whether, in the light of rapid population growth in sub-Saharan Africa (UNAIDS 2018), these groups will grow as proportions of the population.

Heterogeneities in the epidemic are well known (Balkus et al. 2015; de Oliveira et al. 2017; Harling et al. 2014; Magosi et al. 2022), including variations in HIV acquisition risk by age and sex. Age and sex dynamics of heterosexual sexual relationships in SSA, regardless of HIV status, have also been explored (Gregson et al. 2002; Kelly et al. 2003; Ott et al. 2011), usually finding the male partner to be on average several (five or less) years older than the female. Less is understood about which age and sex groups are most likely to transmit HIV, and the age gap between sources and recipients of transmission. This is partially because it is harder to identify sources of transmission than it is to identify recipients. Sources can be identified using traditional epidemiological survey methods, including recruitment of household pairs into a study (Mujugira et al. 2011), characterisation of the sexual contacts of recipients by contact tracing (de Arazoza et al. 2007) or interviewing those recipients (Balkus et al. 2015; Akullian et al. 2017). These can be intensive endeavours, suffering from both recall (Napper et al. 2010) and social desirability (Turner et al. 2009) bias.

Identifying heterogeneities in HIV transmission by age and sex may be critical in informing the next steps of HIV prevention. For example, testing programmes typically yield fewer new diagnoses for men than for women, with consequences for the number of men linked to care. However, programmes aiming to end new heterosexual infections cannot ignore or de-emphasise the role of men in transmission, and must concern themselves with the male population. The risk of being a source of HIV infection will not be evenly distributed amongst males. If subpopulations disproportionately driving new infections can be identified, then the

focus and impact of interventions could be enhanced by tailoring services to meet those men's needs.

Survey-based research into heterogeneities in age and sex in HIV transmission in SSA has produced varied and sometimes contradictory results. A consistent observation, in many studies conducted during the last two decades, has been that a large age gap with a partner is associated with increasing HIV acquisition risk in (usually young) women (Gregson et al. 2002; Pettifor et al. 2005). However, some recent studies did not observe a simple increase in risk with increasing age of the male partner (Akullian et al. 2017; Balkus et al. 2015; Harling et al. 2014; Street, Reddy, and Ramjee 2016), suggesting that a new epidemic environment with widespread ART use has changed transmission dynamics (Harling et al. 2014). In this environment, partnerships most likely to result in HIV infection would be with someone who is old enough to have acquired HIV, but has yet to be diagnosed and started on ART. The most recent studies, however, did find the original association (George et al. 2022; Topazian et al. 2020). Studies of this sort in men are much less common, but those that have been conducted show that, similarly, the risk of HIV acquisition is increased for men with older or same-age partners when compared to those with younger partners (Beauclair et al. 2016; Gregson et al. 2002). Regardless of recipient sex, characterisation of the demographics of sources in this study design is generally rudimentary, with even their ages often simply derived from recipient interviews.

Individual-based mathematical models (IBMs) (Liesenborgs et al. 2019; Bershteyn et al. 2018) and pathogen phylogenetics (Wymant et al. 2017; Kosakovsky Pond et al.

2018) are alternative methods to surveys for describing characteristics of sources of transmission. IBMs model entire transmission networks amongst simulated individuals with the characteristics of interest. In pathogen genomics, recent methodological developments have shown that “deep sequence” analysis of within-host viral genetic diversity can identify pairs of individuals whose HIV infections likely represent direct transmission, as well as the direction of that transmission (Romero-Severson, Bulla, and Leitner 2016; Leitner and Romero-Severson 2018; Romero-Severson et al. 2017). For example the method *phyloscanner (Wymant et al. 2017)*, validated using data on known sexual partnerships (Zhang et al. 2021; Xi et al. 2022; Ratmann et al. 2020), has been used to investigate transmission patterns through the identification of transmission pairs.

A previous IBM study (Akullian et al. 2020) predicted that increasing ART rollout would, due to reduced force of infection, have an “ageing” effect on the epidemic, with the average age of both sources and recipients of transmission expected to increase over time. Amongst phylogenetic studies, de Oliveira et al (de Oliveira et al. 2017) suggested a cycle of transmission in South Africa in which women under 25 are infected by markedly older men, while female-to-male transmission occurred in older individuals of similar ages. Significantly, they did not have access to a method to reconstruct the direction of transmission using individual-level data, and instead assumed that it was determined by the prevalence gradient in the relevant communities (Grabowski and Lessler 2017). More recently Magosi et al. (Magosi et al. 2022) used more sophisticated methods to infer direction of transmission (Wymant et al. 2018) and did not find a similar asymmetry in age differences in Botswana. Instead, they found the age gap in transmission pairs to be small, albeit

with large uncertainty. A similar age gap has been observed in heterosexual transmission in Europe (Kusejko et al. 2018). Monod et al (Monod et al. 2024), working with data from the Rakai Community Cohort Study in Uganda identified a significant upward shift in the age of both partners in male-to-female transmissions between 2003 and 2018, leading to a declining proportion of new infections occurring amongst women aged 15-24 and an increasing one amongst those aged 25-34.

The early months of 2025 saw a sudden suspension of the President's Emergency Plan for AIDS Relief (PEPFAR). The programme was responsible for 49% of international spending, and 34% of all spending, on HIV/AIDS in SSA during 2024 ("UNAIDS HIV Financial Dashboard," n.d.). As of the time of writing its future remains uncertain. A suspension, even if temporary, of such a crucial component of existing HIV care in SSA would likely have a profound impact on health outcomes in the region. ART prevents both progression to AIDS and onwards transmission of the virus. As a result, a cessation or pause in viral suppression for the millions of HIV-positive individuals who have relied on treatment provided under PEPFAR would result in not only a significant increase in the number of individuals progressing to AIDS, but also in incidence of new HIV infections. There is an urgent need to understand the epidemiology of HIV in SSA as we move into a new era (Nichols et al. 2025).

Here, we provide a comprehensive characterisation of the age- and sex-specific transmission dynamics during the large HPTN 071 (PopART) trial in Zambia using two independent methods: an IBM and pathogen genomics. We use these to investigate the potential impact of future interventions focussed on suppressing

transmissions from inferred source populations. We conclude that there is a substantial opportunity to achieve substantial reductions in HIV incidence by improving testing, treatment, and ART adherence of men aged 25-34. We also investigate the effect of an immediate pause of one year of ART provision on these dynamics, finding that it would cause a sudden and dramatic increase in the age of infection for both sexes.

RESULTS

Data

HPTN 071 (PopART) was a cluster-randomised trial of a combination HIV prevention package involving universal testing and treatment in 21 communities, comprising one million individuals, in Zambia and South Africa. It was conducted between 2014-2018 (Hayes et al. 2014, 2019). The intervention involved annual household visits from community health workers offering HIV counselling and rapid testing, followed by linkage to care for HIV-positive clients. The trial had three arms, with arms A and B testing variations of the package and arm C serving as a control arm; for full details see the trial paper. To estimate intervention impact, HIV incidence was measured in a cohort of around 2,000 randomly selected individuals in each community, called the Population Cohort (PC). The IBM ‘PopART-IBM’ was developed as part of the trial to simulate the epidemic and interventions to help understand trial findings and extrapolate them to different settings and timescales (Pickles et al. 2021; Probert et al. 2022). The IBM was calibrated to various sources of age- and sex-stratified data from the trial. In simulating the whole epidemic, the IBM generates a large number of

transmission pairs on a timescale spanning the entire HIV pandemic; for comparability with the phylogenetically identified pairs (described next), from these we selected for analysis a subset which would be eligible to be sampled as part of the phylogenetic study. Across the 1000 replicates of the simulation (see Methods) there were a median of 213,560 transmissions with recipients in this set, with a 95% highest density interval (HDI) of (191,652, 231,239).

An ancillary study to the main trial—the HPTN 071-02 (PopART) Phylogenetics Study—was conducted as part of PopART in the Zambian PopART communities only. The study used phylogenetics to identify and characterise sources of HIV transmission in these communities, thus informing modelling of the impact of different intervention scenarios. For this purpose, genomic deep-sequencing was performed on samples collected over the course of the trial from both HIV-positive individuals in the PC and HIV-positive individuals attending health-care facilities in the study communities who were not on ART at the time of attendance. While enrollment in the former was capped at 45 years of age, there was no upper age limit for inclusion in the latter. Analysing the resulting sequence data with *phyloscanner*, we identified 355 opposite-sex transmission pairs with high confidence in the direction of transmission.

The transmission pairs identified phylogenetically and those simulated from the IBM give us two independent means of characterising transmission during the trial.

Two independent methods agree on sex and age differences in transmission

Male-to-female transmissions made up 56% (198/355) of the phylogenetically identified pairs and 57.9% of the IBM pairs (95% highest posterior density interval, HDI, 56.3% – 59.7%). This is in contrast to males making up, at the start of 2017 in the IBM, 41.6% (39.9%-43.3%) of infected individuals and 44% (42.2%-45.8%) of virally unsuppressed individuals. To estimate the age at transmission of both partners in each phylogenetic pair, we used an estimate for the time of infection of the recipient derived from the HIV-phyloTSI method (Golubchik et al. 2022), together with the dates of birth of both partners. The age, sex, and other characteristics of the phylogenetic pairs are summarised in table 1. Figure 1 shows the distribution of source and recipient ages at the time of transmission, by sex, as estimated by both the IBM and the phylogenetic data. Table 2 shows the median ages, and their interquartile ranges, by sex and direction of transmission in the pairs identified using both methods.

Estimates of the distribution of transmission age gaps (source minus recipient) are shown in figure 2 and summarised in table 2. For phylogenetics there was weak evidence of a difference in the gap between the male and female partners' ages between male-to-female and female-to-male pairs (*t*-test p=0.069). (IBM posterior distributions for the mean and median values of these two age gaps did not overlap at all.) A more detailed illustration of the transmission dynamics can be seen as a who-aquires-infection-from-whom (WAIFW) matrix in figure 3.

The age gap increases with recipient age in female-to-male pairs and decreases in male-to-female pairs

The IBM provides a full transmission tree from which the complete dynamics of transmission can be explored; as inclusion in the dataset is automatic, biased sampling is not an issue. With the phylogenetic results providing independent support for its accuracy, we proceeded to explore its outputs in more detail. Figure 4 shows alluvial plots for the transmission history of individuals who became infected with HIV in the year 2016 in the best-fitting IBM simulation. Orange and green alluvia show viral flow to younger and older groups respectively. For female-to-male transmissions we see that these flows are almost equal, whereas male-to-female transmission are overwhelmingly orange i.e. to younger ages. Thus this shows that there is a clear shift in age for male-to-female transmissions which is absent for female-to-male transmissions, acting to “rejuvenate” the epidemic.

As this finding seemed contrary to the fairly consistent central estimates of the age gap described above, we stratified the gap by recipient age (figure 5a,b). This showed distinctly different patterns for the two directions. In female-to-male transmission pairs the average age gap for the youngest recipients actually had the female partner older, but this rapidly reverses as recipient age increases, with the gap decreasing in a roughly linear fashion. For male-to-female pairs there is also a decline in the gap with increasing recipient age, but the decline levels out in a fashion appearing to be non-linear. These trends are similar in the phylogenetics data (figure 5c,d), although the magnitude of the age gap for the youngest female recipients is larger, with the median gap in women less than 21 being 10.0 years, compared to 6.58 years in the best-fitting IBM simulation.

Age patterns remain stable over time despite steep increases in ART coverage

We next investigated whether a shift in age patterns in the IBM occurred as a result of increasing use of ART (Harling et al. 2014). Figure 7 shows changes in source age, age gap, and proportion of individuals on ART over time for the period 2005–2020 in the best-fitting simulation. Trends are minor or non-existent in the former two variables, with the largest change being in the age of female sources, which decreases from a median of 27.4 years (IQR 22.9–33.6) in 2005 to 26 (22.2–31.6) in 2020. This suggests that age gaps in transmission were largely unaffected by ART.

Rollout of the PopART intervention to individuals aged 35 or under is projected to reduce incidence by 67% by 2040

We used the IBM to quantify the effect of targeting future interventions at different demographic groups. We started from the model state of the entire population of the control communities (those which were not subject to any part of the trial intervention and received standard of care) at the end of 2019 (Pickles et al. 2021). We considered alternative hypothetical scenarios where the PopART intervention was used as a proxy to deliver real-world intensive linkage to care programmes to specific demographic subgroups. We focus a) on younger age groups, since our data shows that they are more likely to be unsuppressed than older age groups living with HIV and b) on younger men, who were shown to drive the rejuvenation of the epidemic (figure 4). We modelled the effect of four different scenarios, starting in the year 2020:

1. Applying the PopART intervention to the entire population

2. Applying the PopART intervention to only the population aged between 18 and 34, with standard of care for other age groups
3. Applying the PopART intervention to only the male population aged between 18 and 34 , with standard of care for other male age groups and females
4. Standard of care

Figure 7 shows the effects of these four scenarios on HIV incidence, by two-year period from 2018–2019 (prior to the introduction of the intervention). The posterior median incidence in 2018–2019 was 0.8 cases per 100 person-years at risk (95% HDI 0.78–1.37). By 2038–2039, overall incidence was projected to decline to 0.42 (0.18–0.71) even in the absence of any PopART intervention. The PopART intervention, when provided to the entire population, further reduces this to 0.1 (0.004–0.26). However, the effect of deploying it only to individuals aged 18-34 has an only slightly smaller effect, with a posterior median incidence of 0.12 (0.014–0.29), while providing it only to men aged 18-34 gave an incidence of 0.22 (0.05–0.44). By the end of 2039, the former intervention achieved 93.8% (65.8%-126.6%) of the reduction between the “standard of care” scenario and full deployment (the stochastic nature of the IBM meaning that the 18-34 intervention could outperform the full intervention in some cases), and the latter 62.1% (23.2%-92.1%). The 5.8% (4.7%-7.1%) of the population that remained virally unsuppressed in 2018-19 would be reduced to 0.37% (0.072%-0.76%) by 2038-39 under the full intervention, to 0.59% (0.023%-1.04%) when deployed only to those aged 18-34 and and to 1.14% (0.45%-1.85%) when deployed only to males 18-34.

Modelling a one-year gap in ART provision

Finally, again starting with the model state in the control communities, we modelled a one-year complete pause in ART provision for the year 2025. Following this, re-engagement with care took place following a subsequent positive test; no deployment of additional resources or targeted interventions to re-integrate individuals with care was modelled. We examined the effect of this cessation in ART on age and sex dynamics, which can be seen in figure 8. The pause is responsible for a substantial increase in incidence (fig 8a), with a posterior median of 1554 new cases in 2024 increasing to 5715 in 2025. Regarding age and sex dynamics (figure 8B), the immediate effect was an increase in the ages of all those involved in HIV transmission. The change in posterior median mean age of all individuals involved in transmissions from 2024 to 2025 was 4.9 years (3.53-6.21) for female recipients, 5.62 (4.27-7.28) for male recipients, 5.56 (4.34-6.99) for female sources, and 6.43 (4.84-7.93) years for male sources. There was also a jump in the age gap in male-to-female transmissions, going from a posterior median mean gap of 4.9 years (3.84-5.82) in 2024 to 6.61 (5.43-7.67) in 2025. No similar effect was seen in female-to-male pairs. After the initial change in 2025, the curves in figure 8 can be seen slowly returning to the patterns of the years 2019-2024, but in the absence of specific strategies to re-engage the HIV-infected population with care, this is not immediate once ART provision is re-established.

Sensitivity analyses

Appendices 1 and 2 describe two supplemental analyses of the phylogenetics data. First we confirmed that the analysis was genuinely informative, differing from results

obtained by choosing pairs of individuals at random as “transmission pairs”. Second we performed a detailed sensitivity analysis of the effects of varying the settings for the *phyloscanner* package. While the sets of transmission pairs obtained using different settings were somewhat variable, the exact choice of pair set had little effect on the main results of this study.

DISCUSSION

In this study we investigated the age and sex patterns of heterosexual HIV transmission in Zambia during the HPTN071 (PopART) trial between 2014 and 2018, using both an individual-based mathematical model (IBM) of the epidemic and pathogen phylogenetics. We focussed on characterising transmission pairs, i.e. jointly analysing sources and the recipients to whom these individuals transmit.

Previous analyses have characterised the population of sources, i.e. the subset of people living with HIV who transmit onwards (Magosi et al. 2022; Monod et al. 2024). For the HPTN071 (PopART) trial we previously reported that male sources were on average a few years older than female sources, that there was no disproportionately large contribution to transmission from external communities or from individuals with drug-resistant viruses, and that there was a disproportionately large contribution to transmission from men aged 25–34 (Hall et al. 2024). These men should be a focus of efforts to prevent onward transmission. Interventions are needed to find them, link and retain them in care, and encourage viral suppression, even if they are more

difficult and costly to reach. In that study we found that linking one man in this age group to care is likely to have a similar effect on incidence as linking two women.

Here, we further developed previous results by characterising transmission pairs, i.e. analysing sources and their recipients jointly, not independently. Notably we estimated the distribution of the age gap within a pair, separately for male-to-female and female-to-male transmission. Most transmissions in either direction had the male partner a few years older than the female (medians between 2.2 and 6.1 years depending on data source and direction). This age gap is similar to reports for the general population in the region (Akullian et al. 2017; Bacaër, Pretorius, and Auvert 2010; Smith et al. 2014; Wolock et al. 2021)

We also observed more complex dynamics (figures 5 and 6). For male recipients, at the youngest ages the age gap is near parity, with the female source being a median of 1.6 years older in the best-fitting simulation for males under 20. These men would tend to take a partner younger than themselves (judging by the patterns in older men). However, such partners, if sexually active at all, are unlikely to be HIV-positive because there is a delay, which we here refer to as the “exposure lag”, between the expected age of first potential exposure to HIV and the expected age of infection. As the age of male recipients increases, the age gap increases.

For female recipients, at the youngest ages the age gap is particularly large, with the male partner being a median of 6.7 years older for women under 20. The exposure lag means that sexual partners who are HIV-positive are older than the average partner, and in addition the contribution from similarly-aged and younger men is

reduced as few of those have reached sexual debut. As the age of female recipients increases, this initially large age gap decreases steadily until it reaches parity for recipients in their early 60s. Women form new partnerships with men on average a few years older than themselves, and this remains true throughout their life (Ott et al. 2011). Thus this decline in the age gap in transmission pairs is likely to be driven by a combination of ART treatment and AIDS-related mortality in older men. For middle-aged and older women, the population of potential male partners will thus be increasingly virally suppressed or HIV-negative. It should be emphasised that, while HIV prevalence increases with age, only a relatively small proportion of new infections occur in older age groups and the contribution of these to the epidemic are relatively minor; in the best-fitting IBM simulation only 4.57% of transmissions were from individuals aged 50 or greater.

A previous paper published by de Oliveira and colleagues (de Oliveira et al. 2017) interpreted transmission patterns in SSA as a cycle in which young women are infected by much older men (the “sugar daddy” phenomenon), then these women age and later transmit back to similarly-aged men. The study was based on phylogenetic evidence but did not include an analysis of the direction of transmission. Our results concur that the widest age gap in transmission to females occurs in the youngest recipient age group. Our best-fitting IBM estimate has male sources a median 5.9 years older than female recipients when the latter are under 25, compared to the mean gap of 8.7 years estimated by de Oliveira et al. The under-25 female age group makes up 50% of female recipients (and 29.4% of all recipients), indicating that this route plays a very substantial but not overwhelming role in overall transmission.

Our results diverge from those of de Oliveira et al. when it comes to female-to-male transmission, as we see the smallest age gaps for the youngest male recipients and a gradual increase with age. Men aged 18 were infected by women a median of 1.3 years older, but by age 25 this gap had the women 1.1 years younger, and by age 40, 7.2 years younger. Our picture of transmission does include a “rejuvenation” effect in which male partners transmit to slightly younger female partners, with female-to male transmissions tending to occur in the same age group. It remains true that having a much older partner poses an increased individual risk of HIV acquisition for a woman under 25, as seen in previous work (George et al. 2022; Topazian et al. 2020) An older partner has had more time to be exposed to HIV themselves, as evidenced by the phenomenon also existing for new infections in men (Beauclair et al. 2016). Another limitation of the de Oliveira et al approach was that collected participants into large age groups rather than consider age as a continuous variable, which may mask subtle dynamics of the age gaps in transmission, an observation also made by others (Xi et al. 2022; Bu et al. 2023)

We found that the increasing rollout of ART between 2005 and 2020 did not change age distributions of transmission, contrary to the suggestion of some studies (Akullian et al. 2017; Balkus et al. 2015; Harling et al. 2014; Street, Reddy, and Ramjee 2016). This was despite an increasing disparity in viral suppression by sex (Monod et al. 2024); in the best-fitting ART simulation only 0.38% more females were on ART than males in 2005, but this had risen to 6.3% by 2017 before declining somewhat to 5.7% in 2019.

We used two entirely independent methods, each with its advantages and disadvantages. The IBM allows us to simulate very large numbers of complete transmission chains, for which we can analyse all individuals with no sampling bias or erroneous identification of pairs. However, the results are dependent on model assumptions (for example, the age mixing matrix (Pickles et al. 2021)), and the individuals involved are simulated. The phylogenetics data are the reverse: there are no model assumptions and the data come from genuine cases of HIV infection, but the sample size, although large compared to similar studies, is several orders of magnitude smaller than the actual population, and inference may be affected by sampling bias and imperfect pair ascertainment. The results of the two methods being largely consistent is a key strength of this study: it gives confidence in both methods, and in additional estimates generated by the IBM for scenarios and periods for which no phylogenetic data exist. Nevertheless, some issues may affect both methods. Notably, neither of our methods attempt to quantify the role of men who have sex with men in transmission. They are not modelled in the IBM, and, as we analysed opposite-sex pairs only in the phylogenetic study, we do not consider their role there either. In addition, while the IBM includes heterogeneity in sexual risk taking behaviour it does not explicitly model key heterosexual populations, such as female sex workers. Membership of such groups is unknown for participants included in the phylogenetic study, and they might also be undersampled. Notably, there are subpopulations that were not modelled in the IBM — such as female sex workers and men who have sex with men — and which were likely undersampled in the phylogenetic study. Our estimates of the effect of interventions (figure 7) rely on these groups not contributing significantly to transmission. We previously estimated

that sex between men makes limited if any contribution to transmission within the population studied here (Hall et al. 2024).

While the decline in the age gap as a female recipient ages is seen in both methodologies, the magnitude of the gap for the youngest women is the main area in which the two methods differ (figure 5), with larger gaps seen in the phylogenetics. It is not entirely clear why this might be. Potentially relevant limitations in the data used to fit the IBM are the absence of trial data for under-18s and the lack of care cascade information from before 2013. The omission of vertical transmission from the model may also be pertinent. Alternatively, there might have been a bias towards finding older-than-average partners for younger women in the phylogenetics data. Despite this disparity, the similarity between the trends in age gaps identified with both methods is striking. The age gap between a female recipient and her male partner at the time of transmission is largest for the youngest women and declines asymptotically towards equality as the recipient's age increases. The youngest male recipients, on the other hand, are infected by a female sources of a similar age to themselves. With increasing recipient age, this gap subsequently widens in the direction of relatively younger source ages.

Monod et al (Monod et al. 2024) modelled a different but related counterfactual scenario. Where we simulated the long-term impact of an intervention based on real-world implementation data, where the hypothetical element is targeted at different demographic groups, they simulated an alternative hypothetical intervention that would either fully close the gap in viral suppression between men and women, half close it, or achieve the UNAIDS 95:95:95 treatment targets in males. Both sets of

scenarios predicted that considerable benefits in reducing female incidence could be achieved by targeting a relatively small number of HIV-positive men. This undertaking would be much less intensive than targeting the much larger number of HIV-negative women at high risk of acquisition.

A limitation of this study is the narrow geographical catchment area of the PopART trial. The results should not be automatically extrapolated to apply to other SSA settings, and indeed the findings of Monod et al. (Monod et al. 2024) in Uganda showed a bimodal distribution in the age of male sources that we did not identify here. We also do not control for bias in the composition of the phylogenetic dataset, both in regard to demographics and to the ability to successfully acquire a sequence from a participant providing a sample, which we assume is random. Nevertheless, the consistent results between the two methodologies should assuage fears about this, as the IBM results are not subject to sampling bias problems.

With the 95:95:95 targets being in reach in many parts of SSA, it is becoming increasingly expensive to link individuals to care using population-wide testing and messaging campaigns. There is widespread agreement that future prevention campaigns need to focus on the remaining 5:5:5 of the population. We estimated that an intensive effort to link people to care, such as the PopART intervention, would result in almost as much reduction in incidence if focussed on those aged between 18 and 34, and would retain two thirds of the benefit when focussing on men between 18 and 34 only. We did not methodically explore potential age cut-offs in this work, and the identification of the optimal demographic groups to target goes beyond the scope of our analysis. Nevertheless, our results show the potential

benefits of focussing on groups who do contribute disproportionately to transmission. Enabling and encouraging younger men to seek and stay in care to the same extent as women would improve the health and life chances of many men, and rapidly reduce infection rates in their partners.

A major disruption in ART provision as a result of changes in the global aid landscape is now underway, with the endpoint unclear. As ART prevents both disease progression and transmission, simultaneous increases in the incidence of AIDS and of HIV infection is to be expected, and the response of the global health community will need to focus on both. In the latter case it is important to anticipate not only an increase in new infections, but in what populations they would occur. Our results suggest a shift in incidence to older age groups, caused by the withdrawal of ART from those previously virally suppressed, at which point their partners, both existing and new, are at greatly increased risk of HIV infection. This effect would linger even after the restoration of ART provision unless procedures were put into place to re-engage those individuals with care.

METHODS

HPTN071 (PopART) trial

The HPTN071 (PopART) trial has been described in detail previously (Hayes et al. 2014, 2019). In summary, the impact of a universally delivered combination prevention package was tested in 21 communities in Zambia and South Africa, between 2013 and 2018 (Hayes et al. 2014, 2019). The total population of the 21

study communities was approximately one million. The intervention comprised annual rounds of home-based HIV testing, including support for linkage to care and retention on ART, promotion of medical male circumcision for HIV-negative men, and additional HIV, tuberculosis (TB), and sexually-transmitted infection (STI) services. It was delivered by teams of Community HIV-care providers (CHiPs). The 21 communities were arranged in seven triplets, matched on baseline HIV prevalence, engagement with care, and geographic location (Hayes et al., 2019). Communities were randomly allocated into one of three arms within each triplet. Arm A communities received the full combination HIV prevention package and offer of immediate ART for individuals that tested HIV-positive. Arm B communities received the full combination HIV prevention package with ART initiation following national guidelines (where ART initiation was dependent upon an individual's CD4 count). Arm C communities received the standard of care. National guidelines on ART eligibility changed over the period of the trial, and thus from late 2016, ART for all HIV positive individuals was standard care in both countries (aligning arms A and B). The primary endpoint was the relative reduction in HIV incidence between arm A or B communities and their matched arm C community between 12 to 36 months after trial start. This was measured in a cohort of 2000 randomly selected individuals aged 18-44 per community, called the Population Cohort (PC), who were followed up in four annual rounds. Rounds of the PC are referred to as the number of months after trial start (e.g. PC0 is 0 months after start).

IBM

For each of the 21 communities in the HPTN017 (PopART) trial, the IBM (Pickles et al. 2021) models every individual in a growing population that reaches approximately 50,000 in 2014, roughly matching the actual community sizes at that time point. Demographic parameters follow historical estimates of mortality and fertility from the UN Population Division (United Nations, n.d.), and future projections of these measures follow the UN's intermediate estimate (United Nations 2015). HIV infections are introduced into the model in 1980, HIV transmission is modelled in serodiscordant couples, and HIV disease progression (in the absence of ART), follows data from the ATHENA cohort in the Netherlands (Cori et al. 2015).

Individuals were categorised into one of three groups of sexual risk-taking behaviour, where membership of a higher risk group reflects both a higher propensity to form partnerships and an increased maximum number of concurrent partnerships. An individual's risk group is assumed not to change through time. Partnerships are formed based upon an individual's age, sex, and sexual-risk taking behaviour, with sexual risk-taking behaviour acting independently of age and sex. Partnership formation and dissolution rates are calibrated using survey data from the PC, and incorporate assortativity by risk group; see (Pickles et al. 2021) for full details. Age mixing was determined using the age of the last reported partner (within the last year) based upon extensive sexual behaviour surveys. To allow for transmissions from areas outside the trial communities, the IBM models two "patches" representing the community and the surrounding area. Partnerships may also form with individuals in the surrounding area, but at a reduced rate (45% of the rate within the community). When serodiscordant, these between-community partnerships are also

at a lower (29% of within-community) risk of HIV transmission, reflecting higher rates of condom usage reported in the PC data.

Testing and linkage to care were modelled as two separate processes, a "background" component representing testing outside the context of the trial (e.g. through clinics and antenatal care) and a "CHiPs" component representing the trial intervention, which was disabled for the control communities. The subsequent care cascade following a positive test was determined by the component from which the test came; for full details see (Pickles et al. 2021).

Parameters for partnership formation and dissolution were set using data from the PC. Other parameters were estimated using Approximate Bayesian Computation (ABC) to fit the model to five sources of data, all stratified by age and sex: 1) HIV prevalence from the Zambian Demographic and Health Surveys (DHS) of 2002, 2007 and 2013; 2) HIV prevalence from the final intervention round of the trial; 3) the proportion of individuals aware of their status among HIV-positive individuals from all three intervention rounds; 4) the proportion of individuals on ART among those who are HIV-positive and aware of their status from all three intervention rounds; and 5) the proportion of individuals virally suppressed amongst all HIV-positive individuals at PC24. For all but 1), which was taken from previous work, fitting used trial data from an intervention community (arm A) in Zambia. The calibration framework fits the epidemic selecting 1000 most likely parameter sets, each of which is used to generate a single simulation. An ABC distance measures distance between observed and simulated summary statistics used in the fitting process.

Point estimates for key quantities presented here are the posterior median and 95% highest posterior density (HPD) interval over 1,000 ABC simulations. We also identified the ‘best-fitting’ IBM simulation - the model replicate whose ABC distance was smallest - and used the corresponding transmission tree when we presented results that require a single realisation of the simulated epidemic.

The IBM was used to model transmission over a 75-year period from 1975 to 2050. To make the outputs comparable to the phylogenetics data, we selected only transmissions for which the recipient would have been eligible to provide a viral sample during the trial, specifically that they were infected with HIV before 2019 and alive in November 2013. (We did not model the viral dynamics that would result in the failure of a HIV-positive individual to provide a viable sequence; effectively we assume that the probability of this happening is independent of participant demographics or time of infection.) We also used only model output from the communities for which extensive phylogenetic data were available; this excluded the entire South African component of the study as well as one triplet of communities in Zambia for which health-care facility phylogenetic data were not collected. All the simulations taken forwards for full analysis were from the nine remaining Zambia communities. Transmissions involving individuals in both the “inside” and “outside” patches of the simulation were included (Pickles et al. 2021).

Phylogenetics

The HPTN 071-02 (PopART) Phylogenetics Study and the sequencing procedure used have been described previously (Hayes et al. 2014, 2019; Hall et al. 2024). In

summary: samples for HIV genome sequencing were obtained from HIV-positive individuals in the PC (during the earliest visit in which they tested positive for the virus, which could be baseline), and individuals who had been approached at the health-care facilities between 2016 and 2018. These individuals were not on ART at the time they were recruited and were in most attending the clinics for ART initiation, although some self-reported previous ART and poor adherence. Whole-genome deep sequencing was done on the Illumina platform using veSEQ-HIV (Bonsall et al. 2020).

The process by which probable transmission pairs were identified has also been previously described (Hall et al. 2024). Briefly, after genome assembly using SPAdes (Nurk et al. 2013) and alignment using *shiver* (Wymant et al. 2018), the collection of viral consensus sequences was clustered using HIV-TRACE (Kosakovsky Pond et al. 2018) with an intentionally generous distance threshold of 0.04 substitutions per site. The deep-sequence data files from all members of every identified cluster were then used as input for separate analyses with *phyloscanner* v1.8.0 (Wymant et al. 2017). The set of genomic windows used in *phyloscanner* were uniformly of 250bp in length, spanning the entire HIV genome, such that the start of each window was 10 bp further along the genome (relative to HXB2) than the previous one. The distance threshold used to identify likely transmission pairs in a single window was of 0.025 substitutions per site on a normalised scale designed to represent average distances on the *gag* and *pol* genes (Ratmann et al. 2019). After adjustment for missing windows, overall assignment of pairs as likely to involve direct transmission was declared where more than 50% of windows suggested it. Same-sex pairs (which indicate at least one missing intermediary under the assumption that transmission is

heterosexual) were then discarded to arrive at a final set; previous work has estimated that the proportion of true male-to-male transmissions was very low amongst trial participants (Hall et al. 2024).

The date of HIV infection of each individual was estimated using HIV-phyloTSI (Golubchik et al. 2022) and these dates, together with recorded dates of birth, were used to estimate the age of source and recipient partners in each pair at the time of the infection of the recipient. While our previous analysis (Hall et al. 2024) investigated only transmissions estimated to have occurred during the time-frame of the PopART trial itself, this filter was not applied here.

Direction of transmission was determined using both the *phyloscanner* score based on the tree topology and comparison of infection dates as estimated by HIV-phyloTSI; see (Hall et al. 2024) for full details.

Appendix 1: Comparison of identified pairs from phylogenetics to random pairs

In this appendix, we explore the presence of a genuine signal of HIV transmission in the phylogenetics dataset, by comparing the results obtained from *phyloscanner* to the results of attaching two opposite-sex individuals in the genomic dataset to each other at random as false transmission pairs.

Specifically, we picked a random selection of 156 male and 197 female participants for whom sequence data were available. (156 and 197 are the number of recipients of each sex in the actual dataset.) We additionally selected these such that their estimated infection dates (as determined by phyloTSI) were no earlier than the first such estimate amongst the recipients in the actual data. For each of these 353 individuals we randomly picked an individual of the opposite sex, whose estimated infection date was earlier than their own, as a shuffled “source” for their transmission. The time of transmission was taken to be the estimated time of infection of the “recipient”. We repeated this process 1,000 times and compare the distribution of age- and sex-related variables across these replicates to those in the real dataset (Appendix 1-Figures 1 to 3).

Recipients. See Appendix 1-Figure 1 for a comparison of the mean and variance of recipients age at the time of transmission, by sex, across the replicates of the shuffled dataset (histograms) to its true value (green line). There is no evidence of a difference with respect to either mean or variance for either recipient sex (all empirical *p*-values >0.178). This provides reassurance that the age distribution of individuals with a sampled source does not differ from that of the complete set of participants providing sequences.

Estimated time since infection showed considerable variation between real and shuffled datasets (Appendix 1-Figure 2), with individuals from shuffled pairs (i.e. those whose sources had not genuinely been sampled) showing greater mean values of the variable (empirical *p*-values <0.001 for both sexes). This is consistent with younger infections being more likely to have a sampled source; elapsed time

from infection would be associated with increasing probability that the source would not be eligible for sampling, either due to migration or commencing ART (as successful ART treatment prevents the acquisition of a sequence).

Sources. In appendix 1-figure 3, mean estimated source ages at the time of transmission are compared. For male sources there is no evidence of a difference in mean age (empirical p=0.41 but female sources are younger in the actual dataset (p<0.001). The variances of source ages also deviate from the null hypothesis for both sexes (p=0.02 for males, p<0.001 for females). Effectively, this test compares the demographics of those individuals identified as sources by the analysis to the general HIV positive population, as the shuffled “sources” need not have genuinely transmitted HIV. It is unsurprising that these variables differed. In this case, a younger age for female sources of transmission when compared to a random sample of HIV-positive women is consistent with onward transmission peaking in younger age groups. A smaller variance in the ages of genuine sources is also consistent with their being a peak age of risk for transmitting the virus (see figure 1).

As with recipients, “sources” in the shuffled dataset tend to have older infections (Appendix 1-Figure 4). In this case this is likely due to a combination of the same rationale outlined for recipients above, and the fact that genuinely identified sources are younger than a random selection of individuals (empirical p<0.001 for both sexes). They would be expected, on average, to have more recent infections because the range of possible times since infection is constrained by a more recent sexual debut.

Age gap. Appendix 1-Figure 5 displays the histograms for the age gap between male and female partners. There is strong evidence for a difference in means for female-to-male transmission (empirical $p<0.001$) but not for male-to-female ($p=0.307$). Meanwhile there are plainly large differences in the variances ($p<0.001$), meaning that genuine transmission pairs have much more closely correlated ages than random pairs.

Discussion. The sampling frame for this study is unusual. We have taken a sample of HIV-positive individuals in the trial communities, and within that sample identified individuals who we are confident acted as sources of transmission for others in the sample. Only 4.96% of sampled individuals have a source identified in this way. As every HIV-positive individual is by definition a recipient of transmission, this 4.96% can be seen as a subsample of the sampled recipient population. If the subsample were unrepresentative, meaning that those recipients for whom we found a source were different to those for whom we did not, it would in turn suggest bias in the characteristics of the sources that we identified for these recipients. That there was no statistical difference between the true subsample of recipients with sampled sources and random collections of recipients of the same size should give us confidence that this bias was not present.

The *phyloscanner* procedure would obviously have failed if it was doing no more than sampling random opposite-sex pairs of individuals in the dataset and designating one of each as the source of the infection in the other. Were this true, then the demographics of the population identified as sources would be no different from those of the entire sample, and the distribution of age gaps in identified pairs

would be no different from that of random pairs of opposite-sex individuals in the dataset. We found strong statistical evidence that neither of these are the case.

Appendix 2: Sensitivity analysis for *phyloscanner* settings

In this section we re-investigate the age and sex dynamics of the set of identified transmission pairs from sets of pairs obtained when some of the key *phyloscanner* settings are varied. For further information regarding these settings, see the *phyloscanner* paper (Wymant et al. 2017) and manual (available at https://github.com/BDI-pathogens/phyloscanner)

Read downsampling

The nature of deep sequencing is such that genome coverage will vary both between samples and in different regions of the genome for the same sample. It may well be the case that, for a given *phyloscanner* window and a putative transmission pair, the number of reads covering the window can differ between the two members of the pair. The exact effects of this on inference have yet to be exhaustively explored, but one *phyloscanner* option allows the user to specify a maximum number of reads to be analysed per window, with the exact choice of reads being randomly selected. (Note that, as many reads in the same window will be identical due to the amplification process during sequencing, regardless of underlying variation within the sample, this process is not the same as limiting the number of distinct genomic *variants*.) A further choice must be made regarding how to proceed if the total

number of reads for one host in one window fails to reach that maximum number. The two options are, firstly, to include all the reads that do exist ("keep") or, secondly, to remove that host from consideration entirely in that window ("blacklist"). (As no inference is conducted from windows in which reads are not present from both hosts, the practical effect of the "blacklist" option is that those windows will be excluded from the analysis.) In the main analysis of this paper we did not use read downsampling; here we describe results from applying maximum read counts of 25 and 50, with both the "keep" and "blacklist" strategies.

As described in the main text, there are 355 pairs found when no read downsampling was applied. Applying the tool always reduces this number, to:

- 296 with a maximum of 50 reads per window and the "keep" strategy
- 274 with a maximum of 50 reads per window and the "blacklist" strategy
- 285 with a maximum of 25 reads per window and the "keep" strategy
- 284 with a maximum of 25 reads per window and the "blacklist" strategy

There was no evidence of a difference in the proportion of pairs with a male source (chi-squared test p=0.699) between the five groups. In this work, we have three key variables of interest: source age at estimated time of transmission, recipient age at estimated time of transmission, and age gap. Each of these was examined separately according to the sex of the inferred source individual, making a total of six variables in practice. We saw no evidence of a difference in any of these variables regardless of the downsampling procedure. (Kruskall-Wallis test p>0.7 in all cases). Their distribution can be seen in appendix 2-figure 1.

Window width

A key *phyloscanner* setting is the width of the sliding window used to subdivide the viral genome. Narrower windows allow the genome to be tiled with more windows and increase the number of reads per window because *phyloscanner* requires reads to fully span a window to be included in it. However, narrow windows have fewer sites at which mutations can be observed, reducing phylogenetic resolution. In the main text the window width was set to 250 base pairs, with each window starting 10bp further along than the previous. We also reduced this to 175bp, with the same 10bp incremental shift, and obtained the set of probable transmission pairs from this variation. The 355 pairs from the main analysis were reduced to 291 here. There was no evidence that these sets differed in respect to the proportion of pairs which were male-to-female (chi-squared test p=0.903) or of our six key quantities (Mann-Whitney U-test p>0.63 in all cases). See appendix 2-figure 2 for the distribution of these values.

Window thresholds for pair identification

As phylogenetic relationships between the reads from each of two individuals will vary across the genome, we apply thresholds on the proportion of windows in which a) the phylogeny suggests that the pair are a transmission pair ("linkage"), and b) the phylogeny suggests a direction of transmission. (These proportions are adjusted for non-informative windows as outlined by (Ratmann et al. 2019)). In the main analysis, individuals are linked as a pair at a proportion threshold of 0.5. For pairs that are

above this threshold, a direction of transmission is called at a proportion threshold of 0.33 (of the total number of windows). We also tested more stringent thresholds, firstly 0.66 for linkage and 0.5 for direction, and then 0.75 for linkage and 0.66 for direction. The 355 pairs from the main analysis were reduced to 260 in the former case and 201 in the latter. There is no evidence of a difference in the proportion of pairs identified as having a male source (chi-squared test p=0.927) or in the distribution of our six quantities of interest (Kruskal-Wallis test, all p>0.3). See appendix 2-figure 3 for the distribution of these estimates.

Patristic distance threshold for per-window pair identification

The final setting that we varied was the patristic distance threshold at which, in a single genomic window, a pair of individuals are inferred as likely to be involved in direct transmission. As phylogenetic branch lengths vary across the genome, this distance is normalised to standardise to expected lengths on the *gag* and *pol* genes. In the main analysis the value used is 0.02 normalised substitutions per site. Here we use values from 0.01 to 0.04 in increments of 0.005.

Increasing this threshold should increase sensitivity, but decrease specificity, of the identification of transmission pairs. For heterosexual HIV transmission, instances where there is only a single unsampled intermediate host can be omitted by simply excluding same-sex pairs (e.g. male-male pairs must be missing at least one unsampled intermediate female). Then false positives must be missing at least two unsampled intermediaries (for example identifying male A → female B → male C →

female D as male A → female D). Reducing the patristic distance threshold will reduce the probability that such false positives are included in the dataset. See appendix 2-table 1 for the number of pairs identified using each threshold.

There is no evidence of an association of the threshold with the proportion of transmission pairs identified with male sources, whether the threshold is treated as a discrete variable (chi-squared test p=0.962), or as continuous (logistic regression p=0.376). Similarly, none of our six key estimates show evidence for an association with the threshold. Again, this is true whether the threshold is treated as a discrete variable (Kruskal-Wallis p>0.8 in all cases) or fit as a continuous predictor of the variable in a linear regression model, with source sex as an interaction term (no coefficients other than source sex having p<0.3). See appendix 2-figure 4 for the actual distribution of their values.

That the key outputs of our analysis are robust to decreasing the distance threshold from 0.02 to 0.01 should offer reassurance towards concerns regarding any biasing effect on the results of this paper caused by the inclusion of false positive transmission pairs.

References

Akullian, Adam, Anna Bershteyn, Daniel Klein, Alain Vandormael, Till Bärnighausen, and Frank Tanser. 2017. “Sexual Partnership Age Pairings and Risk of HIV Acquisition in Rural South Africa.” *AIDS* 31 (12): 1755–64.

Akullian, Adam, Michelle Morrison, Geoffrey P. Garnett, Zandile Mnisi, Nomthandazo Lukhele, Daniel Bridenbecker, and Anna Bershteyn. 2020. “The Effect of 90-90-90 on HIV-1 Incidence and Mortality in eSwatini: A Mathematical Modelling Study.” *The Lancet. HIV* 7 (5): e348–58.

Arazoza, Héctor de, Jose Joanes, Rachid Lounes, Camille Legeai, Stéphan Clémençon, Jorge Pérez, and Bertran Auvert. 2007. “The HIV/AIDS Epidemic in Cuba: Description and Tentative Explanation of Its Low HIV Prevalence.” *BMC Infectious Diseases* 7 (November):130.

Bacaër, Nicolas, Carel Pretorius, and Bertran Auvert. 2010. “An Age-Structured Model for the Potential Impact of Generalized Access to Antiretrovirals on the South African HIV Epidemic.” *Bulletin of Mathematical Biology* 72 (8): 2180–98.

Balkus, Jennifer E., Gonasagrie Nair, Elizabeth T. Montgomery, Anu Mishra, Thesla Palanee-Phillips, Gita Ramjee, Ravindre Panchia, et al. 2015. “Age-Disparate Partnerships and Risk of HIV-1 Acquisition Among South African Women Participating in the VOICE Trial.” *Journal of Acquired Immune Deficiency Syndromes* 70 (2): 212–17.

Beauclair, Roxanne, Stéphane Helleringer, Niel Hens, and Wim Delva. 2016. “Age Differences between Sexual Partners, Behavioural and Demographic Correlates, and HIV Infection on Likoma Island, Malawi.” *Scientific Reports* 6 (November):36121.

Bershteyn, Anna, Jaline Gerardin, Daniel Bridenbecker, Christopher W. Lorton, Jonathan Bloedow, Robert S. Baker, Guillaume Chabot-Couture, et al. 2018. “Implementation and Applications of EMOD, an Individual-Based Multi-Disease Modeling Platform.” *Pathogens and Disease* 76 (5). https://doi.org/10.1093/femspd/fty059.

Bonsall, David, Tanya Golubchik, Mariateresa de Cesare, Mohammed Limbada, Barry Kosloff, George MacIntyre-Cockett, Matthew Hall, et al. 2020. “A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings.” *Journal of Clinical Microbiology* 58 (10). https://doi.org/10.1128/JCM.00382-20.

Bu, Fan, Joseph Kagaayi, Kate Grabowski, Oliver Ratmann, and Jason Xu. 2023. “Inferring HIV Transmission Patterns from Viral Deep-Sequence Data via Latent Typed Point Processes.” *arXiv [stat.ME]*. arXiv. http://arxiv.org/abs/2302.11567.

Cohen, Myron S., Ying Q. Chen, Marybeth McCauley, Theresa Gamble, Mina C. Hosseinipour, Nagalingeswaran Kumarasamy, James G. Hakim, et al. 2016. “Antiretroviral Therapy for the Prevention of HIV-1 Transmission.” *The New England Journal of Medicine* 375 (9): 830–39.

Cori, Anne, Michael Pickles, Ard van Sighem, Luuk Gras, Daniela Bezemer, Peter Reiss, and Christophe Fraser. 2015. “CD4+ Cell Dynamics in Untreated HIV-1 Infection: Overall Rates, and Effects of Age, Viral Load, Sex and Calendar Time.” *AIDS* 29 (18): 2435–46.

Fogel, Jessica M., Ethan A. Wilson, Estelle Piwowar-Manning, Autumn Breaud, William Clarke, Christos Petropoulos, Ayana Moore, et al. 2022. “HIV Drug Resistance in a Community-Randomized Trial of Universal Testing and Treatment: HPTN 071 (PopART).” *Journal of the International AIDS Society* 25 (7): e25941.

George, Gavin, Sean Beckett, Tarylee Reddy, Kaymarlin Govender, Cherie Cawood, David Khanyile, and Ayesha B. M. Kharsany. 2022. “Determining HIV Risk for Adolescent Girls and Young Women (AGYW) in Relationships with ‘Blessers’ and Age-Disparate Partners: A Cross-Sectional Survey in Four Districts in South Africa.” *BMC Public Health* 22 (1): 973.

Golubchik, Tanya, Lucie Abeler-Dörner, Matthew Hall, Chris Wymant, David Bonsall, George Macintyre-Cockett, Laura Thomson, et al. 2022. “HIV-phyloTSI: Subtype-Independent Estimation of Time since HIV-1 Infection for Cross-Sectional Measures of

Population Incidence Using Deep Sequence Data." *medRxiv*, May, 2022.05.15.22275117.

Grabowski, Mary Kathryn, and Justin Lessler. 2017. "Phylogenetic Insights into Age-Disparate Partnerships and HIV." *The Lancet. HIV* 4 (1): e8–9.

Gregson, Simon, Constance A. Nyamukapa, Geoffrey P. Garnett, Peter R. Mason, Tom Zhuwau, Michel Caraël, Stephen K. Chandiwana, and Roy M. Anderson. 2002. "Sexual Mixing Patterns and Sex-Differentials in Teenage Exposure to HIV Infection in Rural Zimbabwe." *The Lancet* 359 (9321): 1896–1903.

Hall, Matthew, Tanya Golubchik, David Bonsall, Lucie Abeler-Dörner, Mohammed Limbada, Barry Kosloff, Ab Schaap, et al. 2024. "Demographics of Sources of HIV-1 Transmission in Zambia: A Molecular Epidemiology Analysis in the HPTN 071 PopART Study." *The Lancet. Microbe* 5 (1): e62–71.

Harling, Guy, Marie-Louise Newell, Frank Tanser, Ichiro Kawachi, S. V. Subramanian, and Till Bärnighausen. 2014. "Do Age-Disparate Relationships Drive HIV Incidence in Young Women? Evidence from a Population Cohort in Rural KwaZulu-Natal, South Africa." *Journal of Acquired Immune Deficiency Syndromes* 66 (4): 443–51.

Hayes, Richard, Helen Ayles, Nulda Beyers, Kalpana Sabapathy, Sian Floyd, Kwame Shanaube, Peter Bock, et al. 2014. "HPTN 071 (PopART): Rationale and Design of a Cluster-Randomised Trial of the Population Impact of an HIV Combination Prevention Intervention Including Universal Testing and Treatment - a Study Protocol for a Cluster Randomised Trial." *Trials* 15 (February):57.

Hayes, Richard, Deborah Donnell, Sian Floyd, Nomtha Mandla, Justin Bwalya, Kalpana Sabapathy, Blia Yang, et al. 2019. "Effect of Universal Testing and Treatment on HIV Incidence - HPTN 071 (PopART)." *The New England Journal of Medicine* 381 (3): 207–18.

Kelly, Robert J., Ronald H. Gray, Nelson K. Sewankambo, David Serwadda, Fred Wabwire-Mangen, Tom Lutalo, and Maria J. Wawer. 2003. "Age Differences in Sexual Partners and Risk of HIV-1 Infection in Rural Uganda." *Journal of Acquired Immune Deficiency*

Syndromes 32 (4): 446–51.

Kosakovsky Pond, Sergei L., Steven Weaver, Andrew J. Leigh Brown, and Joel O. Wertheim. 2018. “HIV-TRACE (TRAnsmission Cluster Engine): A Tool for Large Scale Molecular Epidemiology of HIV-1 and Other Rapidly Evolving Pathogens.” *Molecular Biology and Evolution* 35 (7): 1812–19.

Kusejko, Katharina, Claus Kadelka, Alex Marzel, Manuel Battegay, Enos Bernasconi, Alexandra Calmy, Matthias Cavassini, et al. 2018. “Inferring the Age Difference in HIV Transmission Pairs by Applying Phylogenetic Methods on the HIV Transmission Network of the Swiss HIV Cohort Study.” *Virus Evolution* 4 (2): vey024.

Leitner, Thomas, and Ethan Romero-Severson. 2018. “Phylogenetic Patterns Recover Known HIV Epidemiological Relationships and Reveal Common Transmission of Multiple Variants.” *Nature Microbiology*, July. https://doi.org/10.1038/s41564-018-0204-9.

Liesenborgs, Jori, Diana M. Hendrickx, Elise Kuylen, David Niyukuri, Niel Hens, and Wim Delva. 2019. “SimpactCyan 1.0: An Open-Source Simulator for Individual-Based Models in HIV Epidemiology with R and Python Interfaces.” *Scientific Reports* 9 (1): 19289.

Magosi, Lerato E., Yinfeng Zhang, Tanya Golubchik, Victor DeGruttola, Eric Tchetgen Tchetgen, Vladimir Novitsky, Janet Moore, et al. 2022. “Deep-Sequence Phylogenetics to Quantify Patterns of HIV Transmission in the Context of a Universal Testing and Treatment Trial - BCPP/Ya Tsie Trial.” *eLife* 11 (March). https://doi.org/10.7554/eLife.72657.

Monod, Mélodie, Andrea Brizzi, Ronald M. Galiwango, Robert Ssekubugu, Yu Chen, Xiaoyue Xi, Edward Nelson Kankaka, et al. 2024. “Longitudinal Population-Level HIV Epidemiologic and Genomic Surveillance Highlights Growing Gender Disparity of HIV Transmission in Uganda.” *Nature Microbiology* 9 (1): 35–54.

Mujugira, Andrew, Jared M. Baeten, Deborah Donnell, Patrick Ndase, Nelly R. Mugo, Linda Barnes, James D. Campbell, et al. 2011. “Characteristics of HIV-1 Serodiscordant Couples Enrolled in a Clinical Trial of Antiretroviral Pre-Exposure Prophylaxis for HIV-1

Prevention." *PloS One* 6 (10): e25828.

Napper, Lucy E., Dennis G. Fisher, Grace L. Reynolds, and Mark E. Johnson. 2010. "HIV Risk Behavior Self-Report Reliability at Different Recall Periods." *AIDS and Behavior* 14 (1): 152–61.

Nichols, Brooke E., Elvin H. Geng, Eric Moakley, Andrew N. Phillips, Jeffrey W. Imai-Eaton, John Stover, Edinah Mudimu, and Anna Grimsrud. 2025. "Rapid Development of an Online Tracker to Communicate the Human Impact of Abruptly Halting PEPFAR Support." *Journal of the International AIDS Society* 28 (3): e26433.

Nurk, Sergey, Anton Bankevich, Dmitry Antipov, Alexey A. Gurevich, Anton Korobeynikov, Alla Lapidus, Andrey D. Prjibelski, et al. 2013. "Assembling Single-Cell Genomes and Mini-Metagenomes from Chimeric MDA Products." *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 20 (10): 714–37.

Oliveira, Tulio de, Ayesha B. M. Kharsany, Tiago Gräf, Cherie Cawood, David Khanyile, Anneke Grobler, Adrian Puren, et al. 2017. "Transmission Networks and Risk of HIV Infection in KwaZulu-Natal, South Africa: A Community-Wide Phylogenetic Study." *The Lancet HIV* 4 (1): e41–50.

Ott, Miles Q., Till Bärnighausen, Frank Tanser, Mark N. Lurie, and Marie-Louise Newell. 2011. "Age-Gaps in Sexual Partnerships: Seeing beyond 'Sugar Daddies.'" *AIDS* 25 (6): 861–63.

Pettifor, Audrey E., Helen V. Rees, Immo Kleinschmidt, Annie E. Steffenson, Catherine MacPhail, Lindiwe Hlongwa-Madikizela, Kerry Vermaak, and Nancy S. Padian. 2005. "Young People's Sexual Health in South Africa: HIV Prevalence and Sexual Behaviors from a Nationally Representative Household Survey." *AIDS* 19 (14): 1525–34.

Pickles, Michael, Anne Cori, William J. M. Probert, Rafael Sauter, Robert Hinch, Sarah Fidler, Helen Ayles, et al. 2021. "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 Computational Biology* 17 (9): e1009301.

Probert, William J. M., Rafael Sauter, Michael Pickles, Anne Cori, Nomtha F. Bell-Mandla,

Justin Bwalya, Lucie Abeler-Dörner, et al. 2022. “Projected Outcomes of Universal Testing and Treatment in a Generalised HIV Epidemic in Zambia and South Africa (the HPTN 071 [PopART] Trial): A Modelling Study.” *The Lancet. HIV* 9 (11): e771–80.

Ratmann, Oliver, M. Kate Grabowski, Matthew Hall, Tanya Golubchik, Chris Wymant, Lucie Abeler-Dörner, David Bonsall, et al. 2019. “Inferring HIV-1 Transmission Networks and Sources of Epidemic Spread in Africa with Deep-Sequence Phylogenetic Analysis.” *Nature Communications* 10 (1): 1411.

Ratmann, Oliver, Joseph Kagaayi, Matthew Hall, Tanya Golubchick, Godfrey Kigozi, Xiaoyue Xi, Chris Wymant, et al. 2020. “Quantifying HIV Transmission Flow between High-Prevalence Hotspots and Surrounding Communities: A Population-Based Study in Rakai, Uganda.” *The Lancet. HIV* 7 (3): e173–83.

Romero-Severson, Ethan O., Ingo Bulla, Nick Hengartner, Inês Bártolo, Ana Abecasis, José M. Azevedo-Pereira, Nuno Taveira, and Thomas Leitner. 2017. “Donor-Recipient Identification in Para- and Poly-Phyletic Trees Under Alternative HIV-1 Transmission Hypotheses Using Approximate Bayesian Computation.” *Genetics*, September. https://doi.org/10.1534/genetics.117.300284.

Romero-Severson, Ethan O., Ingo Bulla, and Thomas Leitner. 2016. “Phylogenetically Resolving Epidemiologic Linkage.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (10): 2690–95.

Smith, Jennifer, Constance Nyamukapa, Simon Gregson, James Lewis, Sitholubuhle Magutshwa, Christina Schumacher, Phyllis Mushati, Tim Hallett, and Geoff Garnett. 2014. “The Distribution of Sex Acts and Condom Use within Partnerships in a Rural Sub-Saharan African Population.” *PloS One* 9 (2): e88378.

Street, Renée A., Tarylee Reddy, and Gita Ramjee. 2016. “The Generational Effect on Age Disparate Partnerships and the Risk for Human Immunodeficiency Virus and Sexually Transmitted Infections Acquisition.” *International Journal of STD & AIDS* 27 (9): 746–52.

Topazian, Hillary M., Marie C. D. Stoner, Jessie K. Edwards, Kathleen Kahn, Francesc Xavier Gómez-Olivé, Rhian Twine, James P. Hughes, Myron S. Cohen, and Audrey

Pettifor. 2020. “Variations in HIV Risk by Young Women’s Age and Partner Age Disparity in Rural South Africa (HPTN 068).” *Journal of Acquired Immune Deficiency Syndromes* 83 (4): 350–56.

Turner, Abigail Norris, Alana E. De Kock, Amy Meehan-Ritter, Kelly Blanchard, Mohlatlego H. Sebola, Anwar A. Hoosen, Nicol Coetzee, and Charlotte Ellertson. 2009. “Many Vaginal Microbicide Trial Participants Acknowledged They Had Misreported Sensitive Sexual Behavior in Face-to-Face Interviews.” *Journal of Clinical Epidemiology* 62 (7): 759–65.

UNAIDS. 2016. “The Gap Report.” The Joint United Nations Programme on HIV/AIDS. https://www.unaids.org/sites/default/files/media_asset/UNAIDS_Gap_report_en.pdf.

———. 2018. “The Youth Bulge and HIV.” The Joint United Nations Programme on HIV/AIDS. https://www.unaids.org/en/resources/documents/2018/the-youth-bulge-and-hiv.

“UNAIDS HIV Financial Dashboard.” n.d. Accessed March 11, 2025. https://hivfinancial.unaids.org/hivfinancialdashboards.html#.

United Nations. 2015. “World Population Prospects 2015.” United Nations. https://population.un.org/wpp/Publications/.

———. n.d. “Population Datasets.” https://www.un.org/en/development/desa/population/publications/dataset/index.asp.

Wolock, Timothy M., Seth Flaxman, Kathryn A. Risher, Tawanda Dadirai, Simon Gregson, and Jeffrey W. Eaton. 2021. “Evaluating Distributional Regression Strategies for Modelling Self-Reported Sexual Age-Mixing.” *eLife* 10 (June). https://doi.org/10.7554/eLife.68318.

Wymant, Chris, François Blanquart, Tanya Golubchik, Astrid Gall, Margreet Bakker, Daniela Bezemer, Nicholas J. Croucher, et al. 2018. “Easy and Accurate Reconstruction of Whole HIV Genomes from Short-Read Sequence Data with Shiver.” *Virus Evolution* 4 (1): vey007.

Wymant, Chris, Matthew Hall, Oliver Ratmann, David Bonsall, Tanya Golubchik, Mariateresa

de Cesare, Astrid Gall, Marion Cornelissen, Christophe Fraser, and STOP-HCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration. 2017. “PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity.” *Molecular Biology and Evolution*, November. https://doi.org/10.1093/molbev/msx304.

Xi, Xiaoyue, Simon E. F. Spencer, Matthew Hall, M. Kate Grabowski, Joseph Kagaayi, Oliver Ratmann, and Rakai Health Sciences Program and PANGEA-HIV. 2022. “Inferring the Sources of HIV Infection in Africa from Deep- sequence Data with Semi-parametric Bayesian Poisson Flow Models.” *Journal of the Royal Statistical Society. Series C, Applied Statistics* 71 (3): 517–40.

Zhang, Yinfeng, Chris Wymant, Oliver Laeyendecker, M. Kathryn Grabowski, Matthew Hall, Sarah Hudelson, Estelle Piwowar-Manning, et al. 2021. “Evaluation of Phylogenetic Methods for Inferring the Direction of Human Immunodeficiency Virus (HIV) Transmission: HIV Prevention Trials Network (HPTN) 052.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 72 (1): 30–37.

		All participants			Females			Males		
		Both roles	Sources	Recipients	Both roles	Sources	Recipients	Both roles	Sources	Recipients
All participants		710	355	355	355	157	198	355	198	157
Community number	01	58 (0.08)	29 (0.08)	29 (0.08)	27 (0.08)	14 (0.09)	13 (0.07)	31 (0.09)	15 (0.08)	16 (0.1)
	02	62 (0.09)	31 (0.09)	31 (0.09)	32 (0.09)	19 (0.12)	13 (0.07)	30 (0.08)	12 (0.06)	18 (0.11)
	03	107 (0.15)	51 (0.14)	56 (0.16)	50 (0.14)	18 (0.11)	32 (0.16)	57 (0.16)	33 (0.17)	24 (0.15)
	04	61 (0.09)	31 (0.09)	30 (0.08)	28 (0.08)	10 (0.06)	18 (0.09)	33 (0.09)	21 (0.11)	12 (0.08)
	05	108 (0.15)	53 (0.15)	55 (0.15)	56 (0.16)	24 (0.15)	32 (0.16)	52 (0.15)	29 (0.15)	23 (0.15)
	06	113 (0.16)	57 (0.16)	56 (0.16)	61 (0.17)	26 (0.17)	35 (0.18)	52 (0.15)	31 (0.16)	21 (0.13)
	10	51 (0.07)	24 (0.07)	27 (0.08)	26 (0.07)	11 (0.07)	15 (0.08)	25 (0.07)	13 (0.07)	12 (0.08)
	11	46 (0.06)	24 (0.07)	22 (0.06)	23 (0.06)	10 (0.06)	13 (0.07)	23 (0.06)	14 (0.07)	9 (0.06)
	12	104 (0.15)	55 (0.15)	49 (0.14)	52 (0.15)	25 (0.16)	27 (0.14)	52 (0.15)	30 (0.15)	22 (0.14)
Trial arm	A	220 (0.31)	106 (0.3)	114 (0.32)	108 (0.3)	48 (0.31)	60 (0.3)	112 (0.32)	58 (0.29)	54 (0.34)
	B	217 (0.31)	110 (0.31)	107 (0.3)	111 (0.31)	50 (0.32)	61 (0.31)	106 (0.3)	60 (0.3)	46 (0.29)
	C	273 (0.38)	139 (0.39)	134 (0.38)	136 (0.38)	59 (0.38)	77 (0.39)	137 (0.39)	80 (0.4)	57 (0.36)
Enrolment method	Health care facilities	644 (0.91)	318 (0.9)	326 (0.92)	308 (0.87)	132 (0.84)	176 (0.89)	336 (0.95)	186 (0.94)	150 (0.96)
	Population cohort	66 (0.09)	37 (0.1)	29 (0.08)	47 (0.13)	25 (0.16)	22 (0.11)	19 (0.05)	12 (0.06)	7 (0.04)
Sampling year	2014	26 (0.04)	25 (0.07)	1 (0)	22 (0.06)	21 (0.13)	1 (0.01)	4 (0.01)	4 (0.02)	0 (0)
	2015	10 (0.01)	3 (0.01)	7 (0.02)	7 (0.02)	0 (0)	7 (0.04)	3 (0.01)	3 (0.02)	0 (0)
	2016	223 (0.31)	106 (0.3)	117 (0.33)	111 (0.31)	46 (0.29)	65 (0.33)	112 (0.32)	60 (0.3)	52 (0.33)
	2017	323 (0.45)	160 (0.45)	163 (0.46)	159 (0.45)	73 (0.46)	86 (0.43)	164 (0.46)	87 (0.44)	77 (0.49)
	2018	128 (0.18)	61 (0.17)	67 (0.19)	56 (0.16)	17 (0.11)	39 (0.2)	72 (0.2)	44 (0.22)	28 (0.18)
Marital status at sampling	Never married	89 (0.13)	44 (0.12)	45 (0.13)	44 (0.12)	23 (0.15)	21 (0.11)	45 (0.13)	21 (0.11)	24 (0.15)
	Married/living as married	496 (0.7)	256 (0.72)	240 (0.68)	239 (0.67)	104 (0.66)	135 (0.68)	257 (0.72)	152 (0.77)	105 (0.67)
	Divorced/separated	101 (0.14)	47 (0.13)	54 (0.15)	54 (0.15)	25 (0.16)	29 (0.15)	47 (0.13)	22 (0.11)	25 (0.16)
	Widowed	20 (0.03)	6 (0.02)	14 (0.04)	15 (0.04)	3 (0.02)	12 (0.06)	5 (0.01)	3 (0.02)	2 (0.01)
	Unknown	4 (0.01)	2 (0.01)	2 (0.01)	3 (0.01)	2 (0.01)	1 (0.01)	1 (0)	0 (0)	1 (0.01)
Age at estimated time of transmission (years, mean (SD))		30.1 (8.6)	30.8 (8.4)	29.5 (8.9)	27.5 (8.4)	27.1 (7.3)	27.8 (9.2)	32.7 (8.1)	33.7 (8)	31.5 (8)
Relative age of partner (years, mean (SD))		0 (8.9)	1.3 (8.8)	-1.3 (8.8)	-5.2 (7.2)	-4.4 (6.6)	-5.9 (7.5)	5.2 (7.2)	5.9 (7.5)	4.4 (6.6)

Table 1: Demographics of the individuals in the 355 probable transmission pairs. Age is calculated at the estimated time at which the source infected the recipient. Numbers are given as count and proportion (column-wise) for discrete variables and mean and standard deviation for continuous variables.

Variable	Sexes	IBM		Phylogenetics	
		Median	IQR	Median	IQR
Source age	Males	30.6	(25.9-38.2)	32.5	(28.9-37.3)
	Females	26.6	(22.6-32.5)	24.9	(21.9-30.1)
Recipient age	Males	29.2	(24-37.1)	29.7	(25.3-35.7)
	Females	25	(20.8-31.9)	25.8	(21.9-31.5)
Age gap (source-recipient)	Male source, female recipient	5.1	(1.48-8.77)	6.08	(2.92-10.2)
	Female source, male recipient	-2.21	(-5.21-0.83)	-4.54	(-8.27--0.574)

Table 2: Summary of ages, and age gaps, in transmission pairs identified using both methods. The IBM results are from the best-fitting ABC simulation.

Distance threshold (substitutions per site, normalised to represent average distances on the *gag* and *pol* genes)	**Number of pairs identified**
0.01	257
0.015	315
0.02	355
0.025	386
0.03	417

0.035	441
0.04	459

Appendix 2-Table 1 Number of phylogenetic transmission pairs identified using different patristic distance thresholds.

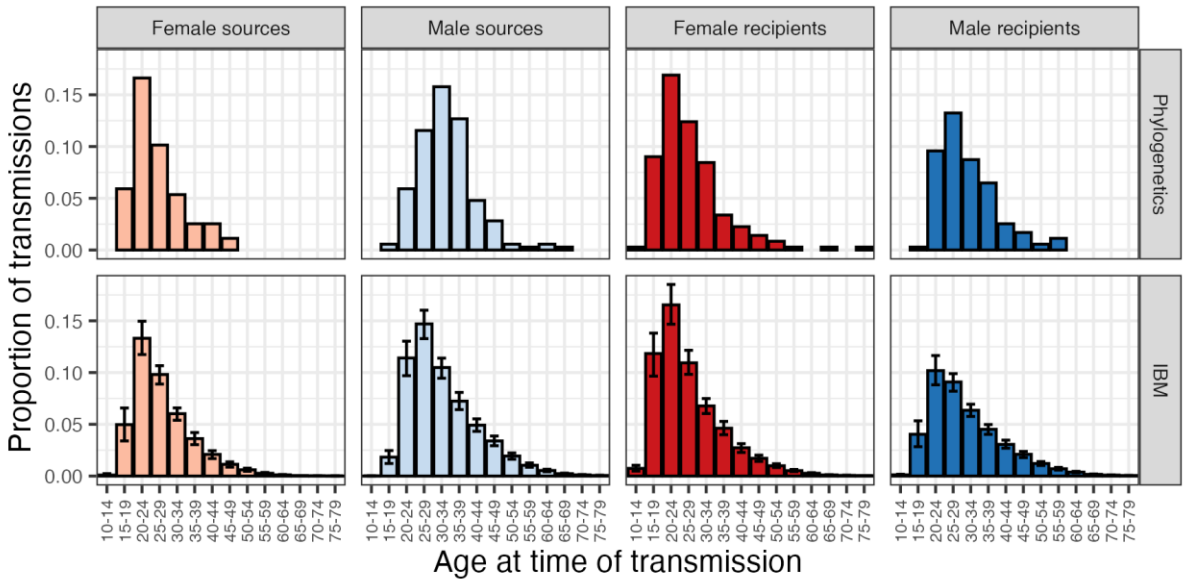


Figure 1: Phylogenetic analysis and the IBM model show similar age distributions for sources and recipients. Histograms of the proportions of transmissions from and to different age groups in the IBM results (top) and the phylogenetics (bottom). Values from the IBM are the posterior mean proportion of transmissions with the 95% highest density interval displayed. As every transmission has a source and a recipient, proportions sum to 1 across the two source histograms

(left, light colours) and also to 1 across the two recipient histograms (right, dark colours).

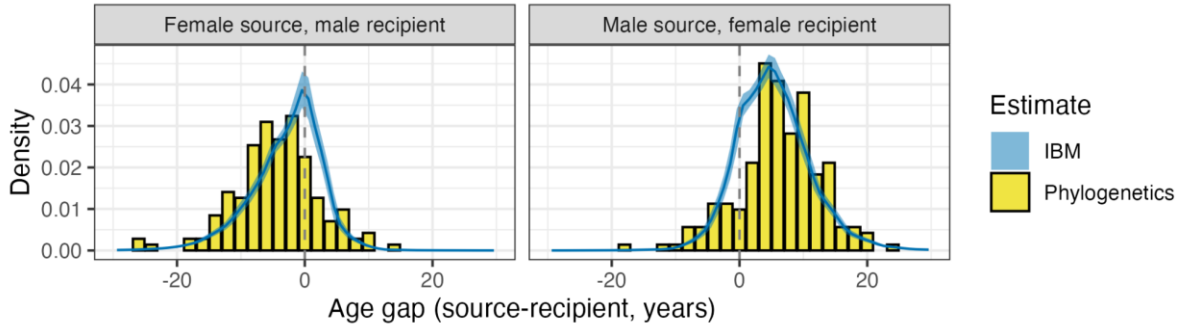


Figure 2: Male partners are on average older in both male-to-female and female-to-male pairs. Distributions of age gaps within transmission pairs from the phylogenetics (yellow, histogram) and the IBM (blue, kernel density estimate). For the IBM, the 95% highest posterior density intervals are given by the shaded area. Female-to-male transmission pairs are displayed on the left and male-to-female on the right. The dashed line indicates equal age.

Figure 3: Transmissions are most common in young couples. Age-based who-acquires-infection-from-whom (WAIFW) matrix for modelled transmissions occurring in 2016. Values are the mean, over the entire year, of the per-year transmission rates from infected individuals in one age group to individuals in another (given by the number of transmissions divided by the size of the source population).

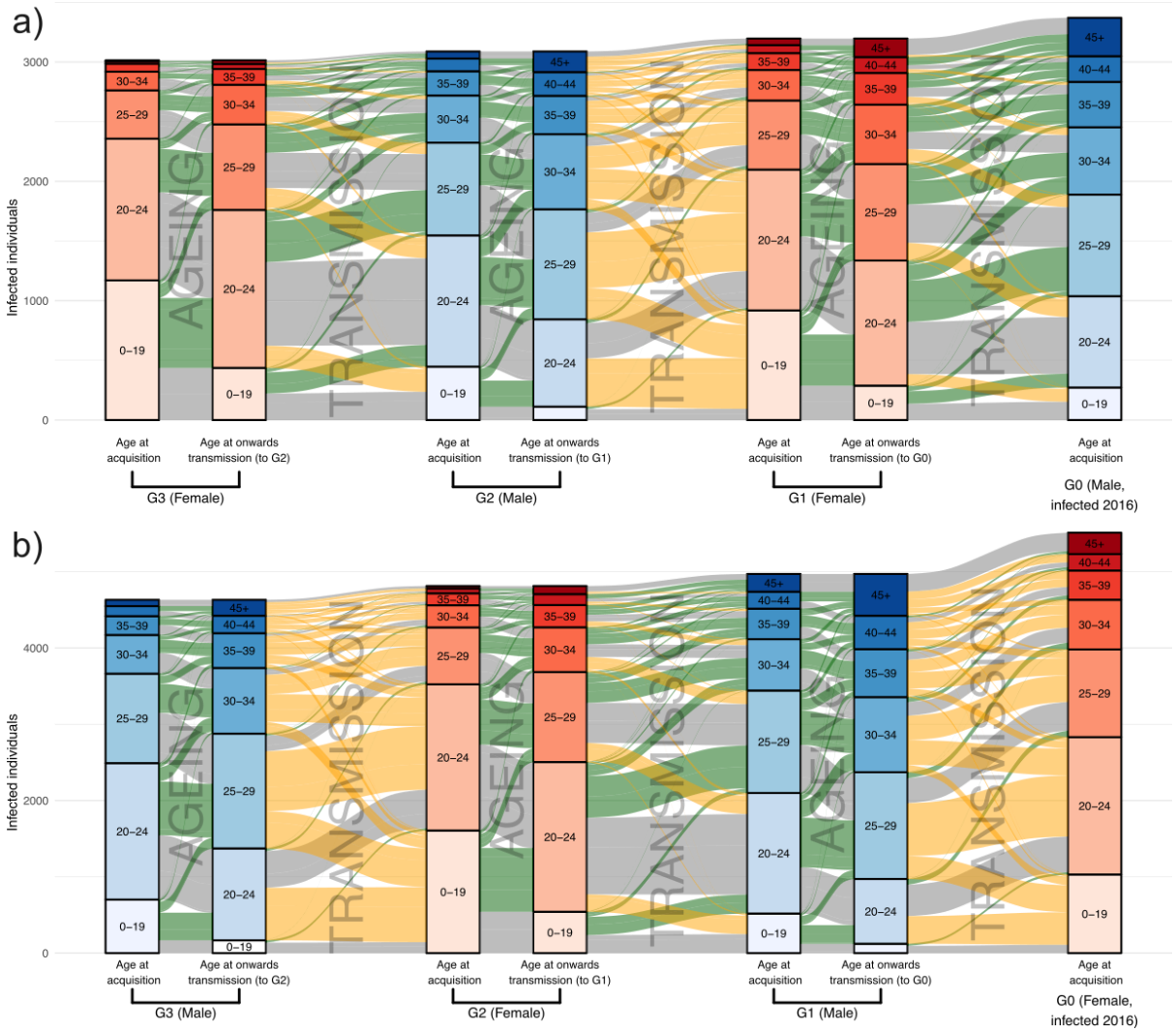


Figure 4: Transmissions to a younger cohort take place predominantly in male-to-female transmissions. Alluvial plots for the transmission history, derived from the IBM, of a cohort of males (panel a) and females (panel b) infected with HIV in 2016. That cohort is G0, shown on the right; to the left of it are the three previous generations of transmission (indicated as G1 to G3) that gave rise to the infections of G0. Male generations are blue and female red. The partitioning within each vertical bar indicates its age distribution. The age distributions of G1 to G3 are shown at both the time of their own infection (left) and the time at which they infected an individual in a subsequent generation (right). Movement to an older age group (by ageing or transmission) is shown by a green alluvium, to a younger group by an orange

alluvium, and to the same age group by a grey alluvium If an individual was responsible for two or more infections in the subsequent generation, then they still contribute a total of 1 individual to their “age at onwards transmission” bar, which is split proportionally between age bands depending on the number of infections they were responsible for while they were a member of each band. (For example, if individual A infected two individuals while A was aged 20–24 and A infected one while they were aged 25–29, A would contribute ⅔ of an individual to the 20–24 bar and ⅓ to the 25–29 bar.) Note that the members of G0 are all members of a particular cohort who were infected with HIV, butall other generations are composed solely of people who transmit HIV, a somewhat different group.

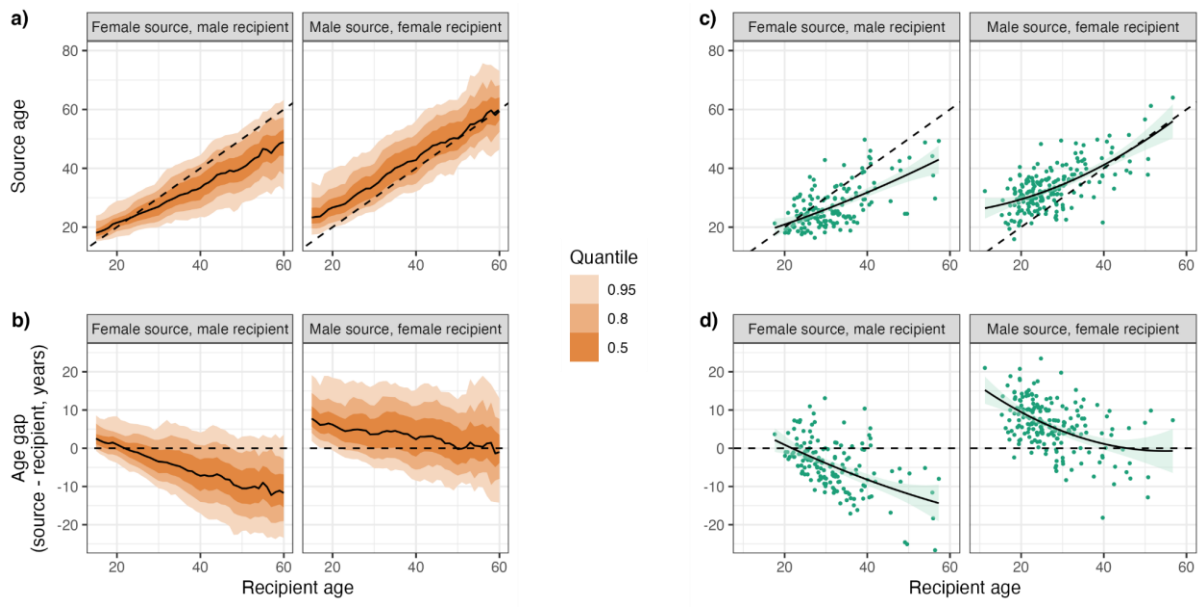


Figure 5: The age gap increases with recipient age in female-to-male pairs and decreases in male-to-female pairs. Relationship between recipient age and source age (a, c) and age gap (b, d) from the IBM (a, b) and the phylogenetics (c, d). The dashed line represents equal ages. The IBM results are from the “best-fitting” simulation (that with the minimum ABC distance) and represent the median value (solid black line) and quantiles (orange shaded areas) of the y-axis variable in 1-year increments. The individual values from the phylogenetics pairs are plotted as green points, with a regression line fitted using quadratic polynomial regression and the 95% confidence region given by the pale green area.

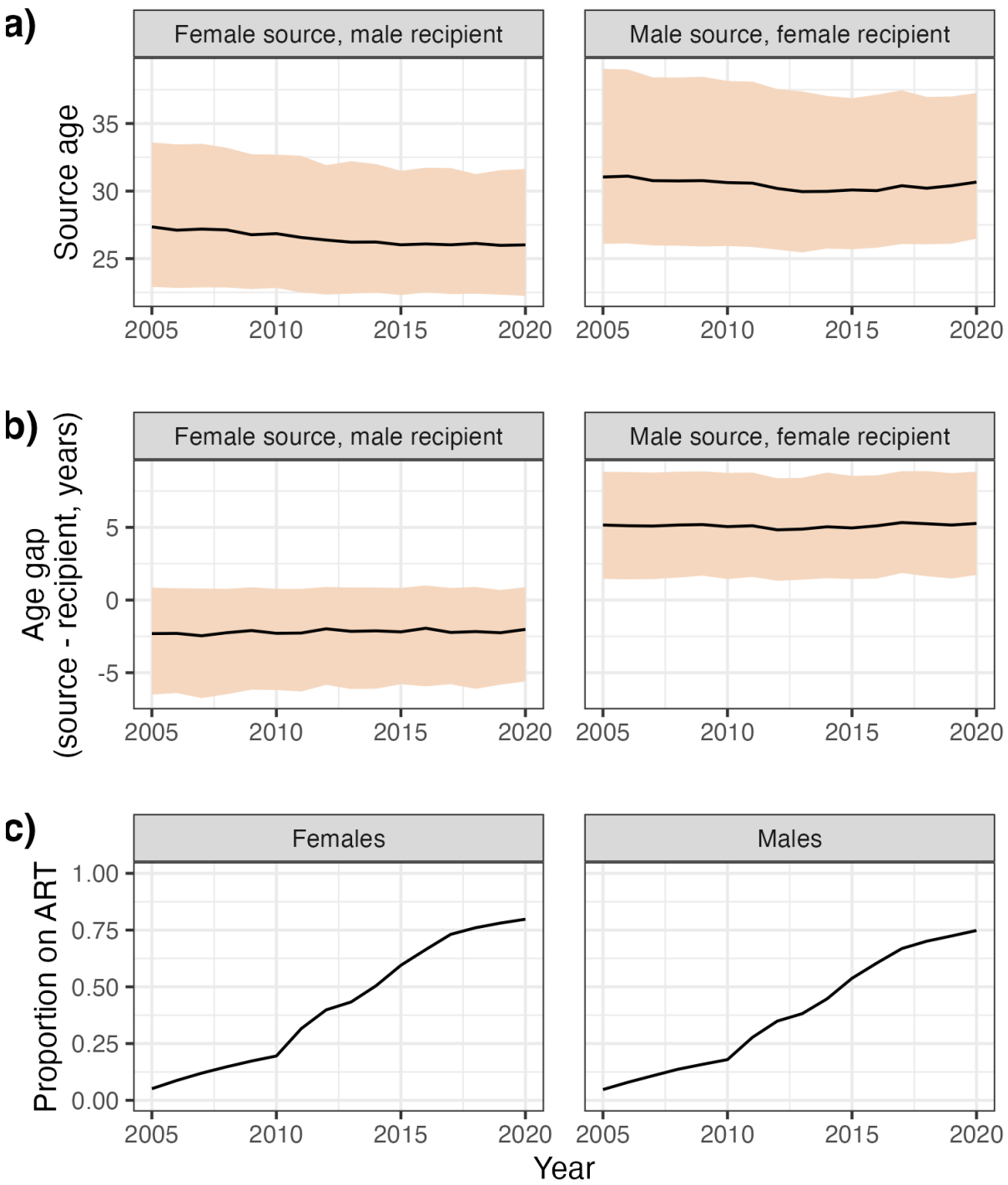
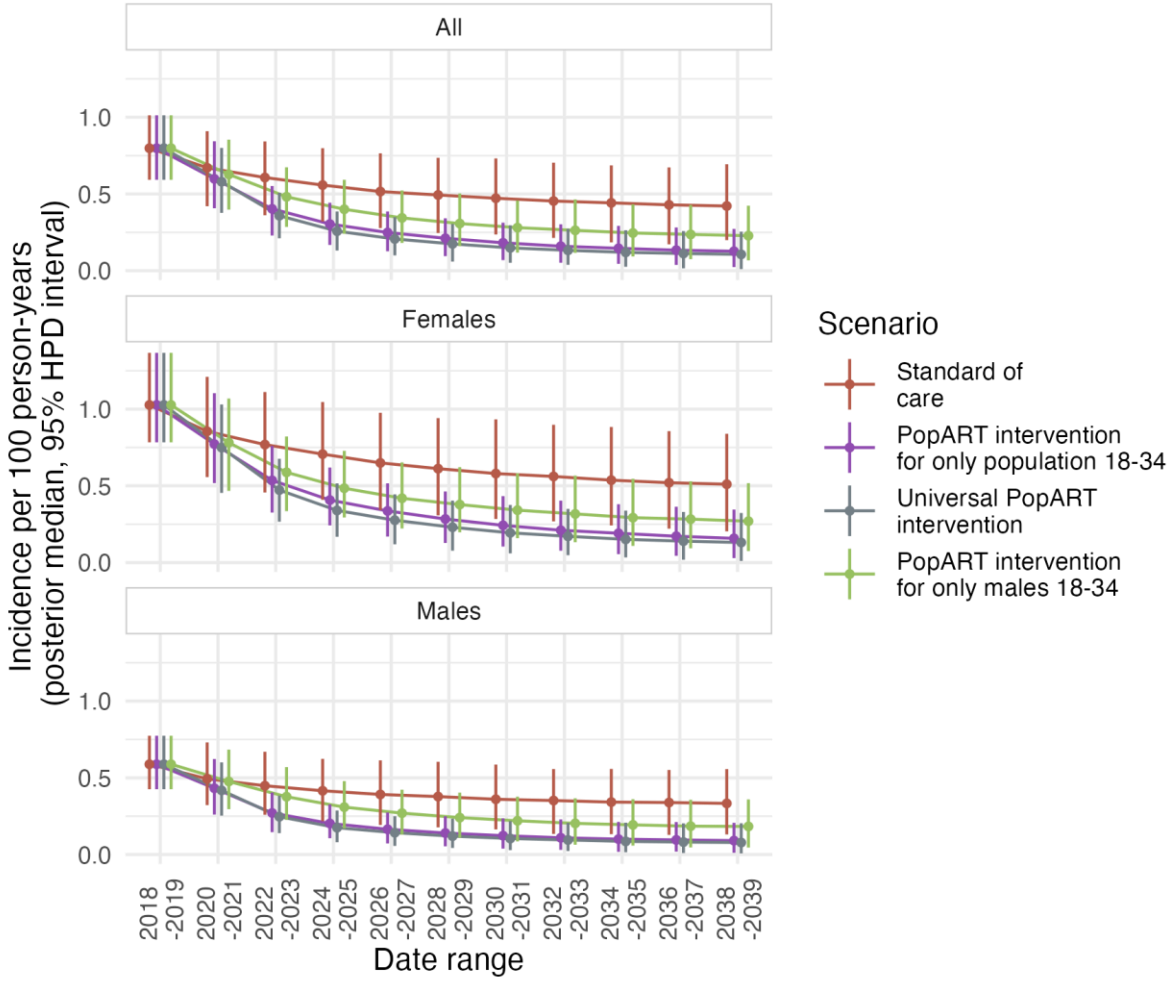


Figure 6: Little change in age gaps over the period of ART rollout. Time trends in age of sources (a), age gaps (b) and proportion of the population on ART (c) over time. Estimates are from the best-fitting IBM simulation. The black line in (a) and (b) is the median as calculated per year, with the shaded area representing the interquartile range.

A


All
Females
Males
1.0
0.5
0.0
Incidence per 100 person-years
(posterior median, 95% HPD interval)
2018
-2019
2020
-2021
2022
-2023
2024
-2025
2026
-2027
2028
-2029
2030
-2031
2032
-2033
2034
-2035
2036
-2037
2038
-2039
Date range
Scenario
Standard of care
PopART intervention for only population 18-34
Universal PopART intervention
PopART intervention for only males 18-34

B

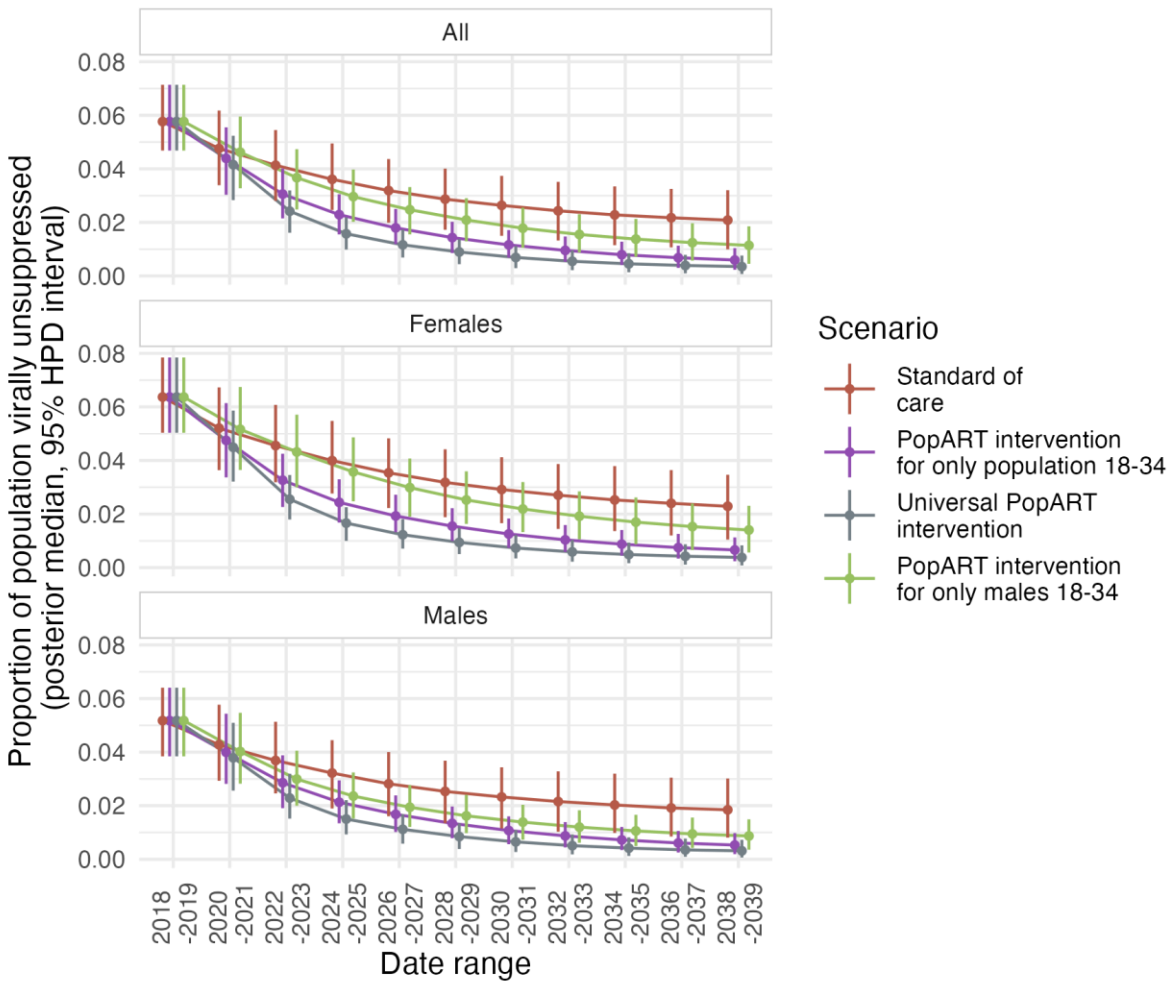


Figure 7: The impact of an intervention focussing on younger people. Projected time trends in HIV incidence (A) and population proportion of people with unsuppressed HIV (B), 2018–2041, in the Zambian control communities of the PopART trial, under various scenarios of future PopART intervention rollout. Posterior median estimates are calculated in two year time periods for the whole population (top), females (middle) and males (bottom). The four scenarios are, firstly, no CHiPS deployment at all (red), provision of the full PopART intervention to the whole population (purple), PopART intervention targeting only those aged under 35 (grey), and PopART intervention targeting only men aged under 35 (green). Vertical bars indicate the 95% highest posterior density interval.

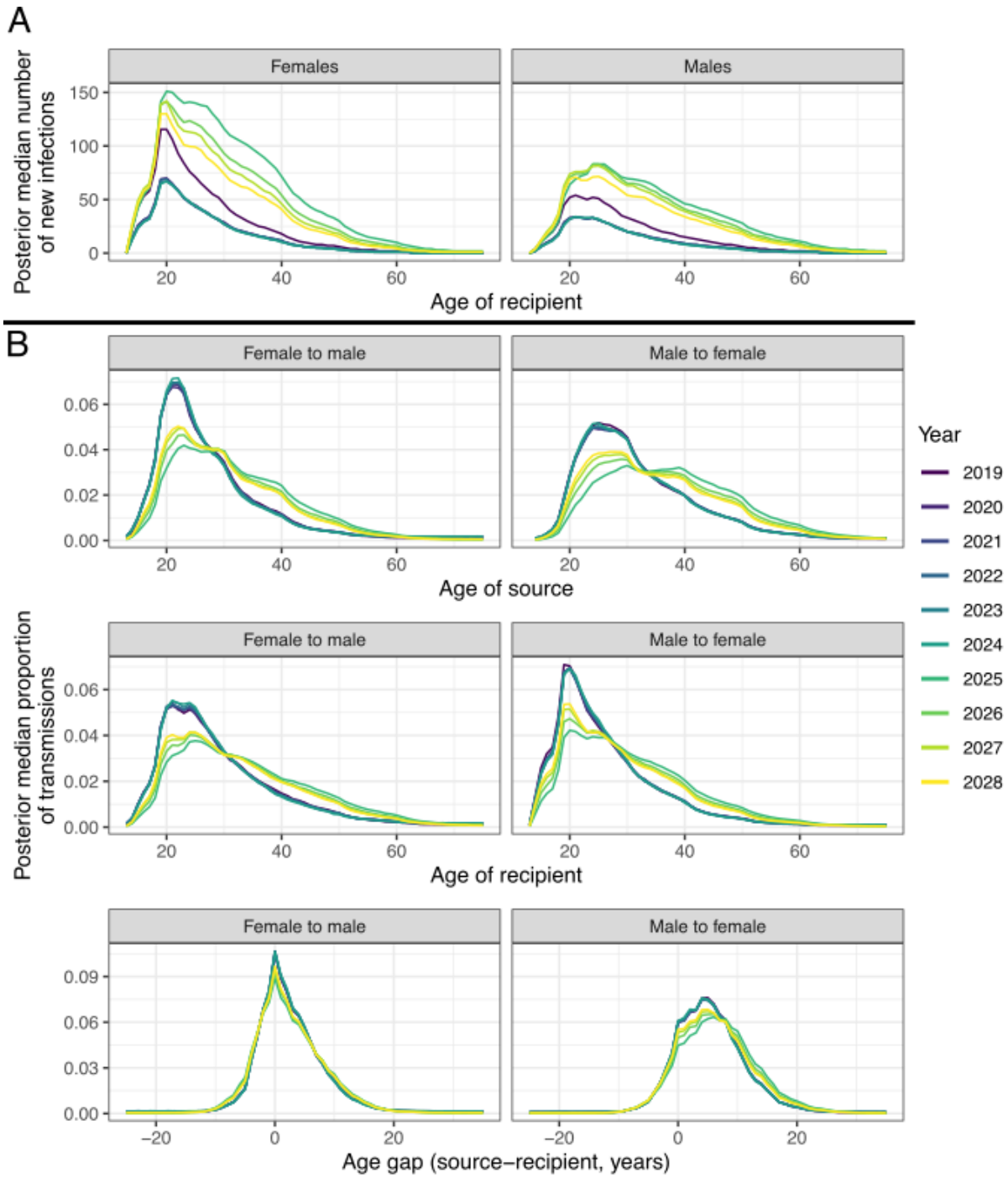
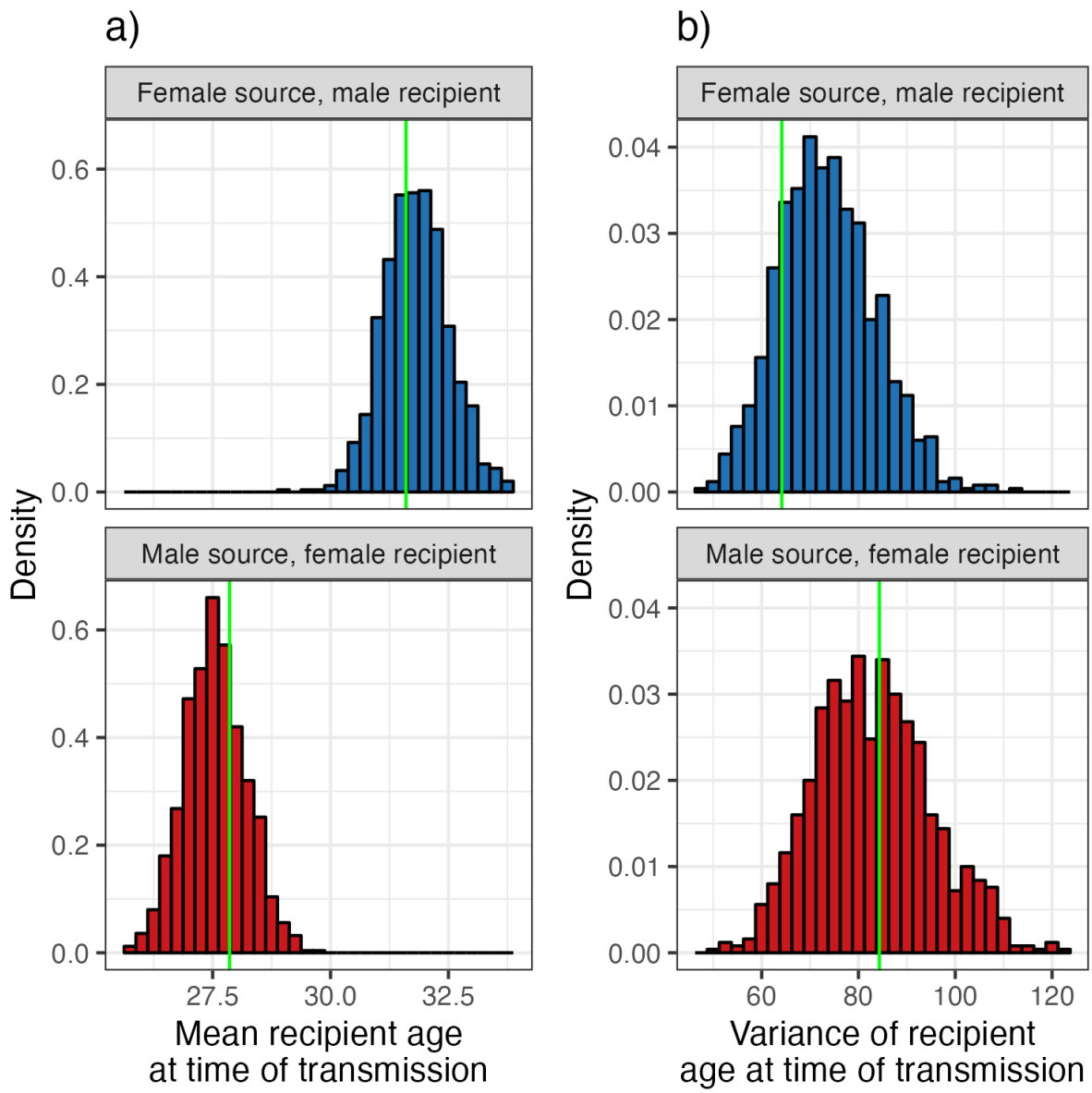
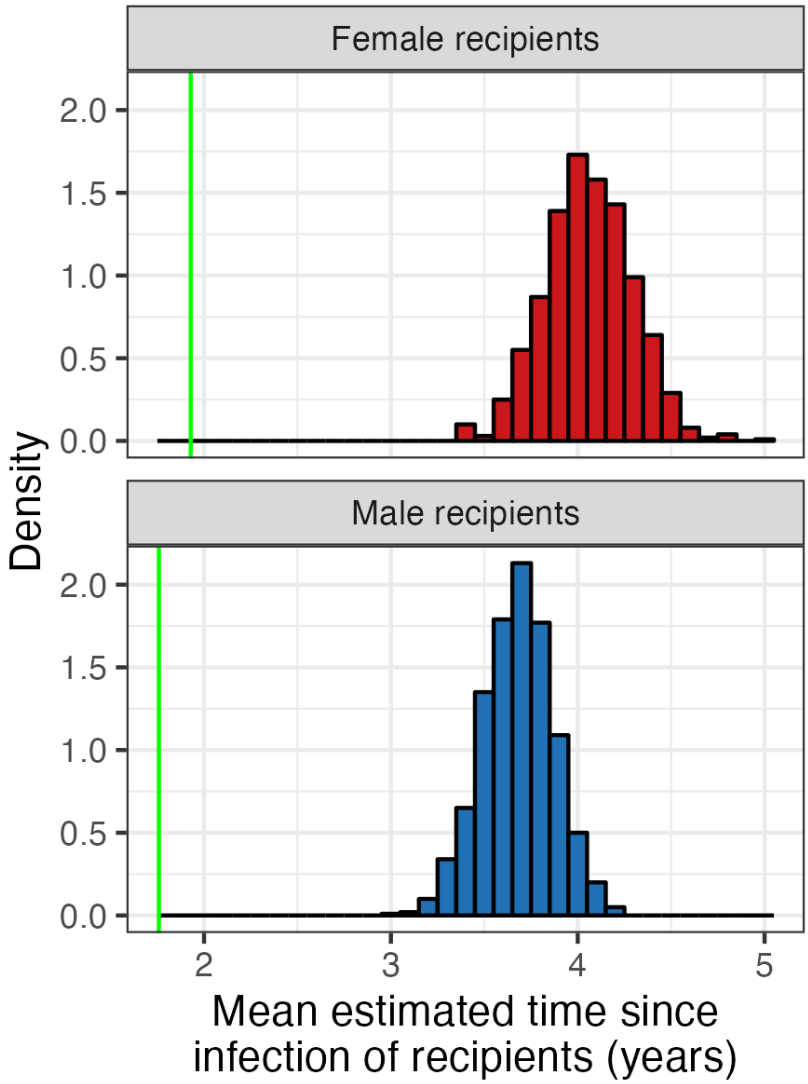


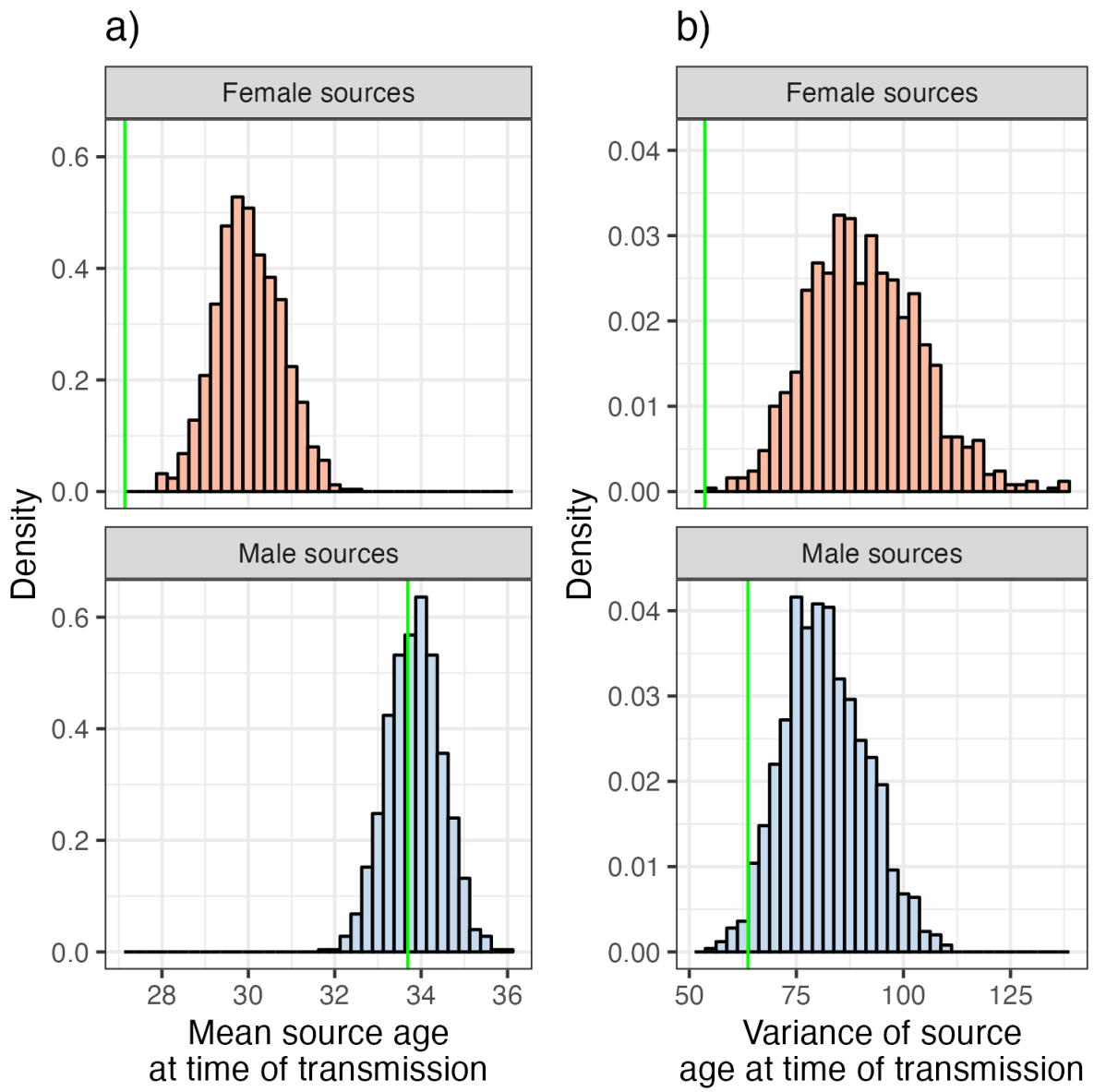
Figure 8: The impact of ART interruption. A: Posterior median new infections by year, by sex and age. B: The age distribution of new transmissions by year in the circumstances of a complete cessation of ART provision during 2025, returning in 2026. Lines chart the posterior median proportion of new infections for sources and recipients of each integer age (top and middle), and age gap rounded to the nearest integer (bottom). In both cases, lines are coloured by year.



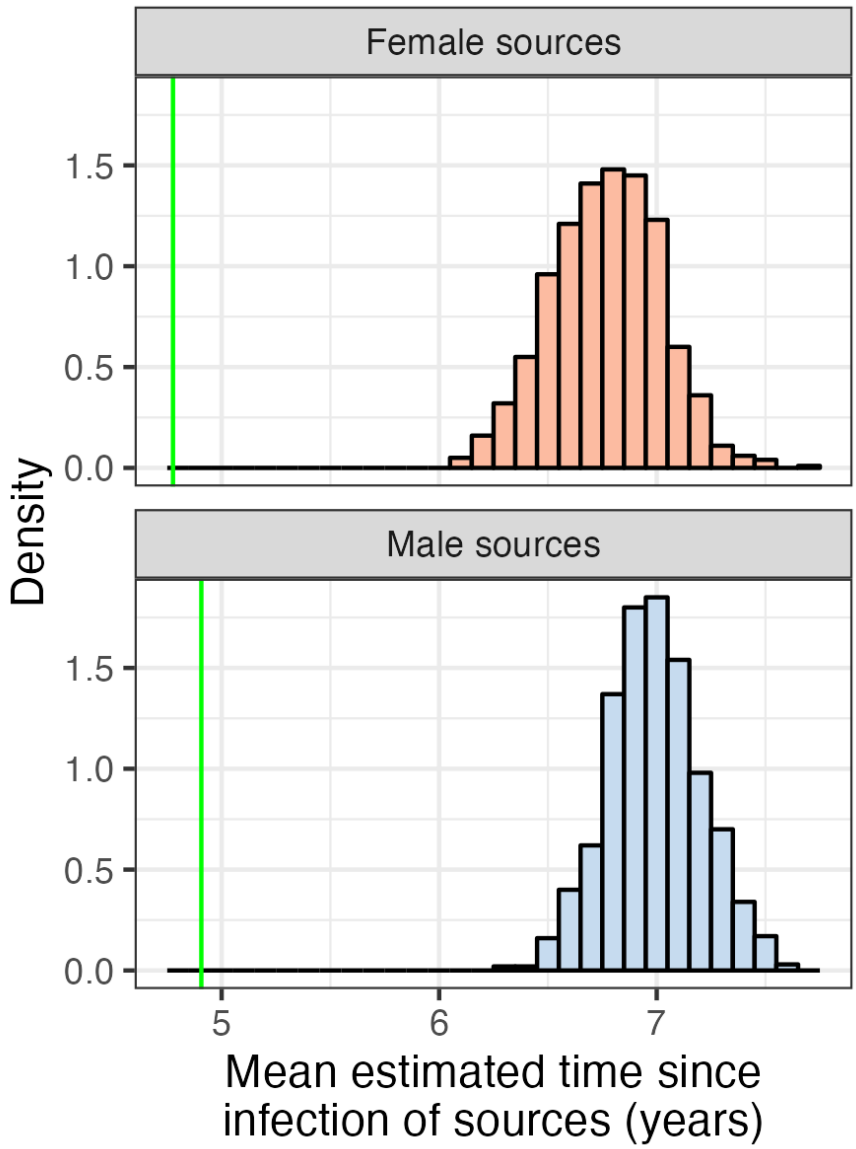
Appendix 1-Figure 1: Empirical distributions of mean (a) and variance (b) of ages of recipients at time of transmission from pairs where “recipients” had randomly chosen “sources” (histograms) compared to the true value from the phylogenetics data (green line).



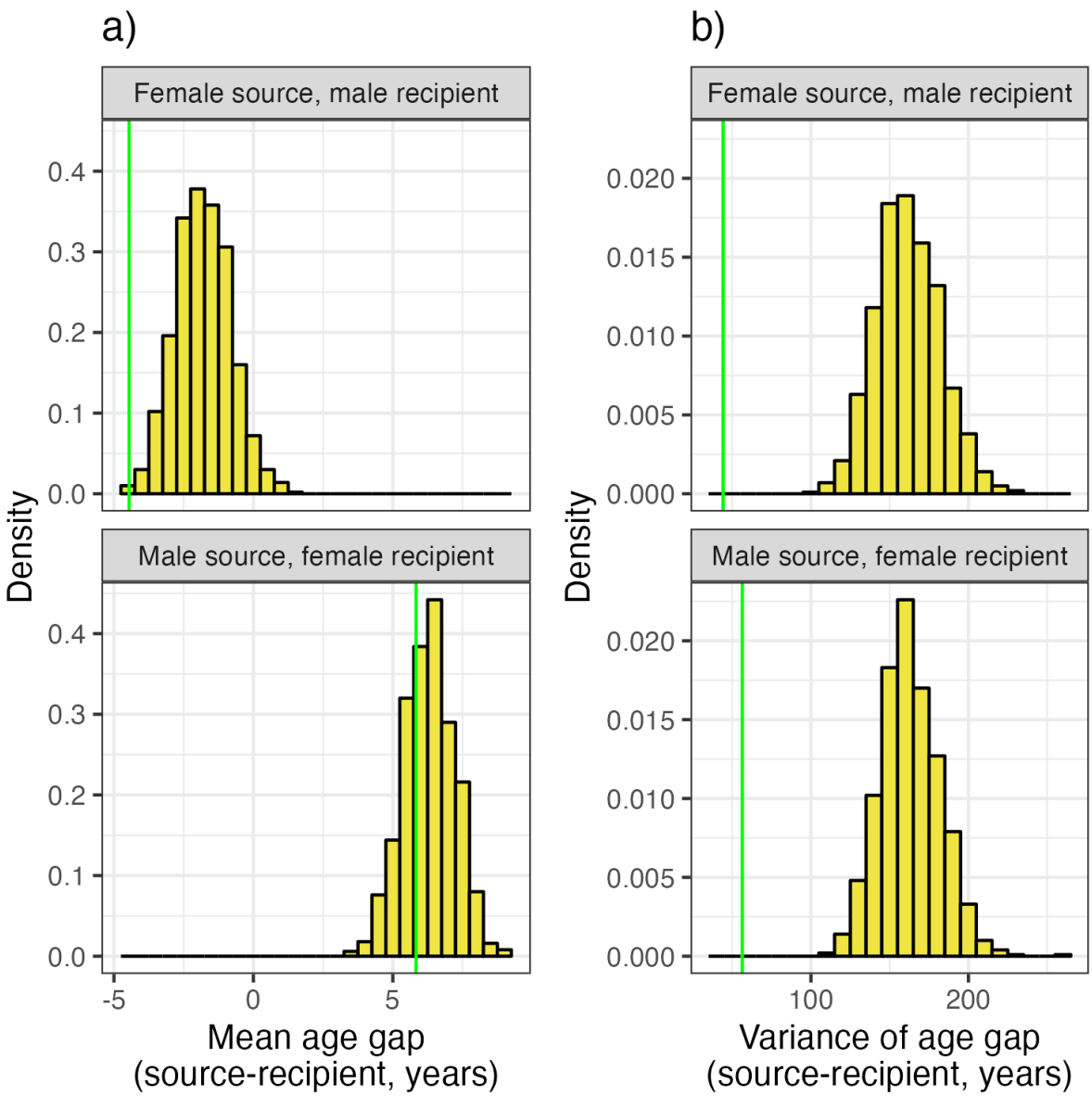
Appendix 1-Figure 2: Empirical distributions of mean estimated time from infection to sampling of recipients from randomised transmission pairs (histograms) compared to the true value from the phylogenetics data (green line).



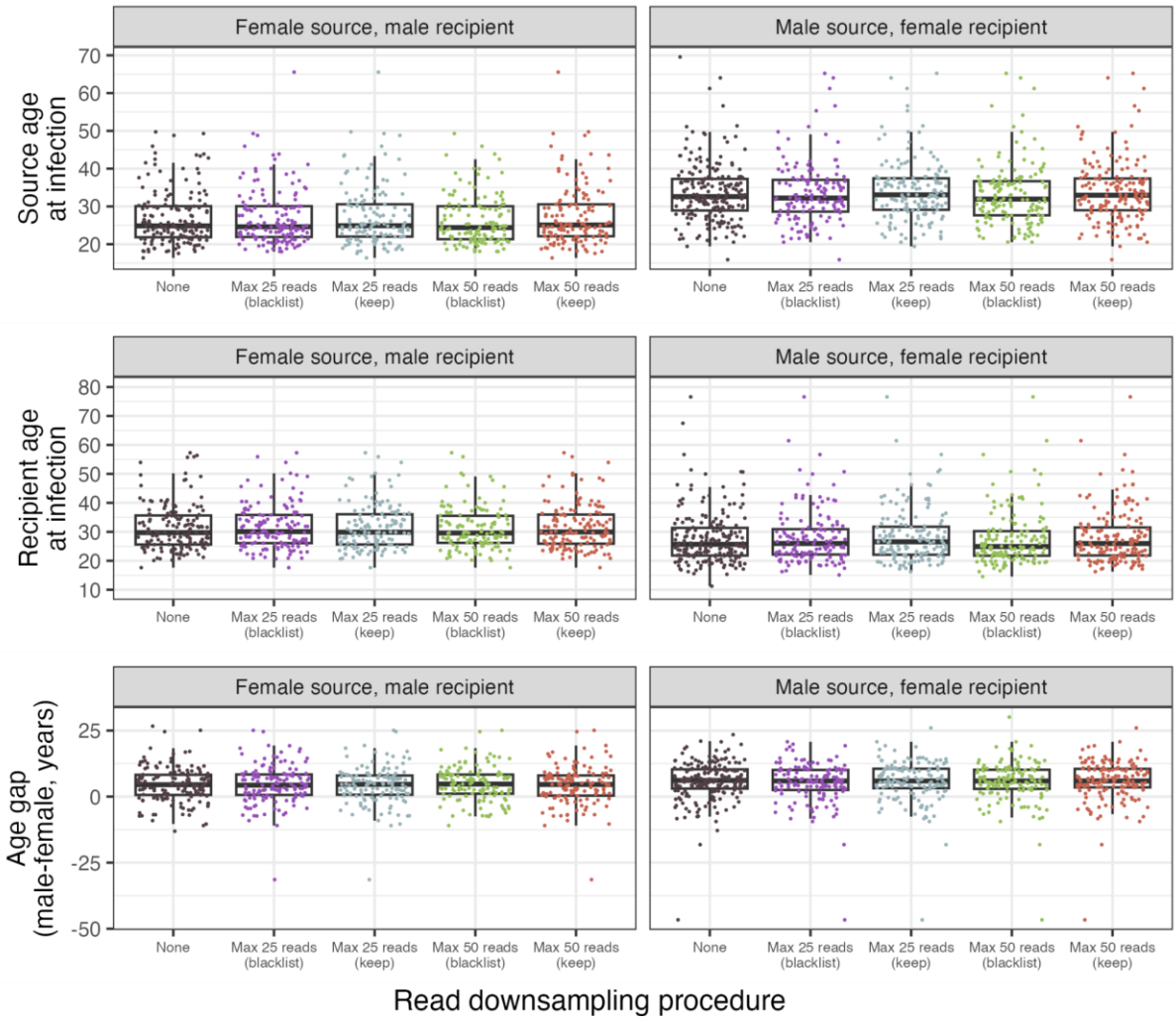
Appendix 1-Figure 3: Empirical distributions of mean (a) and variance (b) of ages of sources at time of transmission from randomised transmission pairs (histograms) compared to the true value from the phylogenetics data (green line).



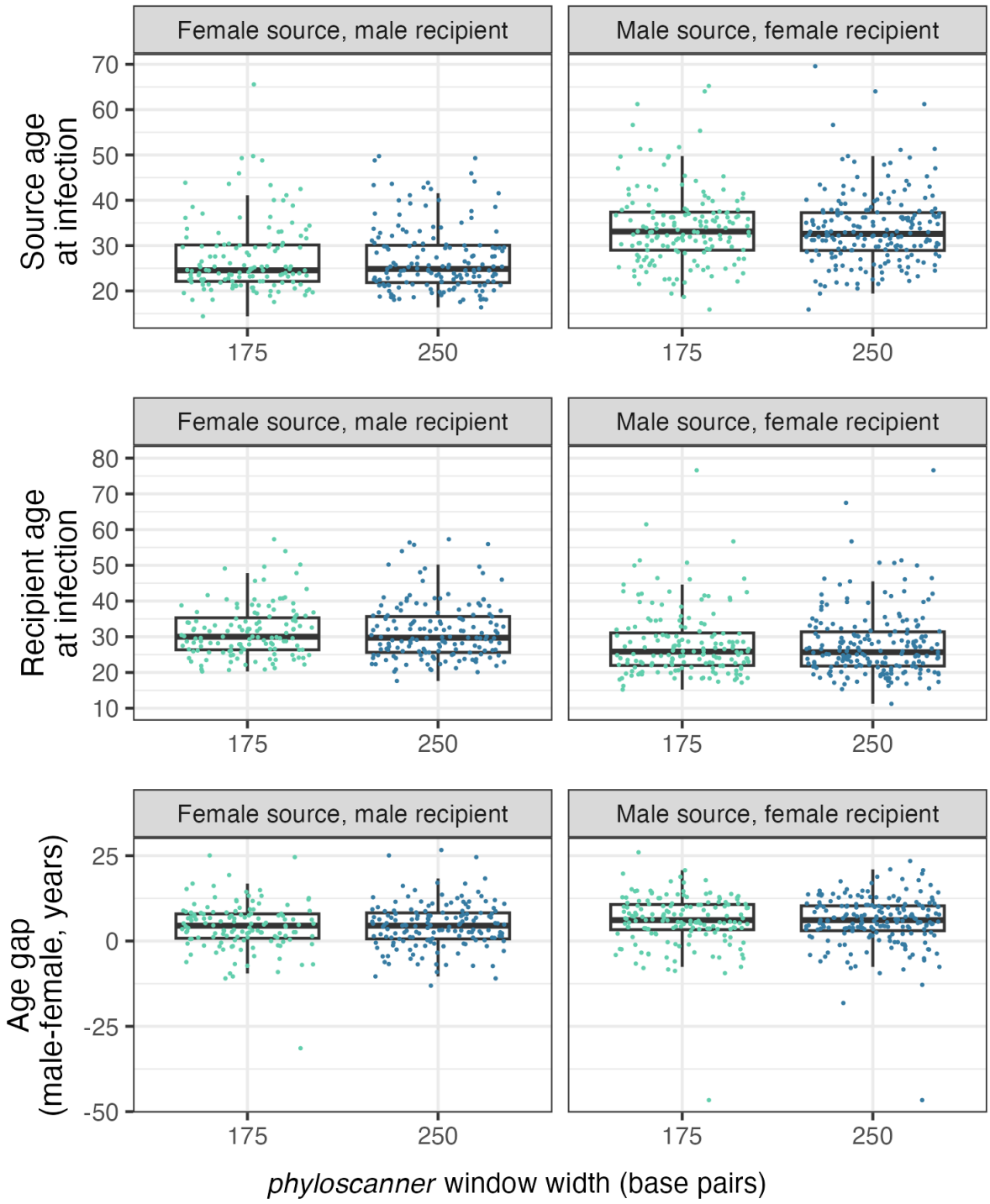
Appendix 1-Figure 4: Empirical distributions of mean estimated time from infection to sampling of sources from randomised transmission pairs (histograms) compared to the true value from the phylogenetics data (green line).



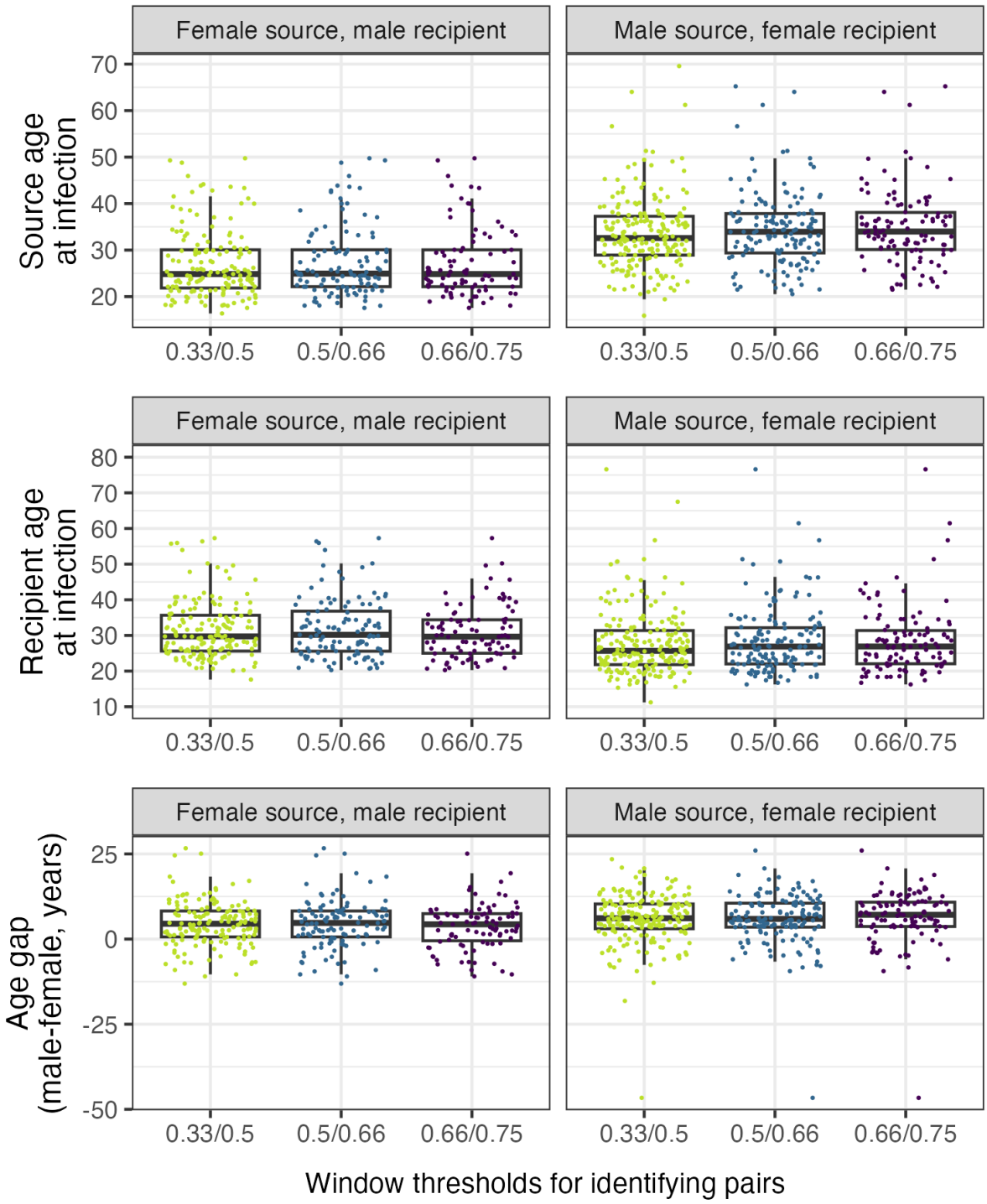
Appendix 1-Figure 5: Empirical distributions of mean (a) and variance (b) in transmission age gap from randomised transmission pairs (histograms) compared to the true value from the phylogenetics data (green line).



Appendix 2-Figure 1: Distributions of the estimated ages of sources at the time of transmission (top), estimated ages of recipients at the time of transmission (middle) and age gap amongst probable transmission pairs identified using *phyloscanner* when the read downsampling procedure is varied. Downsampling to a maximum of *n* reads means that a random sampling of *n* reads is taken for each individual in each genomic window. For windows in which *n* reads do not exist, the two options are “blacklist” - to exclude that host from that window entirely - and “keep” - to use all the reads that are available. The boxplots display the median, 25% and 75% quantiles, and range.

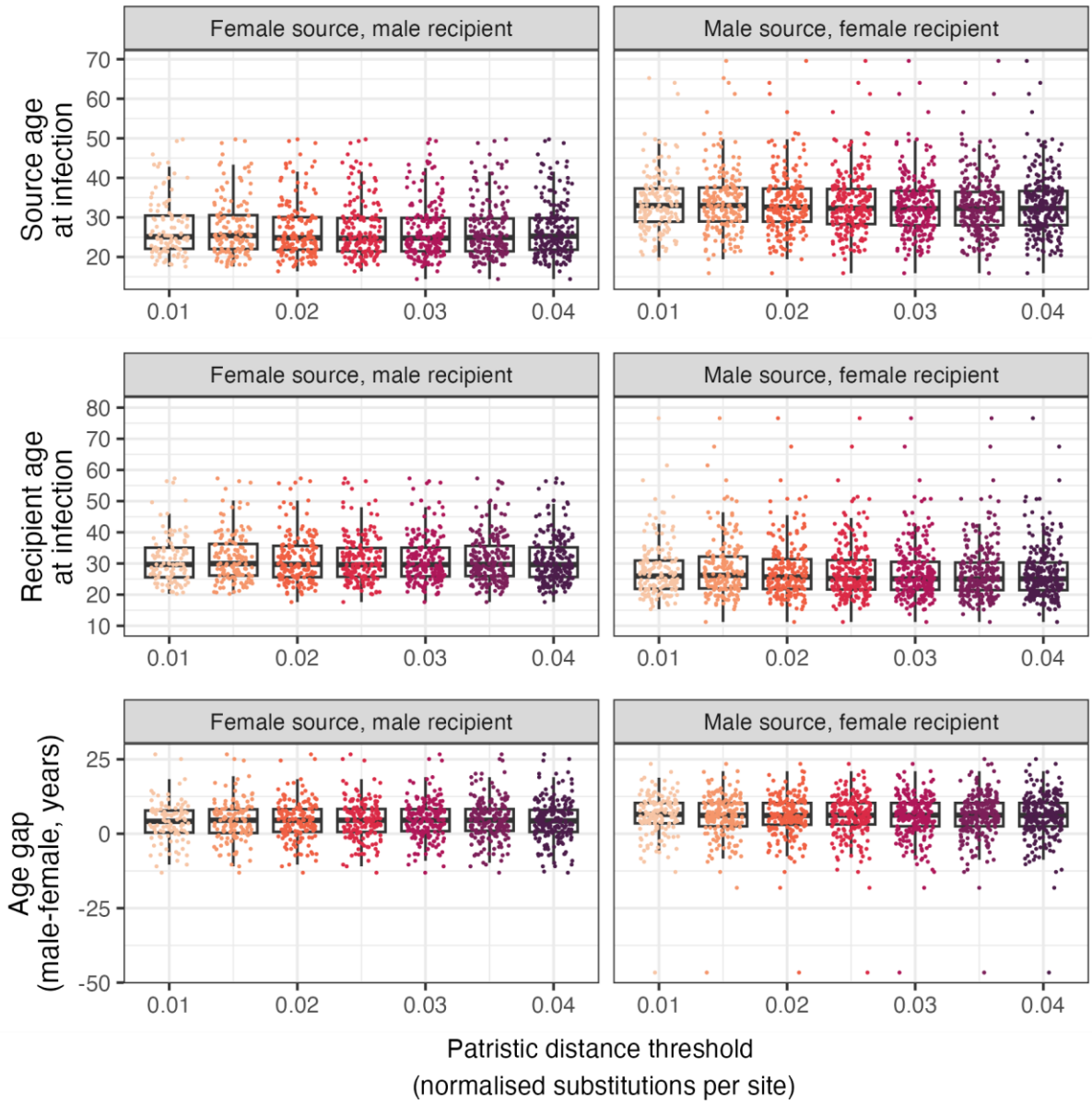


Appendix 2-Figure 2: Distributions of the estimated ages of sources at the time of transmission (top), estimated ages of recipients at the time of transmission (middle) and age gap amongst probable transmission pairs identified using *phyloscanner* with different genomic window widths. The boxplots display the median, 75% quantiles, and range.



Appendix 2-figure 3: Distributions of the estimated ages of sources at the time of transmission (top), estimated ages of recipients at the time of transmission (middle) and age gap amongst probable transmission pairs identified using *phyloscanner*. Different window-

related thresholds for identifying linkage and direction are indicated on the x axis in the format X/Y, where Y is the minimum fraction of genomic windows supporting two samples being from a transmission pair, and X is the minimum fraction of genomic windows supporting transmission having an identifiable direction (conditional on being a transmission pair). The boxplots display the median, 75% quantiles, and range.



Appendix 2-Figure 4 Distributions of the estimated ages of sources at the time of transmission (top), estimated ages of recipients at the time of transmission (middle) and age

gap amongst probable transmission pairs identified using *phyloscanner* with patristic distance thresholds used to identify transmission pairs in an individual window. The boxplots display the median, 75% quantiles, and range.