Supplementary Online Material and Methods

SOM 1 Procedure to identify potential transmitters of recipient MSM

To reconstruct an evidence base of past transmission events amongst MSM in the Netherlands between July 1996 and December 2010, we first identified MSM for whom a narrow infection window could defined (see Materials and Methods in the main text). Next, we considered as potential transmitters all registered infected men that could have in principle infected a recipient. Potential transmitters were defined as infected men in the ATHENA cohort that overlap with the infection window of a recipient MSM. To determine if an infected individual overlapped with an infection window, we need to estimate when the individual in question became infected. Equivalently, we here estimate the time from infection to diagnosis, which we denote by $T_i^{I \to D}$ for individual i. This section describes how individual-level time to diagnosis estimates were obtained. We denote the estimated time to diagnosis for individual i by $\hat{T}_i^{I \to D}$. Estimated infection times are associated with substantial uncertainty and sensitivity analyses were conducted for lower and upper 95% estimates. Findings did not depend substantially on these infection time estimates (figures S16 and S21).

We adapted a previously published method that estimates an individual's time to diagnosis based on certain risk variables at time of diagnosis (*41, 45*). This approach proceeds in two steps. First, HIV surveillance data from an MSM cohort of drug naïve HIV seroconverters are used to estimate the association between the time to diagnosis since the midpoint of the seroconversion interval and risk variables at diagnosis. This association is described with a suitable regression model. Next, the fitted regression model is used to predict the expected time to diagnosis for all infected individuals. Previous work found that CD4 cell count at diagnosis, age at diagnosis, infection route and ethnicity are significantly associated with the time to diagnosis since the midpoint of the seroconversion interval (*45*). Here, ethnicity was not available and infection route was always MSM. Both demographic variables were not considered in this analysis.

The previous method to estimate an individual's time of infection assumes, first, that the time between the midpoint of the seroconversion interval to diagnosis is representative of the unknown time to diagnosis among seroconverting MSM. We denote the time to diagnosis from the midpoint by $\tilde{T}_i^{I \to D}$ for seroconverter i. Second, the previous approach assumes that the approximated time to diagnosis among seroconverting MSM is representative of the time to diagnosis among all infected MSM. Here, we adapt this approach in order to relax both assumptions, using the $\tilde{T}_i^{I \to D}$ as an intermediate step to obtain the final estimate $\hat{T}_i^{I \to D}$.

In the ATHENA cohort, data on 3,025 MSM with a last negative test and date of diagnosis between 2003/01-2010/12 were available to estimate the association between $\tilde{T}_i^{I \to D}$ and risk variables at time of diagnosis. Table S4 characterizes these MSM with a last negative test.

We conducted an exploratory data analysis, shown in figure S24, which indicated that infection status at time of diagnosis (evidence for infection within 12 months prior to diagnosis), age at

diagnosis, status of HIV infection at diagnosis and (to a lesser extent) the first CD4 count within 12 months of diagnosis are associated with $\tilde{T}_i^{I\to D}$ among drug naïve MSM with a last negative test.

For individuals in confirmed recent infection at time of diagnosis, we set

$$\hat{T}_i^{I\to D} = 1 \text{ year.}$$

For all other individuals, we estimated first $\tilde{T}_i^{I\to D}$ from age and first CD4 count at time of diagnosis. Based on the exploratory data analysis shown in figure S24, we fitted the regression model

$$\tilde{T}_i^{I\to D} \sim Gamma(\mu_i, \phi_i),$$

$$\log \mu_i = \beta_0 + R^{Nind}\left(\beta_1^{Nind} C_i^{850} A_i + \beta_2^{Nind} C_i^{250-850} A_i + \beta_3^{Nind} C_i^{250} A_i + \beta_4^{Nind} C_i^{NA} A_i\right) + R^{miss}\left(\beta_1^{miss} C_i^{850} A_i + \beta_2^{miss} C_i^{250-850} A_i + \beta_3^{miss} C_i^{250} A_i + \beta_4^{miss} C_i^{NA} A_i\right)$$

$$\log \phi_i = \gamma_0 + R^{Nind}\left(\gamma_1^{Nind} C_i^{850} A_i + \gamma_2^{Nind} C_i^{250-850} A_i + \gamma_3^{Nind} C_i^{250} A_i + \gamma_4^{Nind} C_i^{NA} A_i\right) + R^{miss}\left(\gamma_1^{miss} C_i^{850} A_i + \gamma_2^{miss} C_i^{250-850} A_i + \gamma_3^{miss} C_i^{250} A_i + \gamma_4^{miss} C_i^{NA} A_i\right)$$

among MSM with a last negative test, where

$\tilde{T}_i^{I\to D}$	time between the midpoint of the seroconversion interval and diagnosis
μ_i	mean of the Gamma distribution for the ith seroconverter
ϕ_i	dispersion of the Gamma distribution for the ith seroconverter
β_k	location parameters
γ_k	shape parameters
R_i^{Nind}	**1**(ith seroconverter with recent infection at diagnosis not indicated)
R_i^{miss}	**1**(ith seroconverter with missing infection status at diagnosis)
C_i^{850}	**1**(ith seroconverter with first CD4 count > 850 cells/ml within 12 months after diagnosis and before ART start)
$C_i^{250-850}$	**1**(ith seroconverter with first CD4 count in [250-850] cells/ml within 12 months after diagnosis and before ART start)
C_i^{250}	**1**(ith seroconverter with first CD4 count < 250 cells/ml within 12 months after diagnosis and before ART start)
C_i^{NA}	**1**(ith seroconverter with no first CD4 count within 12 months after diagnosis and before ART start)
A_i	min(age at diagnosis of ith seroconverter, 45).

All regression coefficients were significant. Figure S25 illustrates the predictions obtained with the fitted multivariable regression model. The fitted regression model explained 53% of the variance in the observed $\tilde{T}_i^{I\to D}$ among MSM with a last negative test.

Rice et al. (*41*) used the expected $\tilde{T}_i^{I\to D}$ as an estimate of the unknown time to diagnosis. To relax the two underlying assumptions noted earlier, we used instead a particular upper quantile α of the estimated probability density function of $\tilde{T}_i^{I\to D}$. Figure S26 illustrates the probability density function of $\tilde{T}_i^{I\to D}$ which was obtained from the parameters of the fitted regression model. The upper quantile parameter α was estimated so that the average $\hat{T}_i^{I\to D}$ is consistent with the average time to diagnosis derived from two mathematical modelling studies between 1996 and 1999. We chose this period in order to validate if the predictive model can reproduce previously estimated reductions in average

time to diagnosis in subsequent years. For this period, Bezemer et al. (*46*) estimated an average time to diagnosis amongst MSM of 3.16 years (95% confidence interval 3.00-3.41 years) in this period. van Sighem et al. (*7*) estimated a mean time to diagnosis of 4.34 years (3.87-5.11 years) by the end of 1999. We chose the quantile parameters $\alpha = 0.109, 0.148, 0.194$ in correspondence to these estimates (figure S27-A). The fact that the chosen quantile parameters are substantially lower than 0.5 indicates that the expected time to diagnosis since the midpoint of the seroconversion interval cannot be considered representative of the time to diagnosis among infected MSM. We used these quantile parameters to obtain central, lower, and upper individual-level time to diagnosis estimates, as shown in Figure S27-B.

Overall, the individual-level predictive model is able to reproduce previously estimated reductions in average time to diagnosis without the addition of time-dependent variables (black lines in Figure S27-B) (*7, 46*). The linear drop in time to diagnosis after 2005 may in part be explained by right censoring in the cohort: as the study endpoint was 2010/12 and the maximum estimated time to diagnosis is around 7 years, we expect that an increasing fraction of men infected since 2004-2005 is not yet diagnosed. In comparison to Rice et al. (*41*), our approach results in larger estimates of time to diagnosis. If the 50% quantile had been used to estimate times to diagnosis, the average time to diagnosis for MSM infected in the period 1996-1999 would have been slightly less than 2 years (figure S27-A).

SOM 2 Procedure to declare potential transmission pairs phylogenetically implausible

HIV sequences cannot prove epidemiological linkage nor the direction of HIV infection (*11, 12*). However, viral sequences can be used to exclude potential transmission events between individuals whose viral sequences are phylogenetically unrelated. There is currently no widely agreed consensus on viral phylogenetic exclusion criteria (*8*).

To guide the viral phylogenetic exclusion criteria adopted in this study, we conducted an evolutionary analysis of sequences from transmitters and recipients in confirmed transmission pairs. This analysis is described in figure S5. In addition, we considered 4,117 pairs of sequences from the same Dutch patient and 201,605 pairs between Dutch patients that died before the last negative antibody test of the other patient. These analyses are described below, and were used to develop exclusion criteria with high specificity (i.e. small type-I error of falsely excluding true transmission pairs). We chose central exclusion criteria for the main transmission analysis and varied lower and upper criteria over the identified range. Sensitivity analyses demonstrate these criteria did not impact substantially on the reported transmission and prevention analyses.

Figure S5 shows the genetic distance between sequences from confirmed pairs in the Belgium transmission chain as a function of time elapsed. It is clear that the genetic distance between sequences from confirmed pairs can exceed typical phylogenetic clustering thresholds, provided the time elapsed is sufficiently large. This analysis indicates that genetic distances of not more than 2% between partial HIV *pol* sequences from true transmission pairs are only expected when the total time elapsed is small. This is typically the case when individuals are frequently followed up as in a controlled, randomized trial (*47*). Among the phylogenetically probable transmission pairs in this study, the maximum time elapsed was 10.87 years. Considering figure S28, the corresponding upper 97.5% quantile of the genetic distances between sequences from true transmission pairs is ~ 7%. To validate the analysis in figure S5, we estimated the genetic distance between sequences from

confirmed transmission pairs in the Swedish transmission chain in the same manner. Figure S28 shows that these genetic distances fall into the 80% probability range estimated from the Belgium transmission pairs. This argues against tight genetic distance thresholds to declare transmission pairs phylogenetically implausible in this study.

To exclude potential transmission pairs, we used the following two criteria:

- Bootstrap clade support. If the potential transmitter did not occur in the same clade as the recipient MSM in sufficiently many bootstrap phylogenetic trees, the pair was excluded. Such bootstrap criteria are frequently used (*8*).
- Phylogenetic incompatibility with direct transmission. We found that within phylogenetic clades with high bootstrap support, branches between the remaining potential transmitters and the recipient MSM were often relatively long (figure S3). With approximately half of all potential transmitters sampled, one explanation is that the actual transmitter did not have a sequence sampled or was not diagnosed by March 2013. Unobserved intermediate transmitters were detected with a coalescent compatibility test that was recently introduced by Vrancken et al. (*13*). The idea behind this test is that viral lineages of a true transmission pair must coalesce at a time when the transmitter was already infected. The test assumes that transmitters are infected with a single virus. The test calculates the probability that the viral lineages from the potential transmitter and the recipient coalesce after the transmitter was infected and before the recipient was diagnosed. The test excludes the potential pair if this coalescent compatibility probability is below a certain threshold. To apply this test, we dated coalescent events within phylogenetic clusters. Specifically, the sampled ancestor birth-death model was used in order to allow for the possibility that transmission might have occurred after the time of sequence sampling. To accommodate temporal variation in model parameters, we implemented a skyline version of the sampled ancestor birth-death model along previous work (*48*).

We then sought to determine thresholds so that potential transmitters are excluded with high specificity (a large proportion of true transmitters to recipients is not excluded). Typically, viral phylogenetic studies aim to identify transmission chains (*23*). This leads to relatively strict thresholds. Here, we aim to exclude pairs of individuals that did not infect each other. This different objective leads to relatively large thresholds.

For the clade frequency criterion, the type-I error is the probability that sequence pairs of a true transmission pair do not co-cluster. As a proxy, we calculated the probability that sequence pairs from the same individual do not co-cluster. Figure S29 shows this probability as a function of the clade frequency threshold. The approximate type-I error is more than 10% for clade frequency thresholds above 85%. To limit this error, we settled on 80% as the central clade frequency threshold, and considered 70% and 85% as the upper and lower thresholds respectively.

To determine the threshold of the coalescent compatibility test below which potential transmission pairs are excluded, we proceeded as for the phylogenetic clustering test.
We approximated the type-I error with the probability that co-clustering sequence pairs from the same individual were excluded by the coalescent compatibility test. Figure S30-A shows this probability as a function of the coalescent compatibility threshold. The approximate type-I error is around 5% for thresholds in the range of 10% to 30%. We chose 20% as the central threshold and considered 10% and 30% in sensitivity analyses.

We also evaluated the power of the test in excluding co-clustering female-female pairs. All female-female pairs were considered as incorrect transmission pairs (*49*). Figure S30-B shows that the coalescent compatibility test excludes more than half of all co-clustering female-female pairs if the compatibility threshold is at least 10%.

To summarize, we adopted the following exclusion criteria:

Exclusion criteria	Clade frequency in bootstrap viral phylogenies	Coalescent compatibility with direct transmission
Central	80%	20%
Lower-I	80%	30%
Lower-II	85%	20%
Upper-I	80%	10%
Upper-II	70%	20%

Viral phylogenetic analyses were remarkably successful in excluding potential transmission events. Across the above exclusion criteria, between 99.94%-99.96% of all potential transmission pairs with sequences available for both individuals could be excluded. Table S3 characterizes the phylogenetically probable transmitters. The difference between using a 7% threshold or no threshold at all was minimal: only 3 more recipients would have been excluded. Sensitivity analyses demonstrate that the findings reported in this study did not vary substantially across these exclusion criteria, and additional genetic distance criteria (figures S14-S22).

SOM 3 Procedure to quantify censoring bias

The observed, probable transmission intervals reported in figure 2 are subject to two main sources of bias. Below, we describe the technical bootstrap procedure to quantify the extent of censoring bias. The idea behind this procedure is described in the Materials and Methods of the main text and figure S6.

Bootstrap techniques proceed by constructing sub-samples from observed data to estimate properties of the observed data that is sampled from the population (*50*). Here, we implemented a bootstrap technique that sub-censors the observed data to estimate the extent of censoring of the observed data. Censoring describes the proportion of infected individuals that have not yet been registered in the ATHENA cohort, irrespective of whether a sequence was sampled or not. To quantify censoring, we considered all potential transmitters (stage A in figure 1) and their "overlap" intervals, during which the potential transmitters overlapped with infection windows of recipients. The probable transmitters and their transmission intervals do not enter the calculations below. We adopt the following notation:

t_E	end of the observation period
t_C	censoring time of potential transmitters
$[t_1, t_2]$	observation period of recipients
$t_C^* = t_C - \delta$	bootstrap censoring time, where $\delta > 0$
$[t_1^*, t_2^*]$	bootstrap observation period, where $t_1^* = t_1 - \delta$ and $t_2^* = t_2 - \delta$.

Here, we set t_E = 2013/03, the time of database closure; t_C= 2010/12, the end of the study period; and $[t_1, t_2]$ to one of the six time intervals 1996/07-2006/06, 2006/07-0207/12, 2008/01-2009/06,

2009/07-2009/12, 2010/01-2010/06, 2010/07-2010/12. The fourth period in table 4 was split into three intervals because of the rapidly increasing impact of censoring towards the present.

For a bootstrap censoring time t_C^*, we can calculate the proportion of non-censored intervals in infection/care stage x to recipients that are diagnosed during the bootstrap observation period $[t_1^*, t_2^*]$,

$$c_E(x, t_1^*, t_2^*, t_C^*) = \frac{\sum_{j \in R(t_1^*, t_2^*)} \sum_{\tau \in V_j(x)} \mathbf{1}\{\tau \; observed \; before \; t_C^*\}}{\sum_{j \in R(t_1^*, t_2^*)} \sum_{\tau \in V_j(x)} \mathbf{1}\{\tau \; observed \; before \; t_E\}},$$

where

$R(t_1^*, t_2^*)$ set of recipient MSM diagnosed in the period $[t_1^*, t_2^*]$,
$V_j(x)$ set of overlap intervals to recipient j that are in stage x.

If the corresponding potential transmitter is not diagnosed before t_C^*, then $\mathbf{1}\{\tau \; observed \; before \; t_C^*\}$ equals zero and otherwise one. This is illustrated in figure S6, where $t_1^* = 2006/06$, $t_2^* = 2007/12$, $t_E = 2013/03$ and t_C^* could be any time between 2008/01 and 2013/03.

We aim to estimate, for the actual censoring time t_C, the proportion of non-censored overlap intervals in stage x to recipients that are diagnosed during the period $[t_1, t_2]$. This can be written as

$$c_\infty(x, t_1, t_2, t_C) = \frac{\sum_{j \in R(t_1, t_2)} \sum_{\tau \in V_j(x)} \mathbf{1}\{\tau \; observed \; before \; t_C\}}{\sum_{j \in R(t_1, t_2)} \sum_{\tau \in V_j(x)} \mathbf{1}\{\tau \; observed \; before \; \infty\}}.$$

We need to assume that the censoring process has not changed within the last Δ^{max} years from t_2. In this case,

$$c_\infty(x, t_1, t_2, t_C) = c_\infty(x, t_1 - \delta, t_2 - \delta, t_C - \delta)$$

for all $\delta < \Delta^{max}$. We need to assume further that all overlap intervals have been observed by t_E. This is only the case when the bootstrap observation period lies sufficiently far back in time, that is $\delta > \Delta^{min}$. In this case,

$$c_\infty(x, t_1 - \delta, t_2 - \delta, t_C - \delta) = c_E(x, t_1 - \delta, t_2 - \delta, t_C - \delta)$$

for all $\delta > \Delta^{min}$. Under these assumptions on δ, the following bootstrap algorithm provides an estimate of the proportion of overlap transmission intervals that are not censored, $c_\infty(x, t_1, t_2, t_C)$.

Bootstrap algorithm
Let B be the number of bootstrap iterations.

For 1:B do

1. Draw δ_b from a uniform distribution with minimum Δ^{min} and maximum Δ^{max}.
2. Compute $\hat{c}_b = c_E(x, t_1 - \delta_b, t_2 - \delta, t_C - \delta_b)$.

Estimate $c_\infty(x, t_1, t_2, t_C)$ with $\hat{c} = \sum_{b=1}^{B} \hat{c}_b$.

We chose Δ^{min} and Δ^{max} as follows. Mathematical modelling indicates that the average time to diagnosis amongst MSM in the Netherlands is ~ 2-3 years in recent years (*7*, *46*). For some individuals, time to diagnosis may be substantially longer and we allowed for up to 4 years. In this case, any δ such that $t_2 - \delta \leq$ 2009/03 should be sufficiently large. Since the most recent t_2 is 2010/12, we have $\delta \geq \Delta^{min} = 2$ years. Further, we assumed that Δ^{max} can be set to 3 years.

Figure S31 shows that estimated censoring bias is extensive: for recipients diagnosed between 2010/07-2010/12, an estimated 20% of overlap intervals from potential transmitters estimated to be in chronic infection are observed. As expected from figure S6, the estimated censoring bias was substantially smaller for overlap intervals of potential transmitters in recent infection at time of diagnosis.

SOM 4 Modelling counterfactual prevention scenarios

We formulated prevention models that moved probable transmitters to less infectious infection/care stages, thereby changing the overall probability that any of the recipient MSM would have been infected to less than one. This section describes these individual-level prevention models and how they were parameterized.

SOM 4.1 Improved testing with conventional assays

Counterfactual testing scenarios re-allocated undiagnosed men to less infectious infection/care stages between diagnosis and ART start. The individual-level testing for prevention model has three parameters

θ_1^{Test}	duration between consecutive HIV tests in months
θ_2^{Test}	additional fraction of probable transmitters that are tested with frequency θ_1^{Test}
θ_3^{Test}	window period of HIV testing assay,

and proceeds as follows to simulate a counterfactual scenario.

A fraction θ_2^{Test} of randomly chosen, undiagnosed probable transmitters are tested in θ_1^{Test} intervals. The first test date was randomly allocated so that the average first test was in mid-2008. We assumed that the window period θ_3^{Test} of conventional assays is exactly 1 month (*51*). Before this window period, all tests were assumed to be negative. After this window period, all tests were assumed to correctly identify HIV status. After a counterfactual, positive test probable transmission intervals before diagnosis were randomly re-allocated to one of the stages between diagnosis and ART start. The re-allocation stage was drawn in proportion to the adjusted number of probable transmitters in that stage. Each re-allocated probable transmission interval was associated with a randomly chosen transmission probability from the new stage. Thus, the testing for prevention model changes the probability of secondary infections from undiagnosed men to a lower probability of secondary infections from diagnosed, untreated men.

To parameterize this model, we reviewed testing behaviour amongst uninfected MSM in the Netherlands, recipient MSM, and probable transmitters to the recipient MSM in this study. The duration between consecutive tests, θ_1^{Test}, was set to 12 months throughout. 38% of uninfected MSM

in the Netherlands reported to test annually in the EMIS 2010: The European Men-Who-Have-Sex-With-Men Internet Survey (*52*). Amongst MSM diagnosed between 2009/07-2010/12, 26.8% had a last negative test within 12 months prior to diagnosis. Amongst probable transmitters to recipient MSM diagnosed between 2009/07-2010/12, 17.3% had a last negative test within 12 months prior to diagnosis. Figure S32 shows that this low proportion of probable transmitters with a last negative test was not sensitive to the choice of infection time estimates or phylogenetic exclusion criteria. Figure 4 reports estimates of the proportion of transmissions that could have been averted by overall testing coverage γ_{target}^{Test}. Given a proportion $\gamma_{current}^{Test}$ of probable transmitters that are already testing annually, we determined θ_2^{Test} through the relationship

$$\gamma_{target}^{Test} = \gamma_{current}^{Test} + (1 - \gamma_{current}^{Test})\theta_2^{Test}.$$

Based on figure S32, we set $\gamma_{current}^{Test}$=0.17.

SOM 4.2 Improved testing with specialized assays that detect early infection before the presence of HIV antibodies

Counterfactual testing scenarios with specialized assays that can detect early infection before the presence of HIV antibodies, were simulated as the scenarios based on conventional assays, except that the window period was set to zero (*51*).

SOM 4.3 Antiretroviral pre-exposure prophylaxis

Counterfactual PrEP scenarios prevented randomly chosen, uninfected men from becoming infected. The individual-level PrEP prevention model has two parameters,

θ_1^{PrEP}	fraction of individuals that take PrEP
θ_2^{PrEP}	probability that an individual taking PrEP is not infected,

and proceeds as follows to simulate a counterfactual scenario.

A fraction θ_1^{PrEP} of recipients that test negative is randomly chosen to take PrEP by mid 2008. The intervention was assumed to be efficacious on a randomly chosen fraction θ_2^{PrEP} of those. This fraction was removed from the newly infected recipients (infection probability set from 1 to 0). In addition, a fraction θ_1^{PrEP} of probable transmitters was randomly chosen to take PrEP since they first tested negative. Test dates were simulated as in SOM 4.1. The intervention was also assumed to be efficacious on a randomly chosen fraction θ_2^{PrEP} of those. This fraction was removed from the infected probable transmitters (lowering infection probabilities of the corresponding recipients to below 1). The PrEP prevention model averts secondary infections amongst recipients as well as primary infections of probable transmitters that were uninfected at time of testing.

We parameterized θ_2^{PrEP} based on findings from the iPrEX, PROUD and ANRS Ipergay studies (*18, 19, 37*). The iPrEX trial demonstrated an overall reduction in HIV incidence of 44% (95% confidence interval 15-63%) of daily oral tenofovir-based PrEP amongst MSM from diverse settings (*18*). The PROUD study demonstrated a reduction in HIV incidence of 86% (58%-96%) of daily oral

single-pill PrEP amongst predominantly white, high risk MSM recruited from sexual health clinics in the United Kingdom (*19*). Reports from the ANRS Ipergay study indicate a reduction in HIV incidence of 86% (40%-99%) amongst MSM in France and Canada who follow an on demand dosing scheme 2-24 hours before sex (*37*). Reflecting the more recent PROUD and Ipergay trials, θ_2^{PrEP} was for a single simulated counterfactual scenario drawn from a Beta distribution with mean of 86% and 95% interquartile range 40%-99%. Uncertainty in this parameter is the main reason why confidence intervals associated with prevention strategies that include PrEP are larger than those without in figure 4. For the sensitivity analysis reported in figure S12, reflecting the iPrEX trial, θ_2^{PrEP} was for a single simulated counterfactual scenario drawn from a Beta distribution with mean of 44% and 95% interquantile range 20%-70%.

SOM 4.4 Treatment as prevention

Counterfactual treatment as prevention (TasP) scenarios re-allocated diagnosed, untreated men to less infectious infection/care stages after ART start. The individual-level TasP prevention model has one parameter

θ_1^{TasP} time to first viral suppression,

and proceeds as follows.

In case of immediate ART provision, all diagnosed but untreated probable transmitters started ART. Corresponding probable transmission intervals were randomly re-allocated to stages after ART start, with the exception of the intervals between diagnosis and time to first viral suppression θ_1^{TasP}. These intervals were always re-allocated to be 'before first viral suppression'. Each re-allocated probable transmission interval was associated with a randomly chosen transmission probability from the new stage. Thus, the TasP prevention model changes the probability of secondary infections from diagnosed, untreated men to a lower probability of secondary infections from treated men.

In case of ART provision when CD4 progress below 500 cells/ml, only the probable transmission intervals after diagnosis with CD4 progression to below 500 cells/ml were randomly re-allocated.

To parameterize this model, available Kaplan-Meier estimates of the percentage of patients with initial suppression to below 100 copies/ml were used (*7*). An estimated 50% of all patients diagnosed between 2007/01-2010/12 reached first viral suppression in 3.6 months, and θ_1^{TasP} was set to this value.

SOM 4.5 Combinations

Counterfactual combination prevention scenarios were evaluated through combination of the single intervention models. To evaluate test-and-treat prevention interventions, we first applied the testing for prevention model, followed by the treatment as prevention model. The PrEP prevention model was always linked to an HIV testing component. To evaluate PrEP in combination with test-and-treat interventions, we first applied the PrEP+test prevention model, followed by the treatment as prevention model.

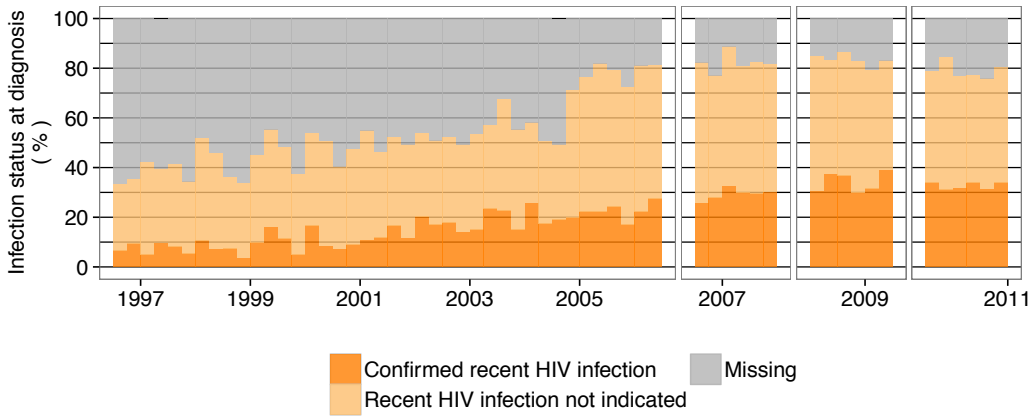


Figure S 1 Number of identified recipient MSM by 3-month intervals. MSM were confirmed to be in recent HIV infection at time of diagnosis if one of the following were reported: a last negative HIV-1 antibody test in the 12 months preceding diagnosis, an indeterminate HIV-1 western blot, or clinical diagnosis of acute infection. MSM with confirmed recent infection were considered as recipient in the viral phylogenetic transmission and prevention study. To evaluate trends over time, recipient MSM were stratified into four time periods as illustrated by the four blocks in the figure.

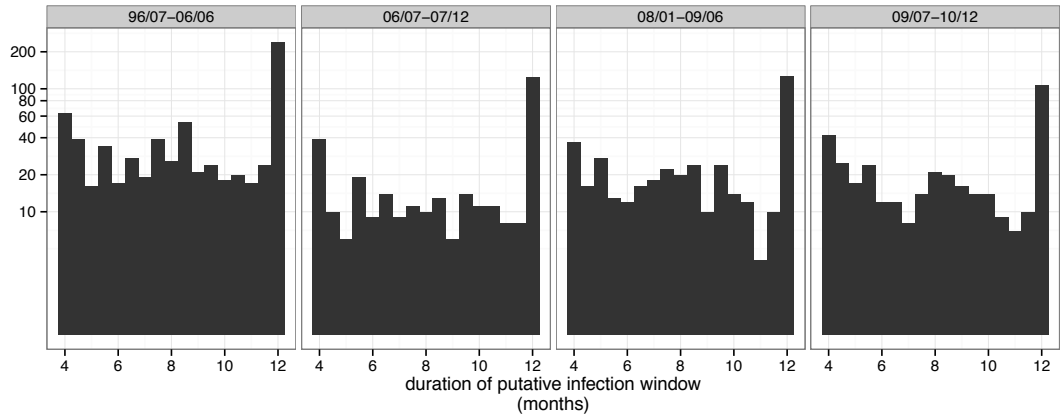


Figure S 2 Duration of infection windows of recipient MSM. Infection windows were at most 12 months long, reflecting the definition of recency of HIV infection. Where available, last negative HIV antibody tests were used to shorten infection windows. We assumed that the window period of HIV antibody tests is approximately 4 weeks, so that the last negative test had to be within 11 months preceding diagnosis in order to reduce the duration of the infection window.

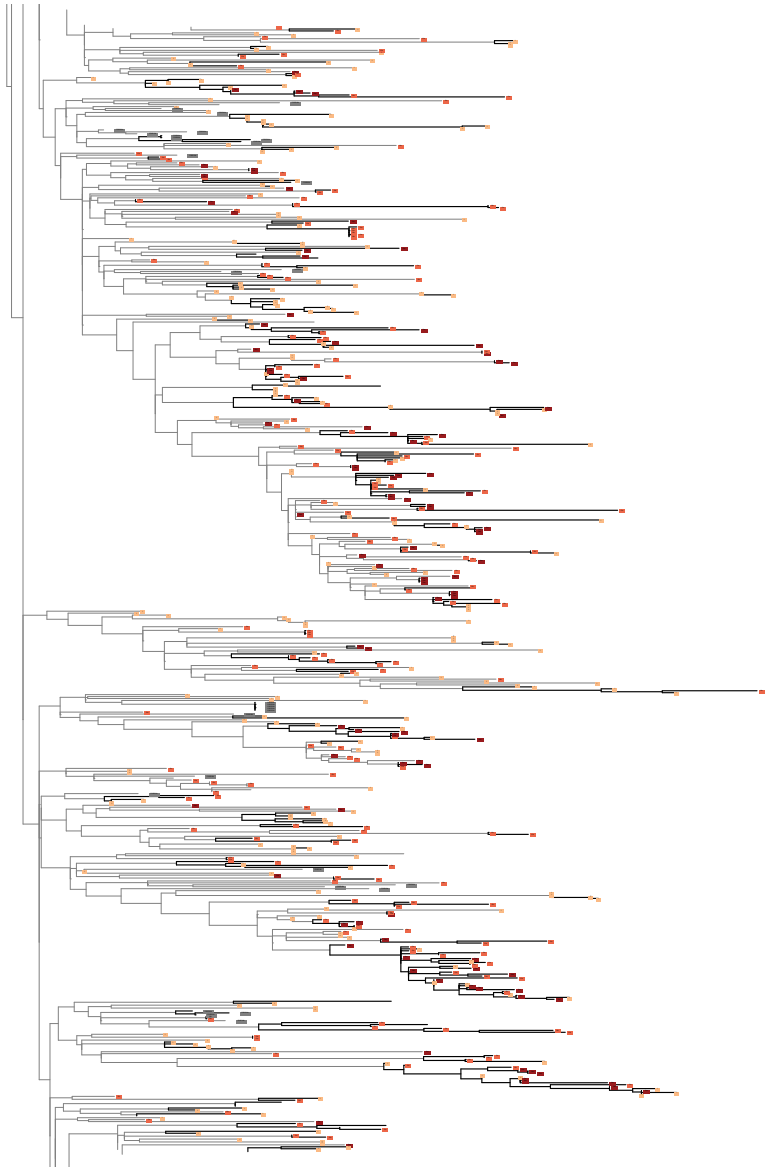


Figure S 3 Snapshot of the reconstructed viral phylogeny. Dutch sequences were enriched with subtype B sequences from the Los Alamos HIV sequence database because multiple subtype B lineages were likely imported into the Netherlands (*7*). Sequences were aligned with ClustalX v2.1 (http://www.clustal.org/clustal2/) using default parameters, and the alignment was manually curated. Primary drug resistance mutations listed in the IAS-USA March 2013 update were masked in each sequence. The viral phylogeny of the enriched ATHENA sequences was reconstructed under the GTR nucleotide substitution model with the ExaML maximum-likelihood method (*42*). Each clade in the viral phylogeny was annotated with the frequency with which it occurred among all bootstrap trees. Sequences from the Los Alamos sequence database are shown in grey. Sequences from men in recent infection at diagnosis are shown in dark red. Sequences from men for whom recent infection at diagnosis was not indicated are shown in orange. Sequences from men with unknown infection status at diagnosis are shown in yellow. Sub-clades that occurred in 400 out of 500 bootstrap trees are shown with thicker branches. Estimated branch lengths are in units of substitutions per site.

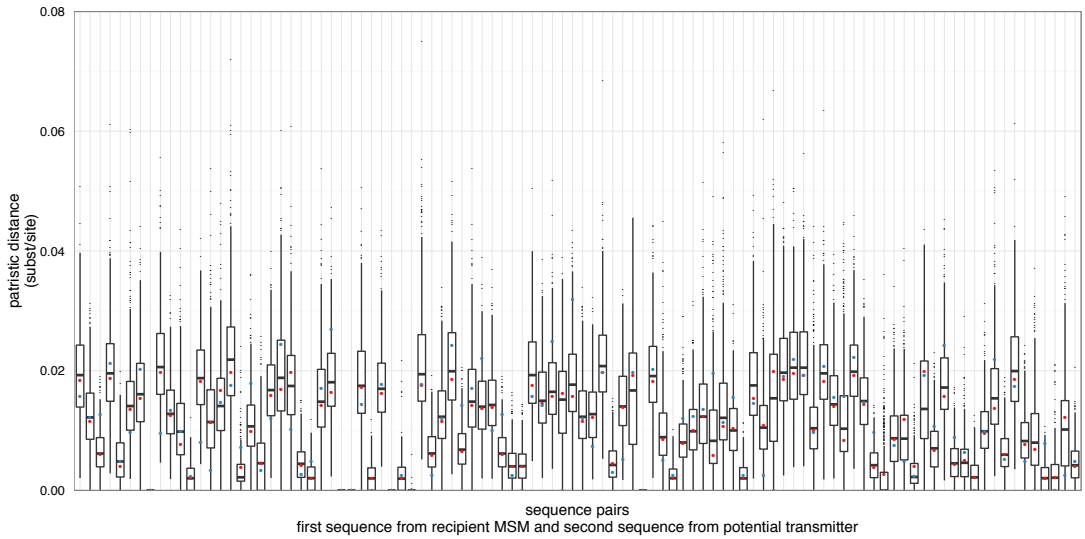


Figure S 4 Uncertainty in the estimated genetic distance between sequences from the transmitter and recipient of potential transmission pairs. For illustration purposes, 100 pairs with a median genetic distance below 2% were selected. Genetic distances (sum of the average number of nucleotide substitutions per site) were calculated from the reconstructed viral phylogeny on the sequence alignment (red dot), and from reconstructed viral phylogenies on bootstrap sequence alignments (boxplot, bar: median, box: interquartile range; whiskers: 95% quantile range). The genetic distance calculated on the tree with overall highest likelihood is shown as a blue dot. Uncertainty in genetic distances was accounted for in transmission analyses through bootstrap resampling.

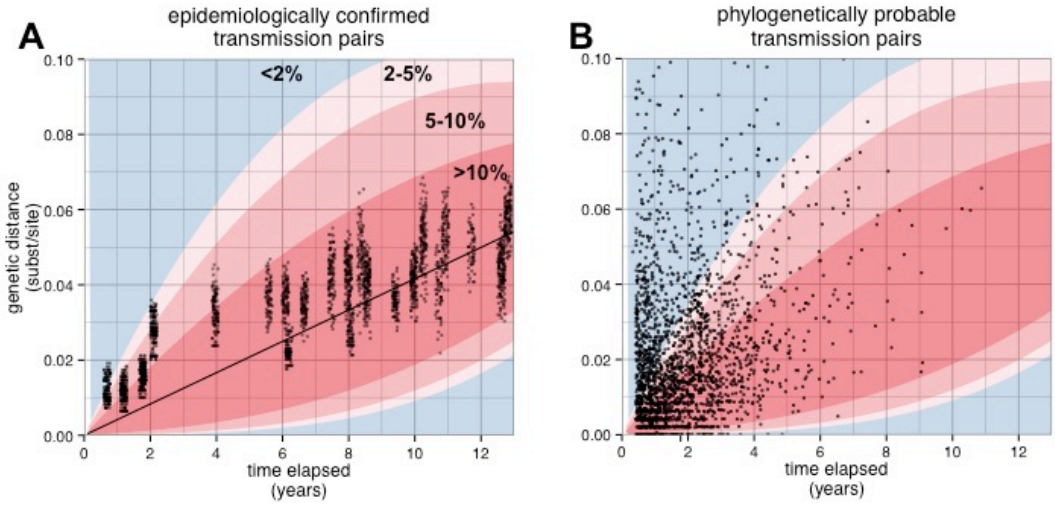


Figure S 5 Genetic distance between sequence pairs from previously published, epidemiologically confirmed transmitter-recipient pairs, and sequence pairs from the phylogenetically probable transmission pairs in this study. (A) Aligned sequences from the Belgium transmission chain were obtained from the authors (*13*). Drug-resistance sites were masked in each sequence. Patient A developed multi-drug resistance (*13*), and sequences from patient A were not considered. The viral phylogeny among all sequences was constructed with the maximum-likelihood methods (*42*) under the GTR nucleotide substitution model. Genetic distances between pairs of sequences from the confirmed transmitter and the confirmed recipient were calculated from the reconstructed viral phylogeny. Infection windows were determined through in-depth patient interviews, and made available by the authors (*13*). The time elapsed between sequences from a transmission pair was calculated as the time from the midpoint of the established infection window of the recipient to the sampling date of the transmitter, plus the time from the midpoint to the sampling date of the recipient. Genetic distances between confirmed pairs were strongly correlated with the time elapsed (Spearman correlation $\boldsymbol{\rho}$=0.84, n=2,807). We fitted the probabilistic molecular clock model

$$d_i \sim Gamma(\mu_i, \phi_i)$$
$$\mu_i = \beta t$$
$$\phi_i = \gamma_0 + \gamma_1 t,$$

where

d_i genetic distance between sequence pair i
μ_i mean of the Gamma distribution for the ith pair

ϕ_i dispersion of the Gamma distribution for the ith pair
β evolutionary rate
γ_k dispersion parameters

with regression techniques. The estimated model parameters were

$$\hat{\beta} = 0.00416, \quad \hat{\gamma}_0 = 1.008, \quad \hat{\gamma}_1 = -0.0523.$$

The fitted model explained 28% of the variance in the genetic distances between sequences from confirmed transmission pairs. Quantile ranges of the probabilistic molecular clock model are shown in red. (B) The fitted model was then applied to the 2,343 phylogenetically probable transmission pairs in this study to express the relative probability that a phylogenetically identified transmitter was the actual transmitter to a recipient. To reflect uncertainty in the genetic distance between probable transmission pairs, calculations were repeated on genetic distance values sampled from the distributions shown in figure S4. The time elapsed between sequences from phylogenetically probable pairs was calculated as the time from the midpoint of the infection window of the recipient to the sampling date of the transmitter, plus the time from the midpoint to the sampling date of the recipient. Transmission probabilities clearly varied between probable transmitters.

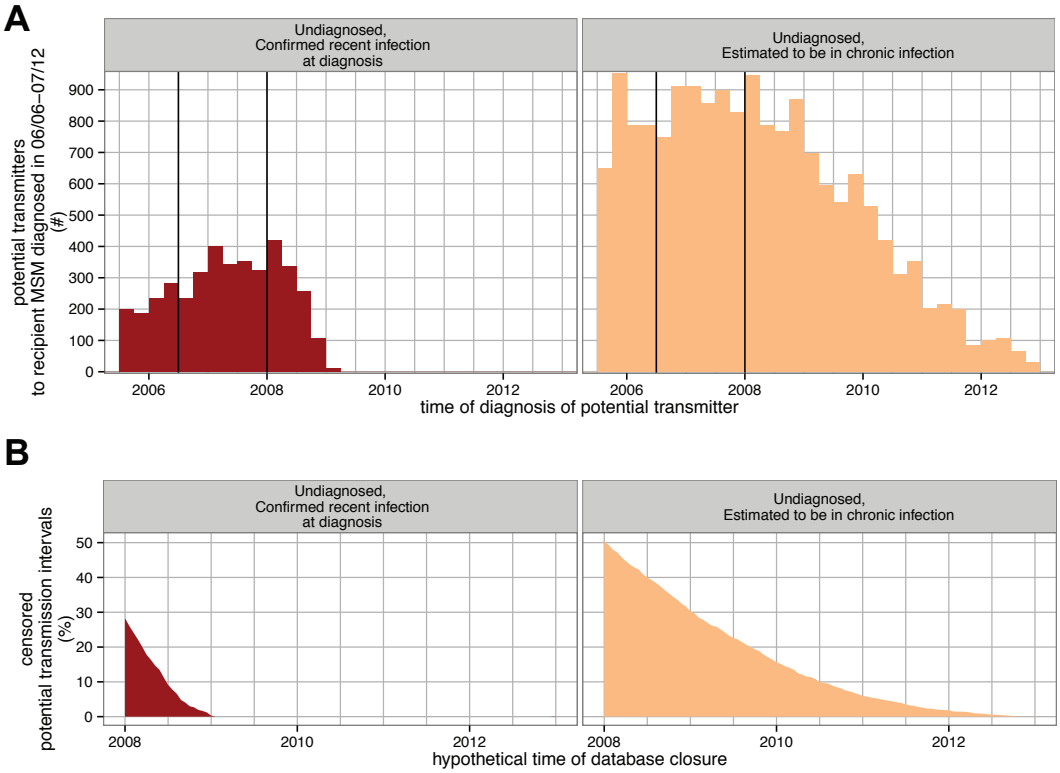
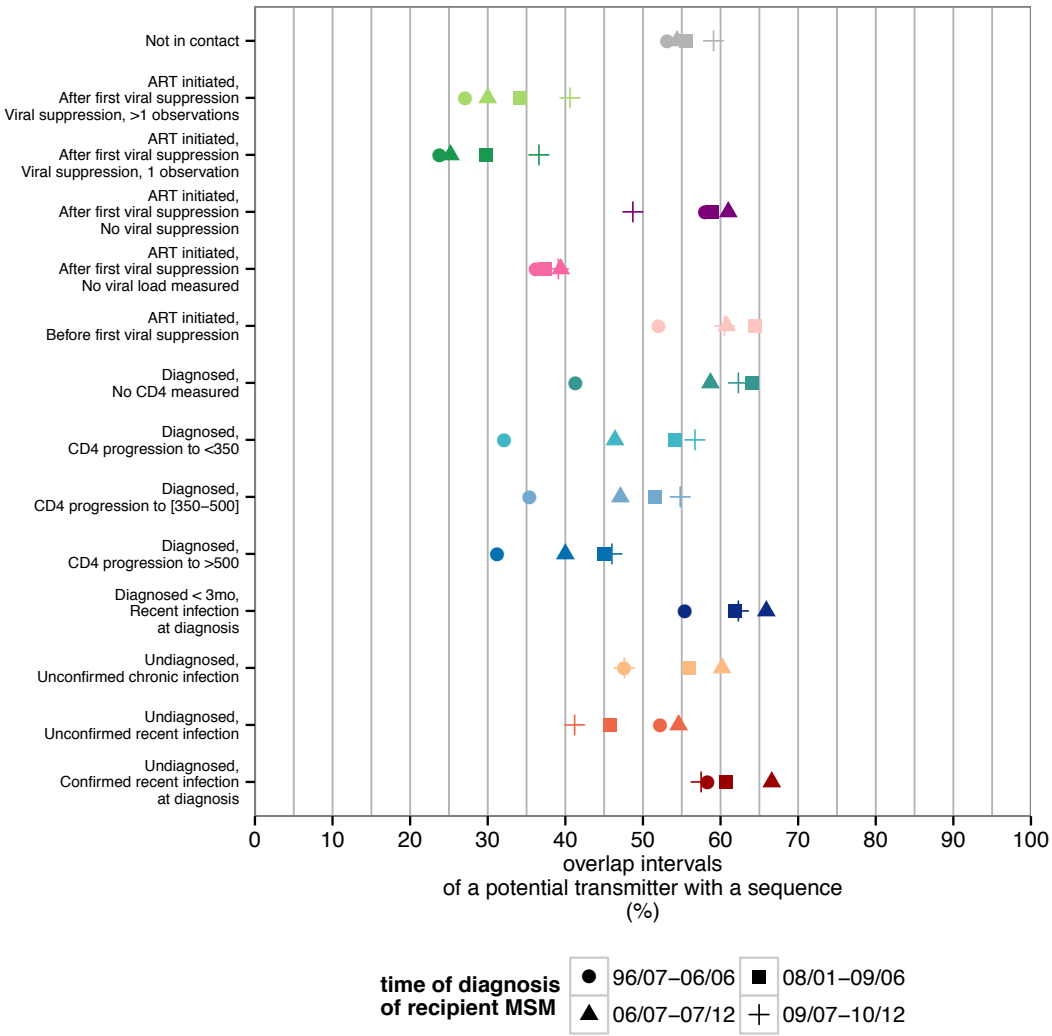


Figure S 6 Right censoring at past, hypothetical database closure times. (A) Distribution of time of diagnosis of potential transmitters to recipients that are diagnosed between $t_1^* = 2006/06$ to $t_2^* = 2007/12$. (Left) Histogram of the time of diagnosis of potential transmitters with confirmed recent infection at diagnosis. Infection windows of the recipients start the earliest in June 2005, and so do the putative transmission intervals between potential transmitters and their recipients ("overlap intervals"). Therefore, all potential transmitters with an overlap interval before diagnosis must be diagnosed after June 2005. This explains the abrupt start of the histogram after June 2005. (Right) Histogram of the time of diagnosis of potential transmitters estimated to be in chronic infection. Potential transmitters in undiagnosed, chronic infection at the putative transmission time may be diagnosed several years after their recipient. **(B)** Estimated proportion of censored overlap intervals at hypothetical database closure times after $t_2^* = 2007/12$. Considering a hypothetical closure time, say $t_C^* = 2008/12$, we considered potential transmitters with date of diagnosis after t_C^*. Next, we counted the overlap intervals of the hypothetically censored potential transmitters in each stage. Then we determined the proportion of these intervals among all intervals by stage. This proportion is plotted against hypothetical closure times, and quantifies the proportion of intervals that would have been censored, had the database been closed at the hypothetical closure time. A bootstrap algorithm described in the supplementary online material was used to extrapolate these estimates to the actual database closure time.



320 **Figure S 7 Sequence sampling probabilities by stage in the infection and care continuum.** To characterize sequence coverage by stage in the infection/care continuum, we considered potential transmitters with and without a sequence, and their "overlap" intervals during which they overlapped with infection windows of recipients. Then, the proportion of overlap intervals whose potential transmitter had a viral sequence sampled was calculated, and plotted by stage and time of diagnosis of the recipient. Colour codes are as in figure 2 in the main text. Typically, sampling probabilities increased with calendar time. Reflecting preferential sequencing for drug resistance testing, intervals with viral load measurements below 100 copies/ml were sampled least frequently, while those above 100 copies/ml were sampled twice as often. Intervals with a lower CD4 count were more likely to be sampled than those with a higher CD4 count. Intervals of transmitters in confirmed recent infection at diagnosis were also more likely to be sampled than those without, reflecting participation of the former in sub-studies of the ATHENA cohort (*7*).

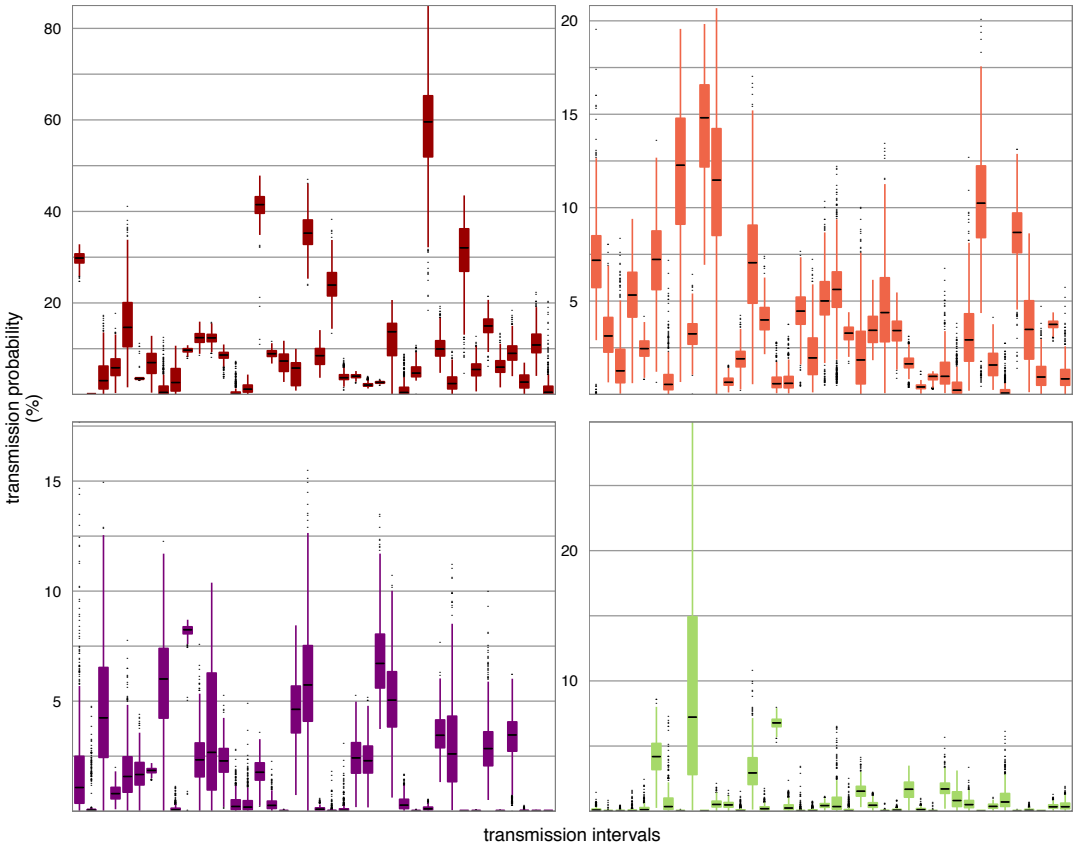


Figure S 8 Invidividual-level variation in phylogenetically derived transmission probabilities by infection/care stages. Transmission probabilities for observed transmission intervals τ were calculated as described in Materials and Methods, and are shown for a random sample of 40 observed transmission intervals for four infection/care stages. Colour codes match those in figure 2 in the main text. Uncertainty in the estimated phylogenetic transmission probabilities is indicated with boxplots (black bar: median, box: 50% interquartile range, whiskers: 95% interquartile range). Substantial individual-level variation in transmission probabilities indicates that a relatively large number of past transmission events are needed in order to reliably quantify sources of transmission.

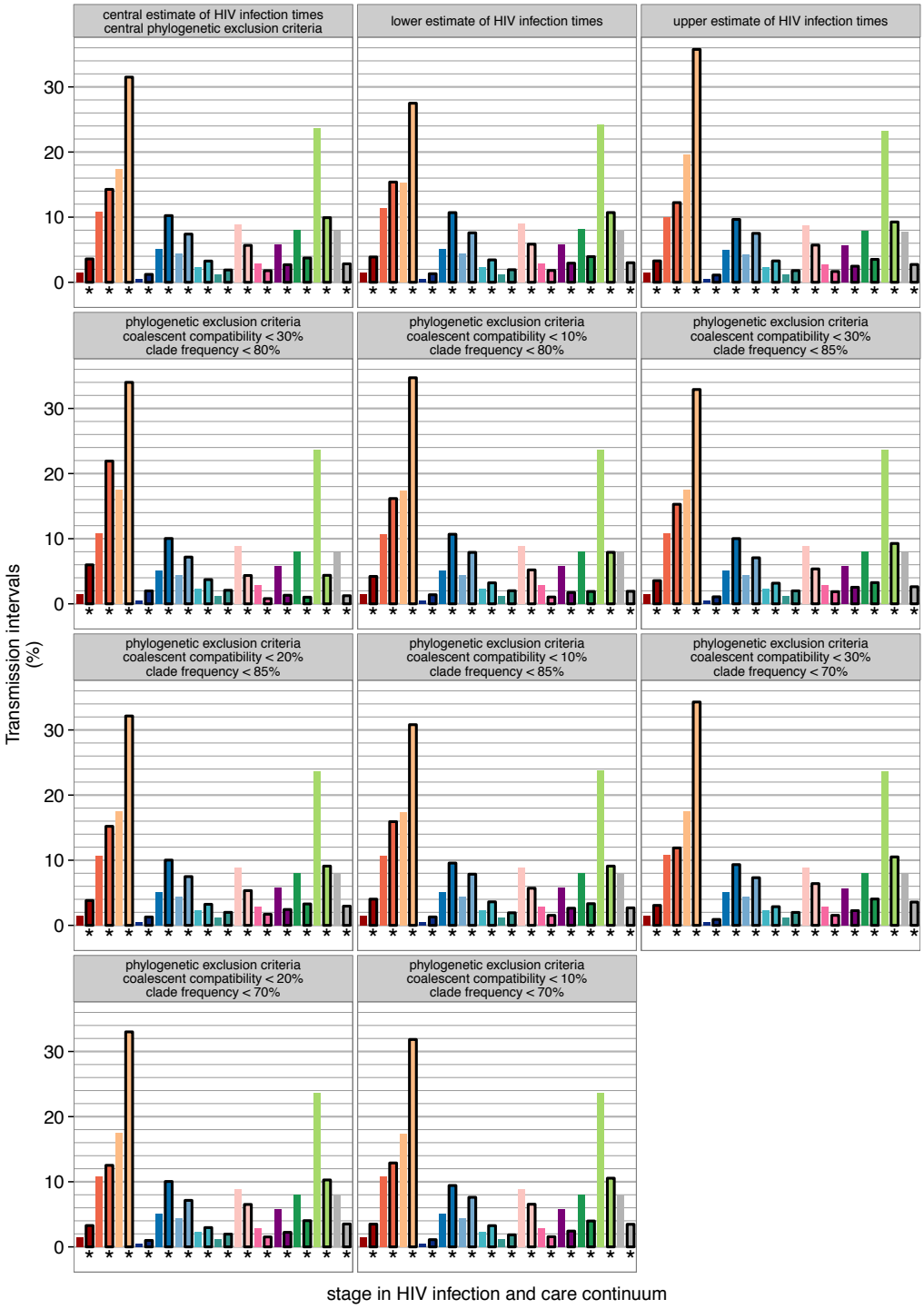


Figure S 9 Frequency of infection/care stages among phylogenetically probable transmitters. Phylogenetic exclusion criteria and infection time estimates were varied as described in the panels. Colour codes are as in figure 2 in the main text. Overall, infection/care stages before ART start were overrepresented amongst phylogenetically probable transmitters (marked by an asterix), while all stages after ART start were underrepresented amongst phylogenetically probable transmitters (marked by an asterix). Periods with no contact for at least 18 months to HIV care services were also underrepresented amongst phylogenetically probable transmitters, likely reflecting that a large proportion of potential transmitters that are listed in the ATHENA cohort but had no contact for 18 months moved abroad or died.

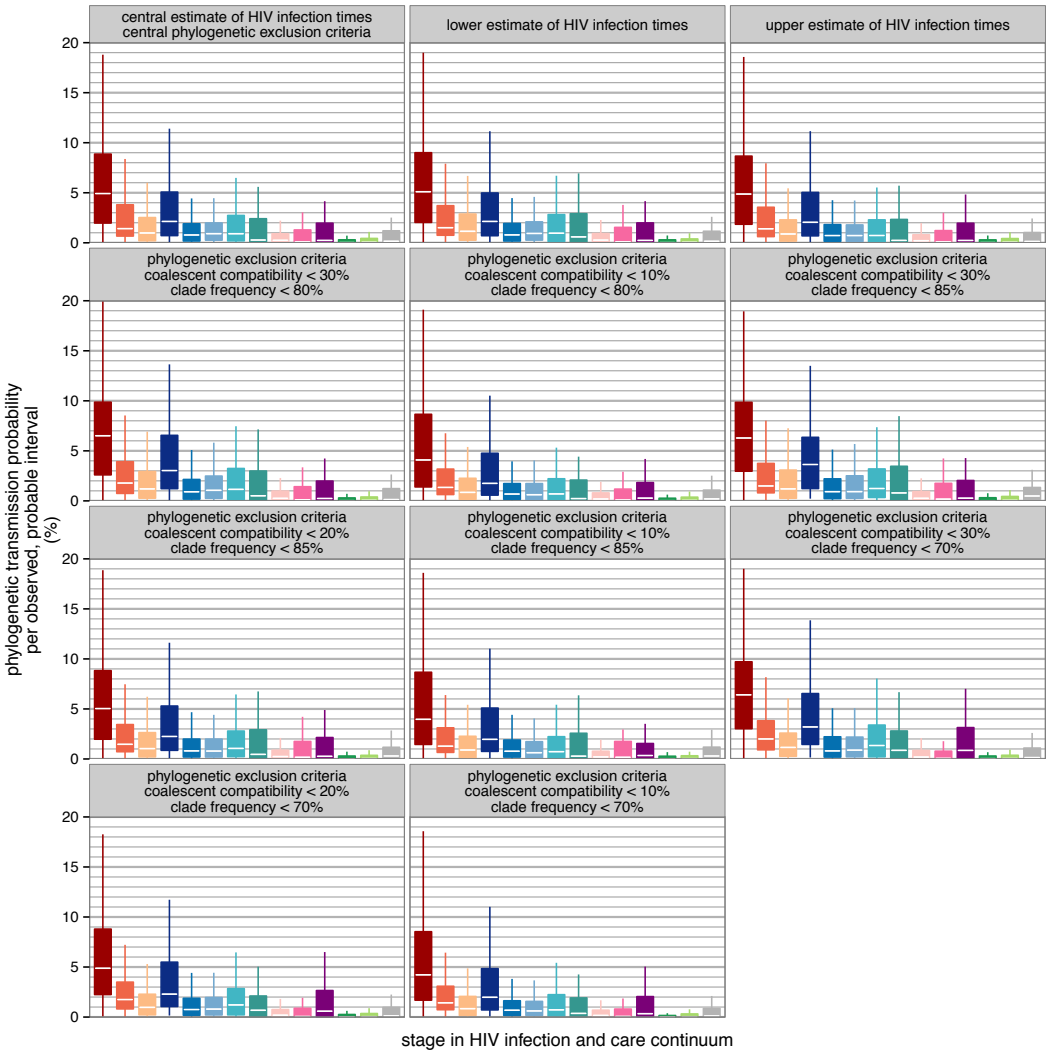


Figure S 10 Phylogenetically derived transmission probabilities of observed transmission intervals. Phylogenetic exclusion criteria and infection time estimates were varied as described in the panels. Colour codes are as in figure 2 in the main text. Overall, transmission probabilities were small, with a mean of 2.1% (25% quantile: 0.9%, 75% quantile: 2.5%, 97.5% quantile: 11.6%). However, when grouped by infection/care stage, the phylogenetic transmission probabilities were highly informative as to how transmission rates change with progression through the infection and care continuum.

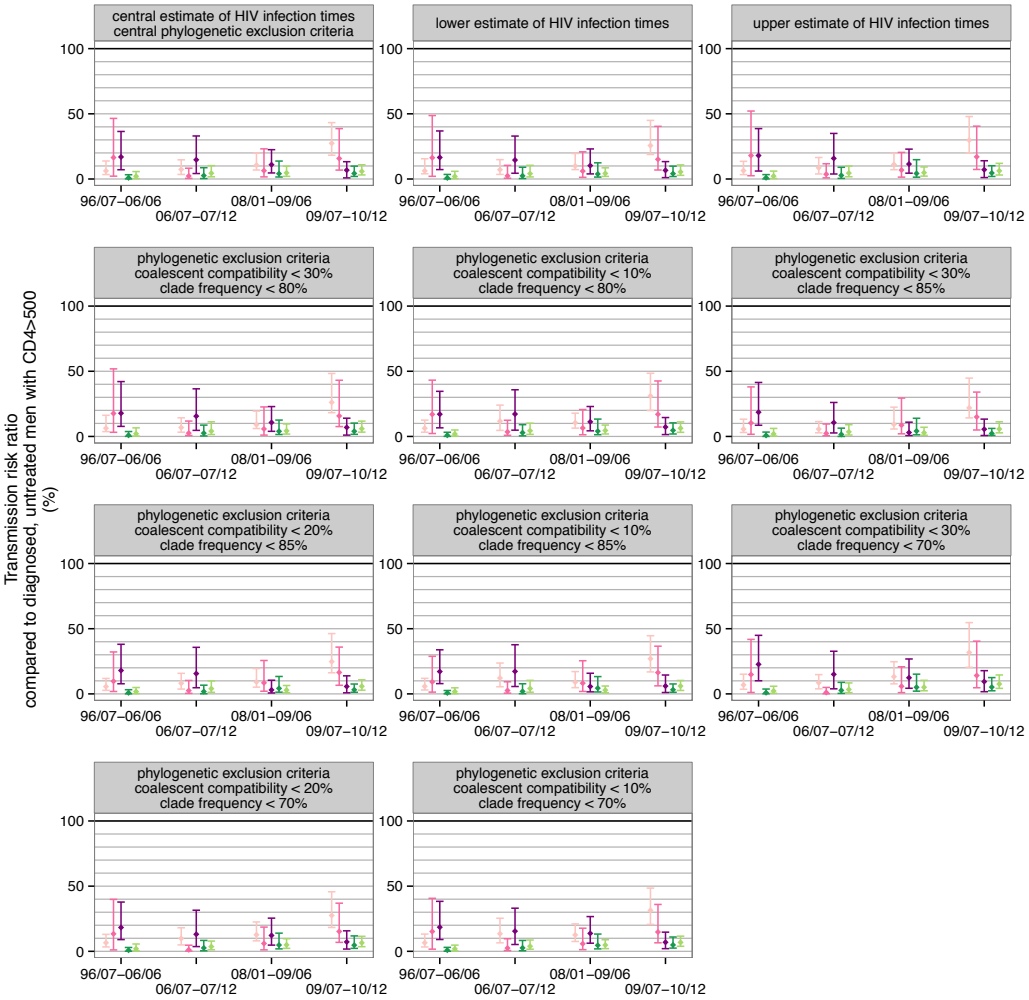


Figure S 11 Transmission risk ratio from men after ART start, compared to diagnosed untreated men with CD4 > 500 cells/ml. Colour codes are as in figure 2 in the main text.

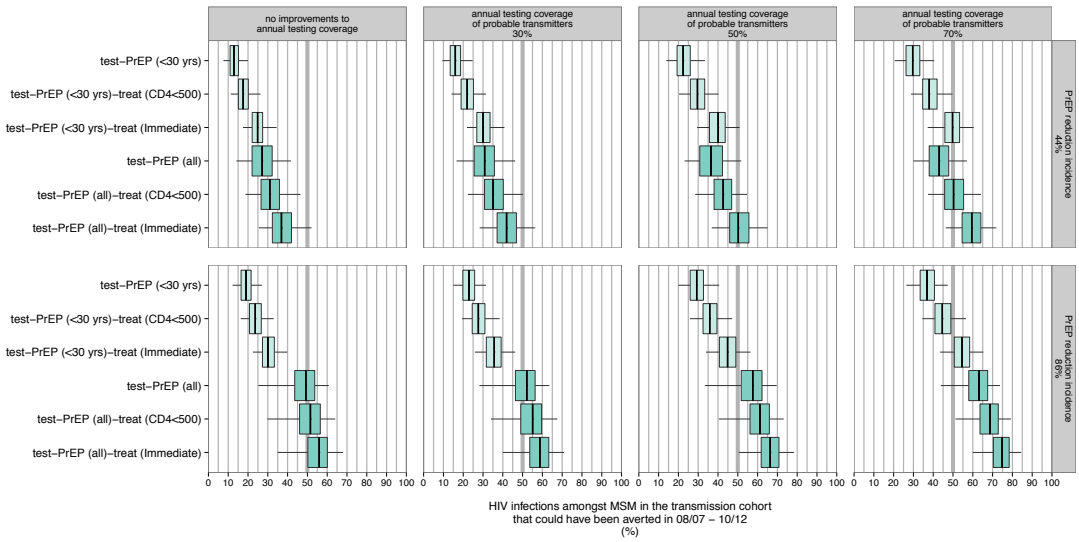


Figure S 12 Sensitivity analysis on the impact of PrEP with lower efficacy. Estimated proportion of infections between mid 2008 to 2011 that could have been averted through the listed interventions, assuming an 44% efficacy of PrEP as reported in the iPrEX trial, and an 86% efficacy of PrEP as reported in the more recent Ipergay and PROUD trials.

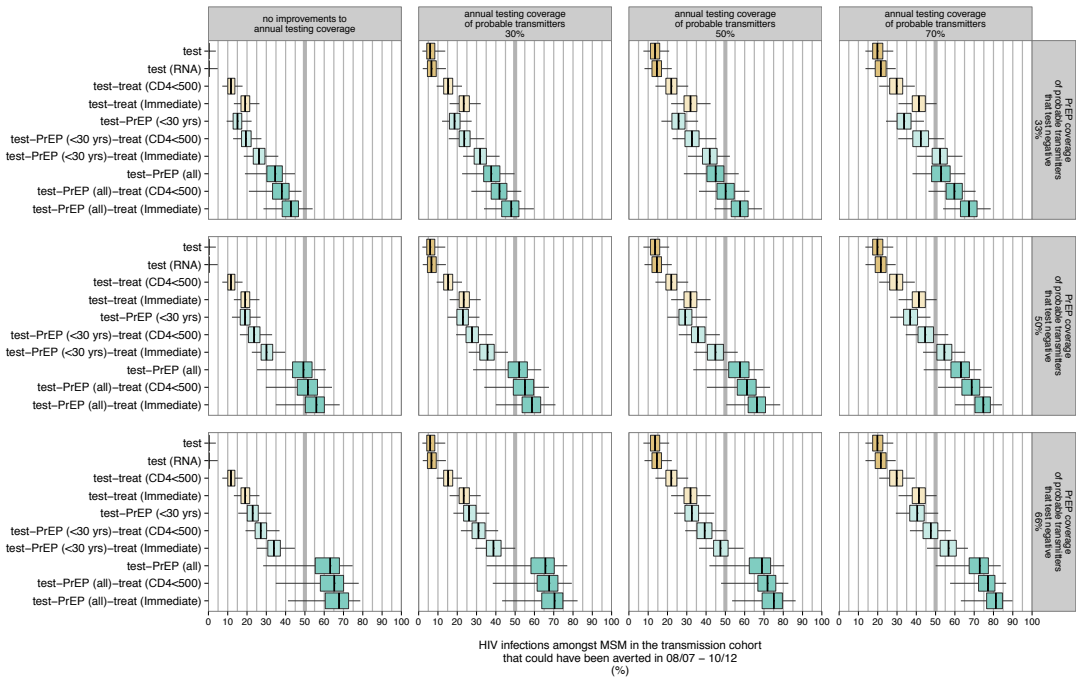


Figure S 13 Sensitivity analysis on the impact of lower or higher PrEP coverage.

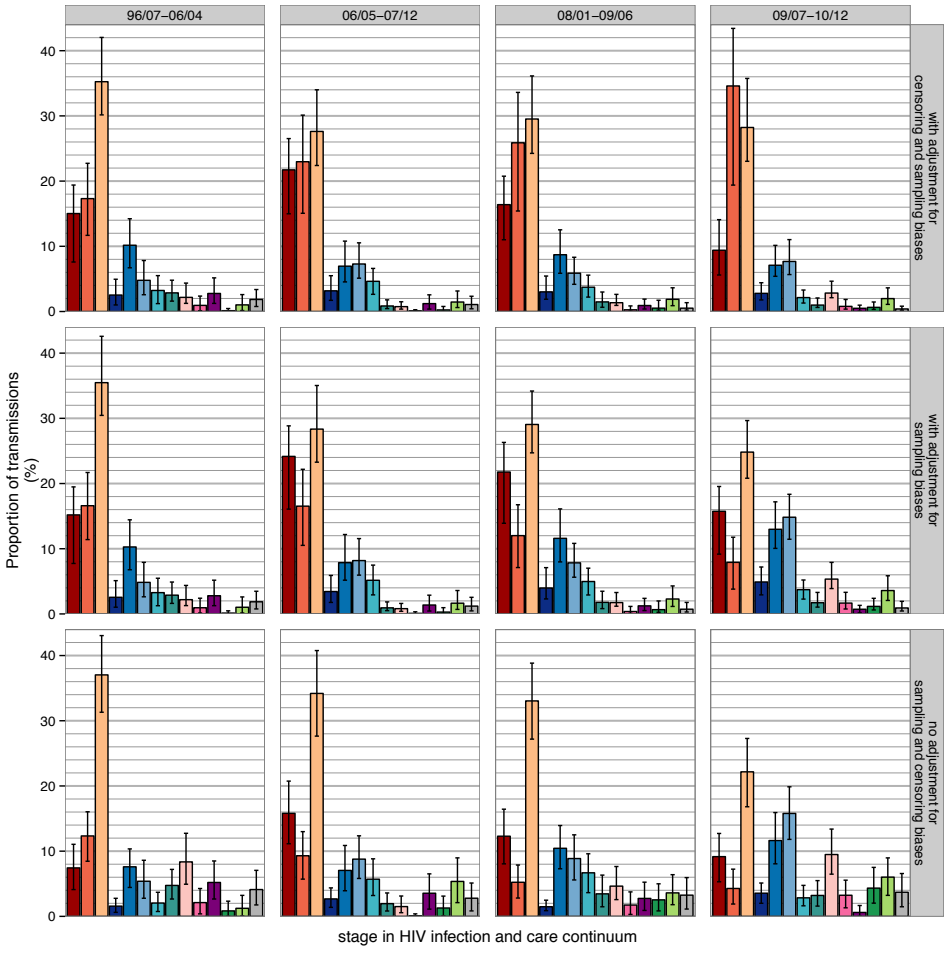


Figure S 14 Impact of sampling and censoring adjustments on the estimated proportion of transmissions from stages in the infection and care continuum. The proportion of transmissions attributable to infection/care stages was calculated as described in the Materials and Methods (top row), with equal sequence sampling probabilities in the missing data model (middle row), and with $m_j(z)$ =0 (bottom row, see Materials and Methods). Colour codes are as in figure 2 in the main text. With no corrections to censoring and sampling bias, the proportion of transmissions attributable to undiagnosed men declines to less than 40% by 2009/07-2010/12. With no

corrections to censoring bias but corrections for sequence sampling bias, the proportions of stages with large per capita transmission probabilities increase most. Those stages with a large number of missing intervals are not necessarily amplified most, because each of these intervals may be associated with a relatively small transmission probability. In particular, the estimated proportion of transmissions from undiagnosed men in recent infection at diagnosis is large, even though the total number of added, missing intervals in this stage is comparatively small. This comparison confirms, first, that sequence sampling and censoring bias can have an extensive impact on population based phylogenetic analyses. Second, this comparison demonstrates that it is not intuitive how corrections to sequence sampling and censoring biases impact on such analyses when relative transmission rates also vary across risk groups.

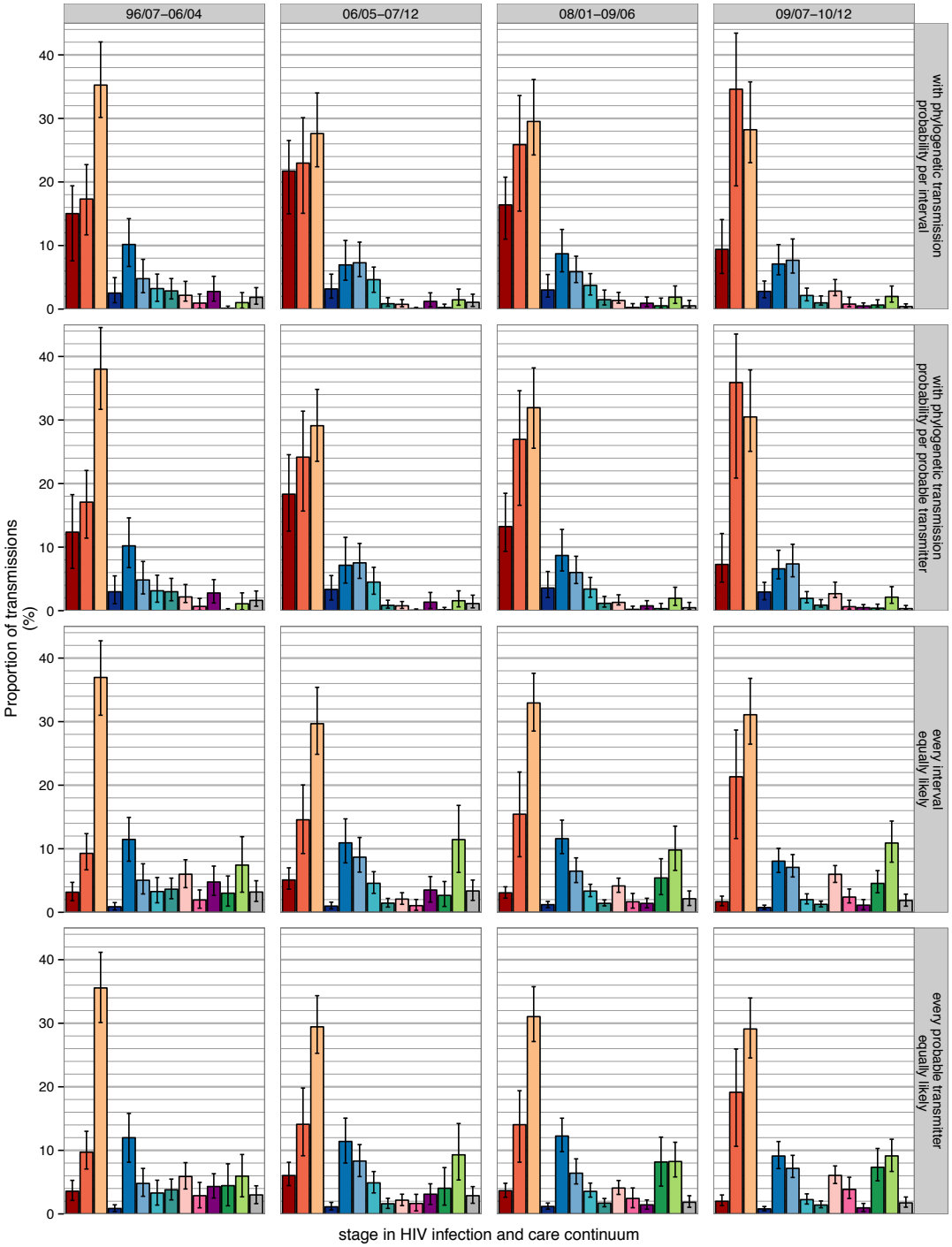


Figure S 15 Impact of phylogenetic transmission probabilities on the estimated proportion of transmissions from stages in the infection and care continuum. The proportion of transmissions attributable to infection/care stages was calculated as described in the Materials and Methods (top row), without adjusting for differences in the number of intervals per pair ($\omega_{ij\tau} = \varphi_{ij}\omega_{ij}$, second row), with transmission from every probable transmitter equally likely ($\omega_{ij\tau} = 1/\tau_{ij}$, third row), and transmission from every interval equally likely ($\omega_{ij\tau} = 1$, bottom row). Colour codes are as in figure 2 in the main text. Setting $\omega_{ij\tau} = \varphi_{ij}\omega_{ij}$ (second row) had no substantial impact on the estimated proportions. In the last two cases, the proportion of transmissions from undiagnosed men in recent

infection at time of diagnosis is very small, because the high infectiousness during this short stage in the infection and care continuum is ignored. In addition, the proportion of transmissions from men after ART start is much higher because the low infectiousness during these stages in the infection and care continuum is ignored. Ignoring differential transmission probabilities among probable transmitters may complicate interpretation of viral phylogenetic cluster association studies.

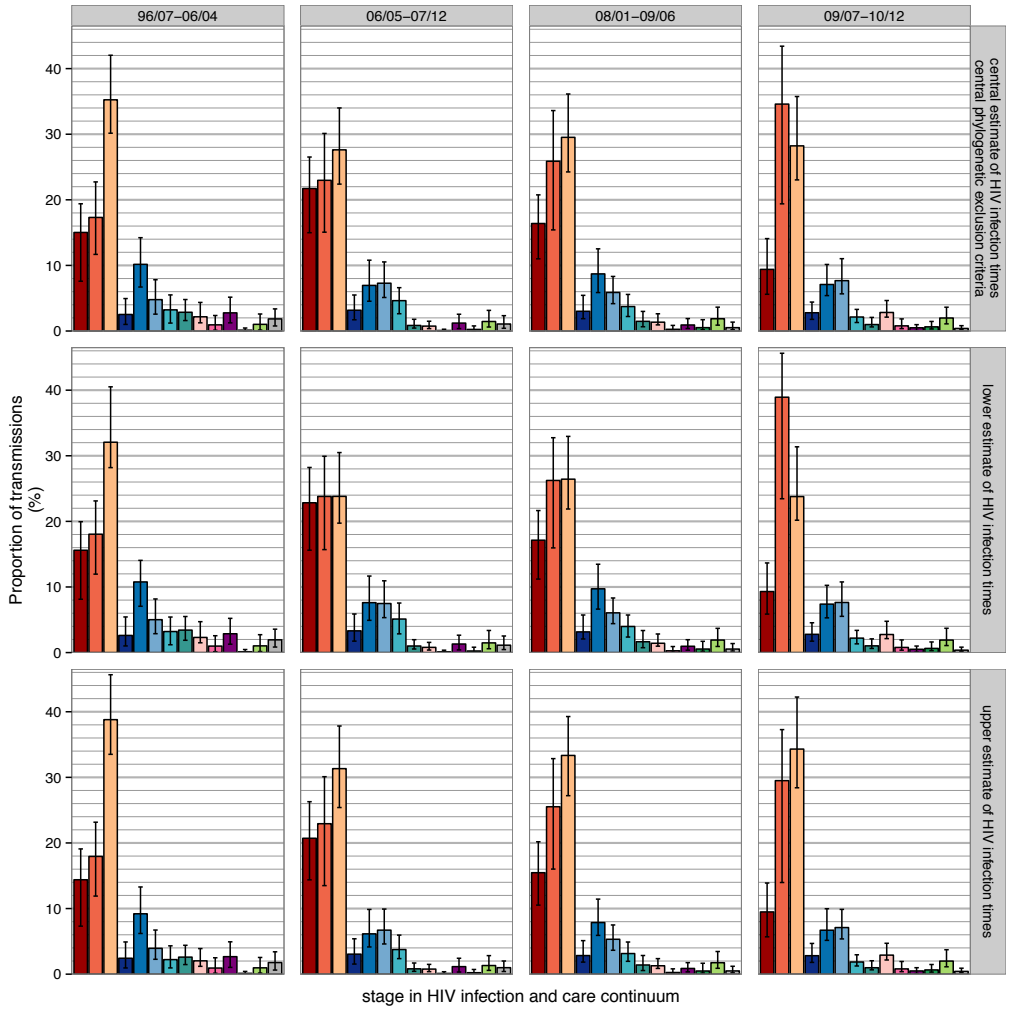


Figure S 16 Impact of infection time estimates on the estimated proportion of transmissions from stages in the infection and care continuum. The proportion of transmissions attributable to infection/care stages was calculated as described in the Materials and Methods (top row), based on the potential transmitters identified with lower 95% estimates of HIV infection times in table S2 (middle row), and upper 95% estimates of HIV infection times (bottom row). Colour codes are as in figure 2 in the main text. The estimated proportions did not vary substantially.

390

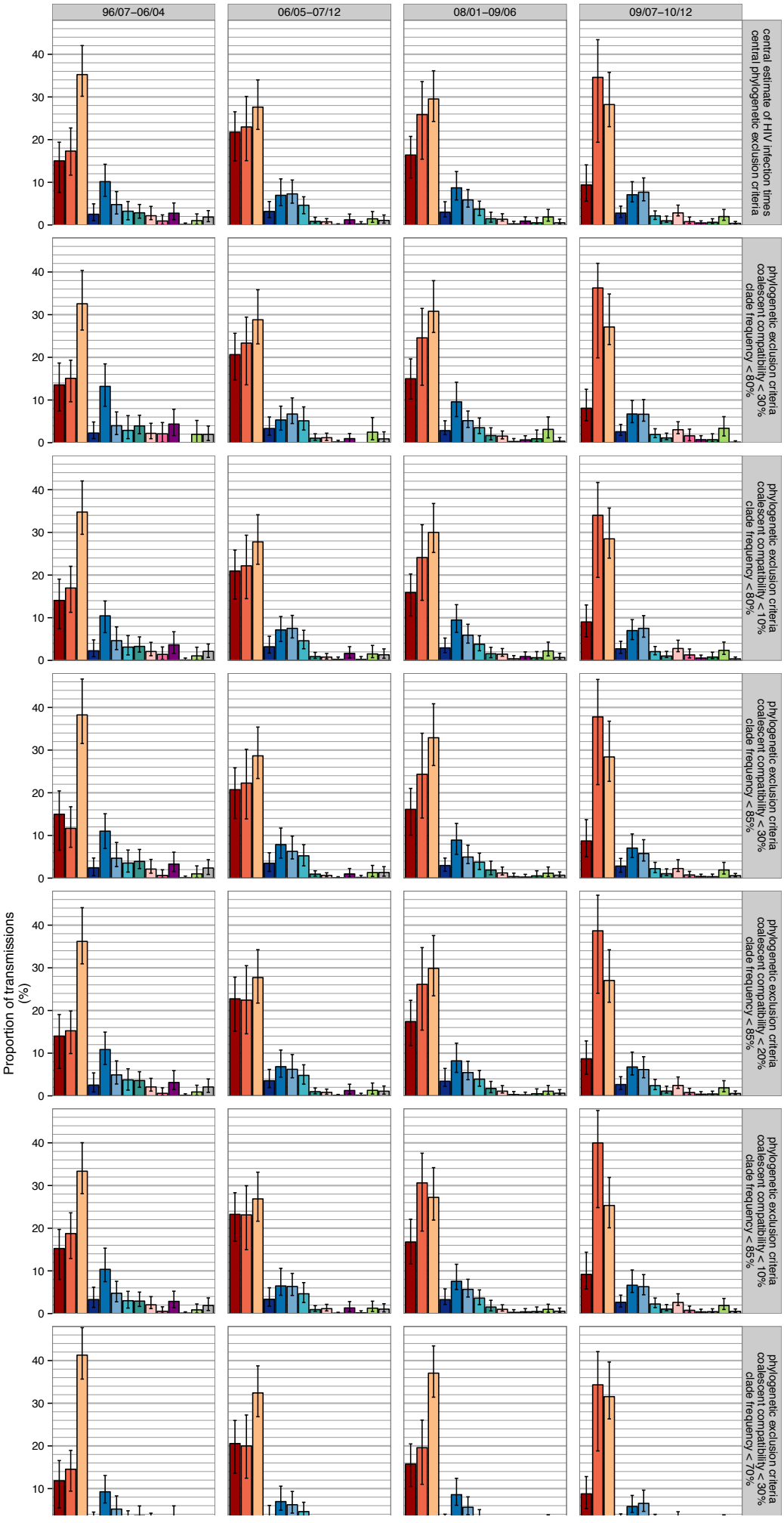

96/07–06/04
06/05–07/12
08/01–09/06
09/07–10/12
central estimate of HIV infection times
central phylogenetic exclusion criteria
phylogenetic exclusion criteria
coalescent compatibility < 30%
clade frequency < 80%
phylogenetic exclusion criteria
coalescent compatibility < 10%
clade frequency < 80%
phylogenetic exclusion criteria
coalescent compatibility < 30%
clade frequency < 85%
phylogenetic exclusion criteria
coalescent compatibility < 20%
clade frequency < 85%
phylogenetic exclusion criteria
coalescent compatibility < 10%
clade frequency < 85%
phylogenetic exclusion criteria
coalescent compatibility < 30%
clade frequency < 70%
Proportion of transmissions (%)
0
10
20
30
40

Figure S 17 Impact of phylogenetic clustering criteria on the estimated proportion of transmissions from stages in the infection and care continuum. The proportion of transmissions attributable to infection/care stages was calculated as described in the Materials and Methods (top row), and then using alternative upper and lower phylogenetic exclusion criteria as described in the row panels. Colour codes are as in figure 2 in the main text. The estimated proportions did not vary substantially.

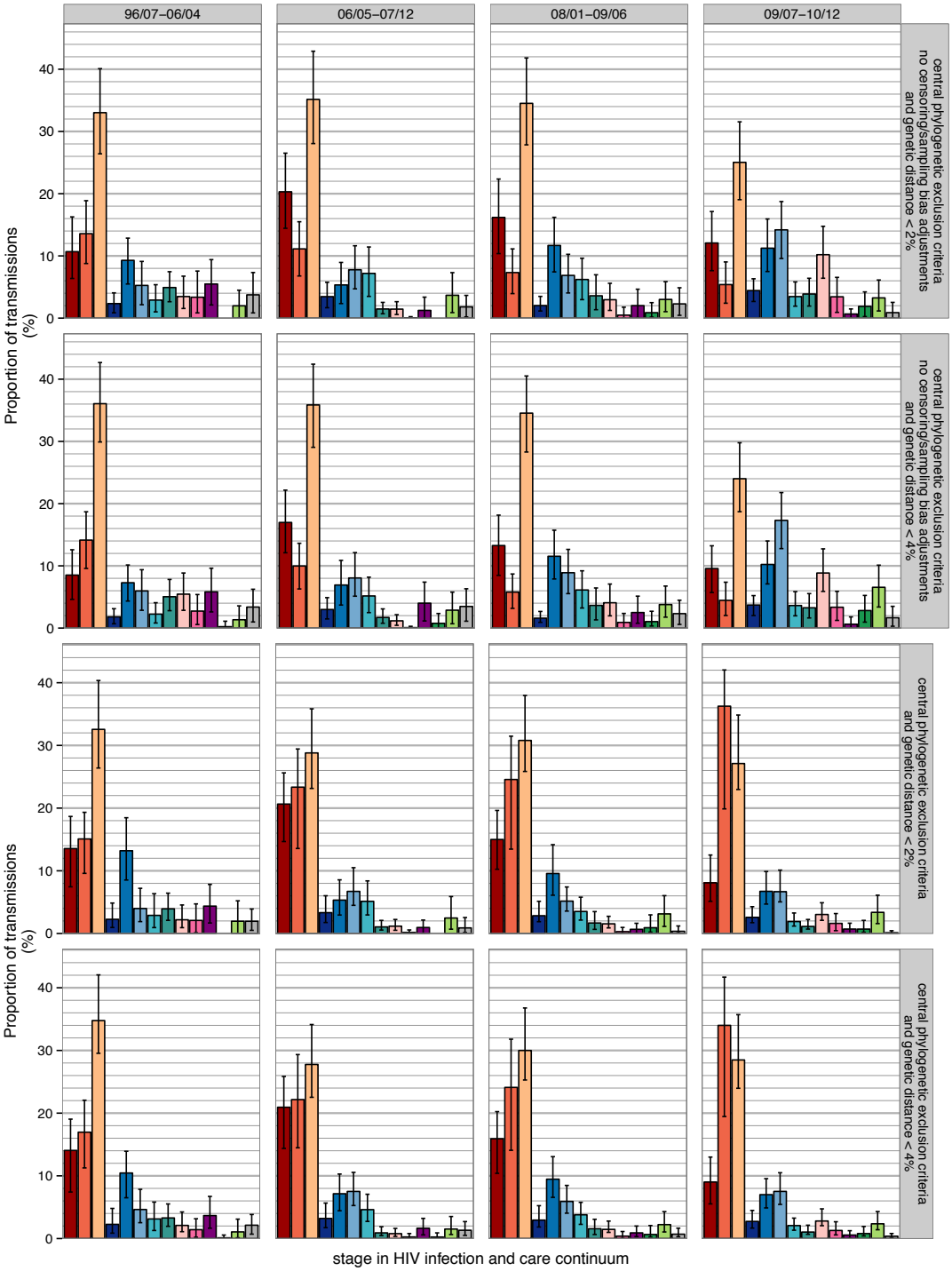


Figure S 18 Impact of additional genetic distance criteria on the estimated proportion of transmissions from stages in the infection and care continuum. The proportion of transmissions attributable to infection/care stages was calculated as described in the Materials and Methods, but potential transmitters were also excluded with additional genetic distance criteria described in the row panels. Colour codes are as in figure 2 in the main text. Before censoring and sampling biases were adjusted, an additional 2% genetic distance criterion lead to a slight increase in the proportion of transmissions attributable to men in their first year of infection, while the additional 4% genetic distance criterion lead to estimates that are comparable to those obtained without the genetic distance criterion. After censoring and sampling biases are adjusted, the estimated proportions did not differ substantially from those obtained without an additional genetic distance criterion.

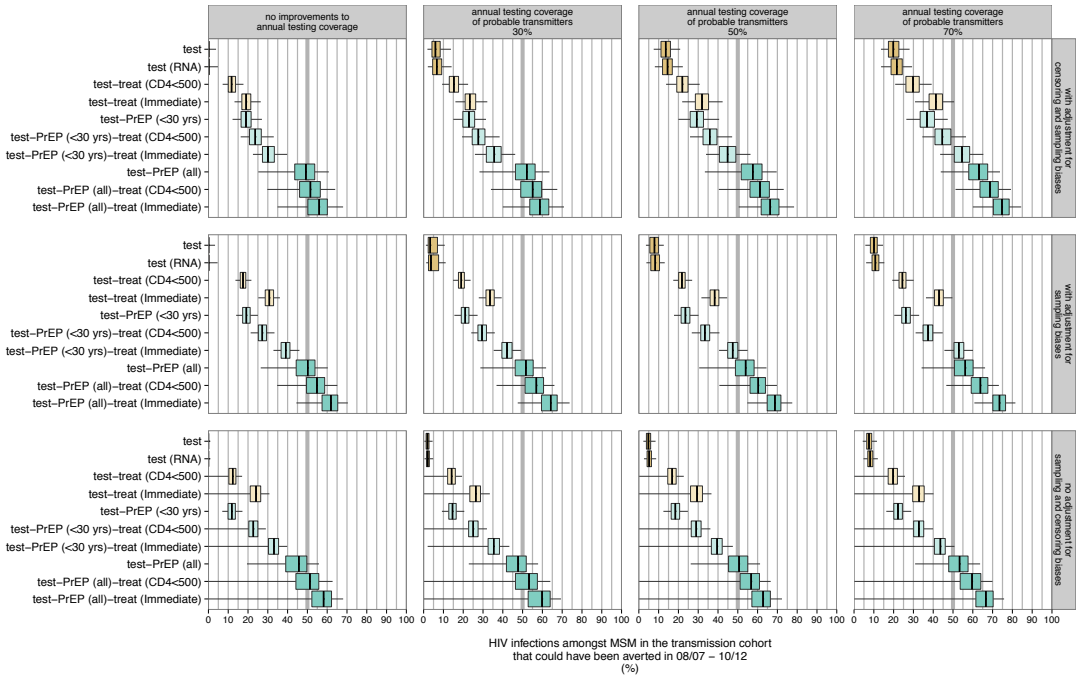


Figure S 19 Impact of sequence sampling and censoring adjustments on the estimated proportion of averted infections. The proportion of transmissions that could have been averted was calculated as described in the Materials and Methods (top row), with equal sequence sampling probabilities in the missing data model (middle row), and with $\boldsymbol{m_j(z)}$ =0 (bottom row). With no corrections to censoring and sampling bias, the estimated proportion of undiagnosed men is less than 40% in 2008/07-2010/12. Correspondingly, the estimated impact of test-and-treat is higher when compared to the central estimate. However, even under this extreme case of model misspecification, interventions including test-and-PrEP are associated with the largest reductions in HIV incidence. The estimated $\boldsymbol{a(H)}$ differ from the central estimate by at most 10%. The case where sampling bias is adjusted but censoring bias is ignored is overall similar to the central estimates. This comparison indicates that the evaluation of the short-term impact of prevention strategies is robust to extensive differences in how sequence sampling and right censoring biases are adjusted for.

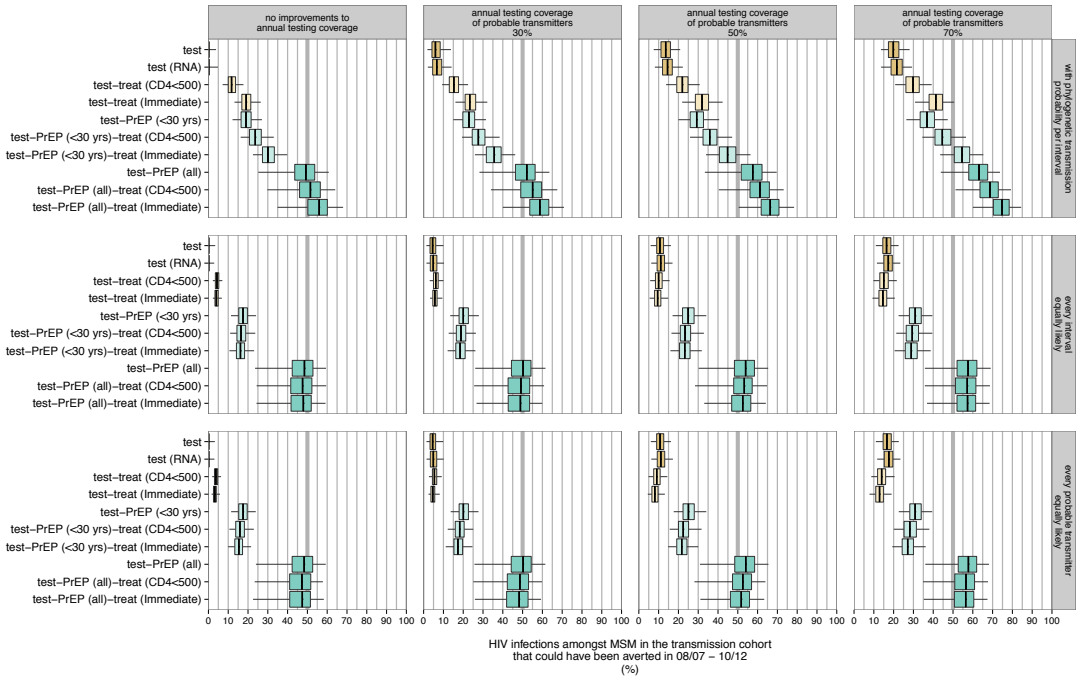
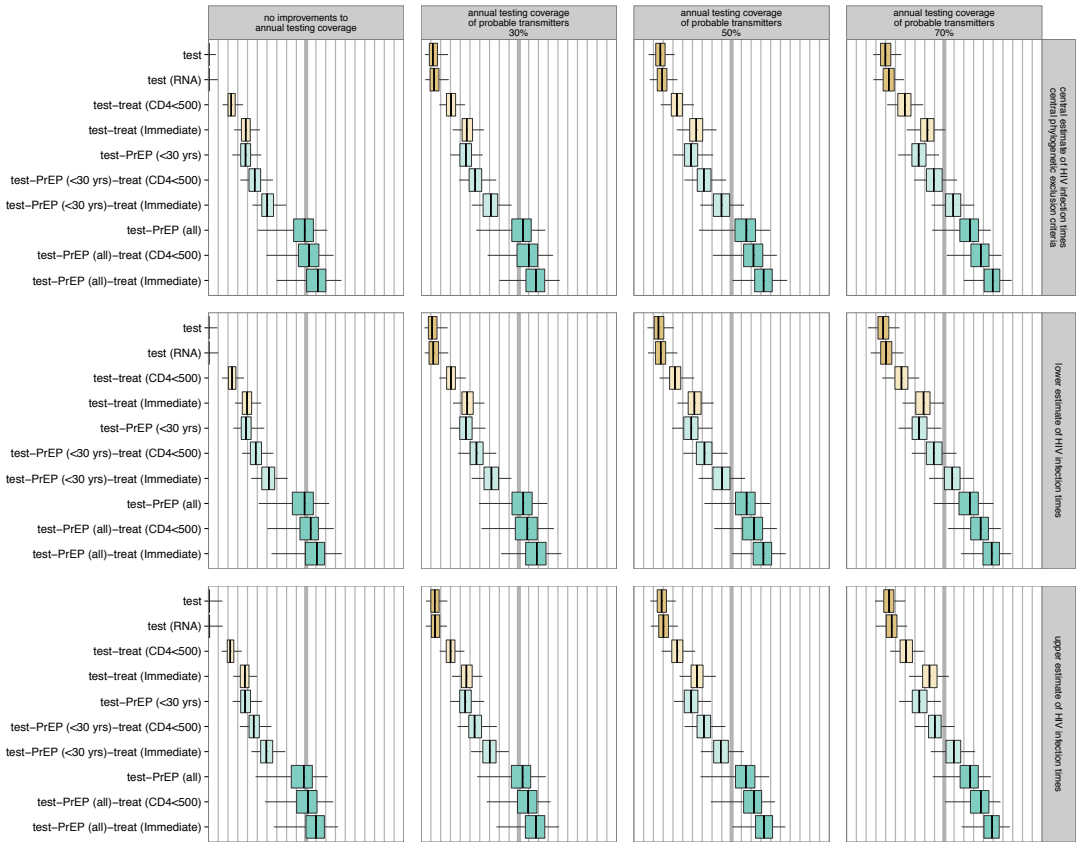
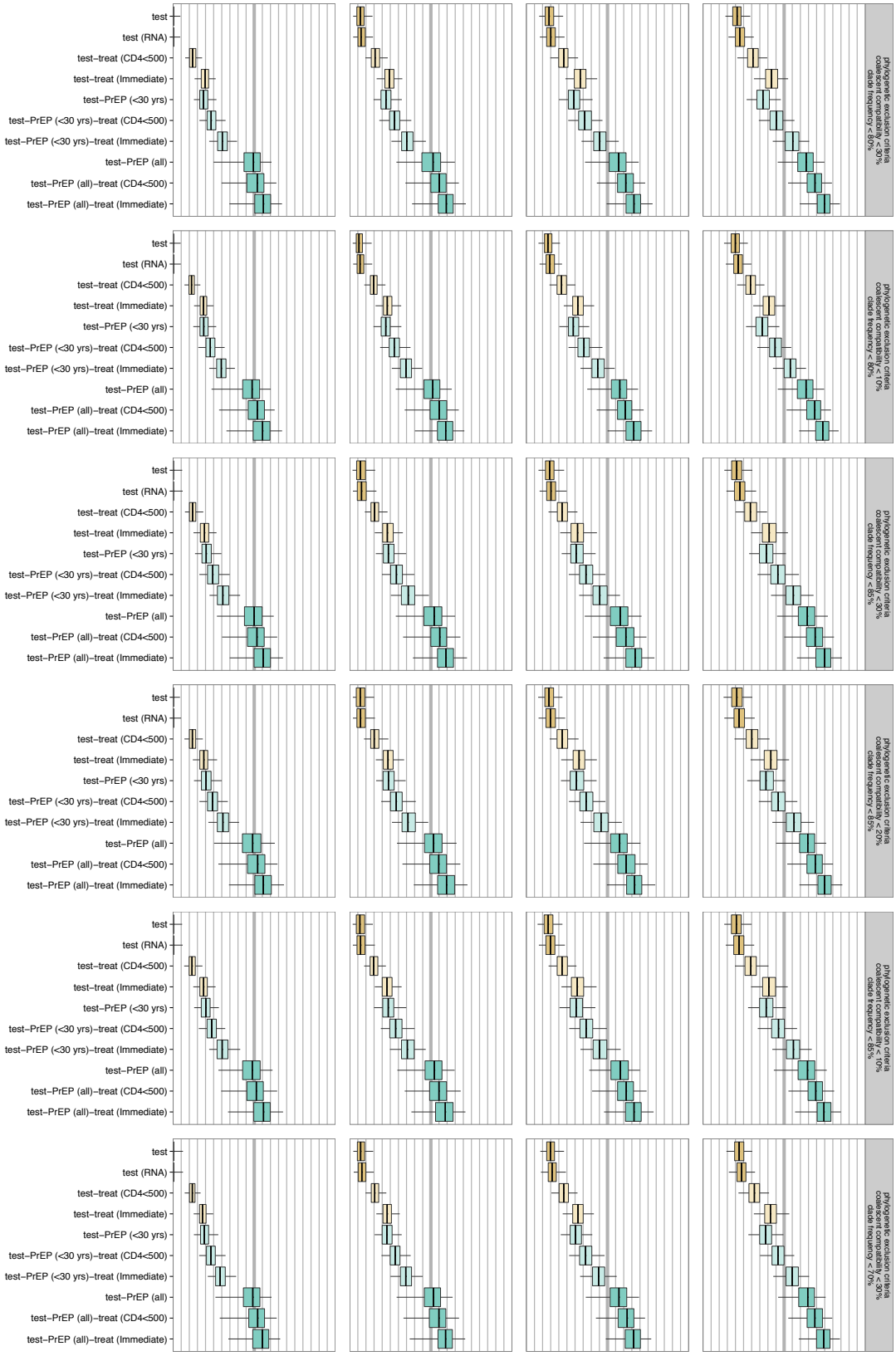


Figure S 20 Impact of phylogenetic transmission probabilities on the estimated proportion of averted infections. The proportion of transmissions that could have been averted was calculated as described in the Materials and Methods without adjusting for differences in the number of intervals per pair ($\boldsymbol{\omega_{ij\tau} = \varphi_{ij}\omega_{ij}}$, top row), with transmission from every probable transmitter equally likely ($\boldsymbol{\omega_{ij\tau} = 1/\tau_{ij}}$, middle row), and transmission from every interval equally likely ($\boldsymbol{\omega_{ij\tau} = 1}$, bottom row). It is clear that if men

after diagnosis are considered as infectious as undiagnosed men, then there is no secondary benefit in moving these individuals to stages further down in the HIV infection and care continuum.




test
test (RNA)
test–treat (CD4<500)
test–treat (Immediate)
test–PrEP (<30 yrs)
test–PrEP (<30 yrs)–treat (CD4<500)
test–PrEP (<30 yrs)–treat (Immediate)
test–PrEP (all)
test–PrEP (all)–treat (CD4<500)
test–PrEP (all)–treat (Immediate)
phylogenetic exclusion criteria
cophenetic distance > 30%
clade frequency < 80%
phylogenetic exclusion criteria
cophenetic distance > 30%
clade frequency < 85%
phylogenetic exclusion criteria
cophenetic distance > 30%
clade frequency < 70%

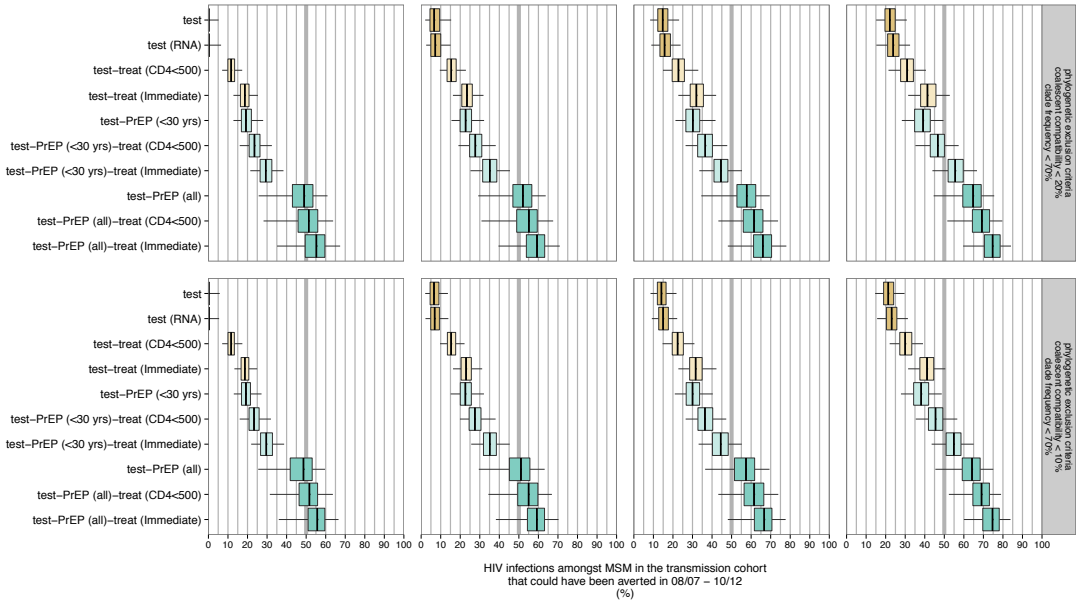


Figure S 21 Impact of infection time estimates and phylogenetic exclusion criteria on the estimated proportion of averted infections. The proportion of transmissions that could have been averted was calculated as described in the Materials and Methods, but using alternative upper and lower phylogenetic exclusion criteria as described in the row panels. The estimated proportions averted did not vary substantially.

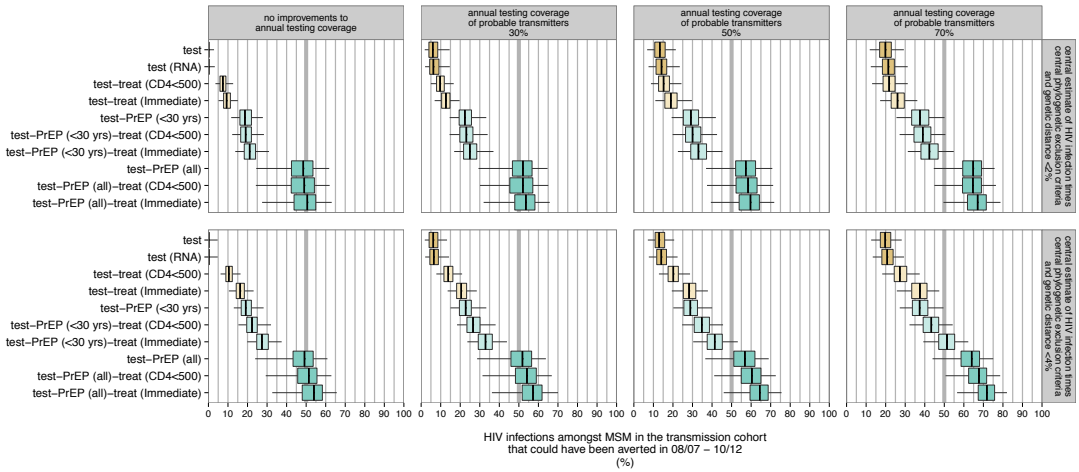


Figure S 22 Impact of additional genetic distance criteria on the estimated proportion of averted infections per biomedical intervention. The proportion of transmissions that could have been averted was calculated as described in the Materials and Methods, but potential transmitters were also excluded with additional genetic distance criteria described in the row panels. The estimated short-term impact of prevention strategies is insensitive to an additional 4% genetic distance criterion. With an addition 2% genetic distance criterion, the predicted impact that test-and-treat strategies could have had on reducing incidence is lower than without a genetic distance criterion.

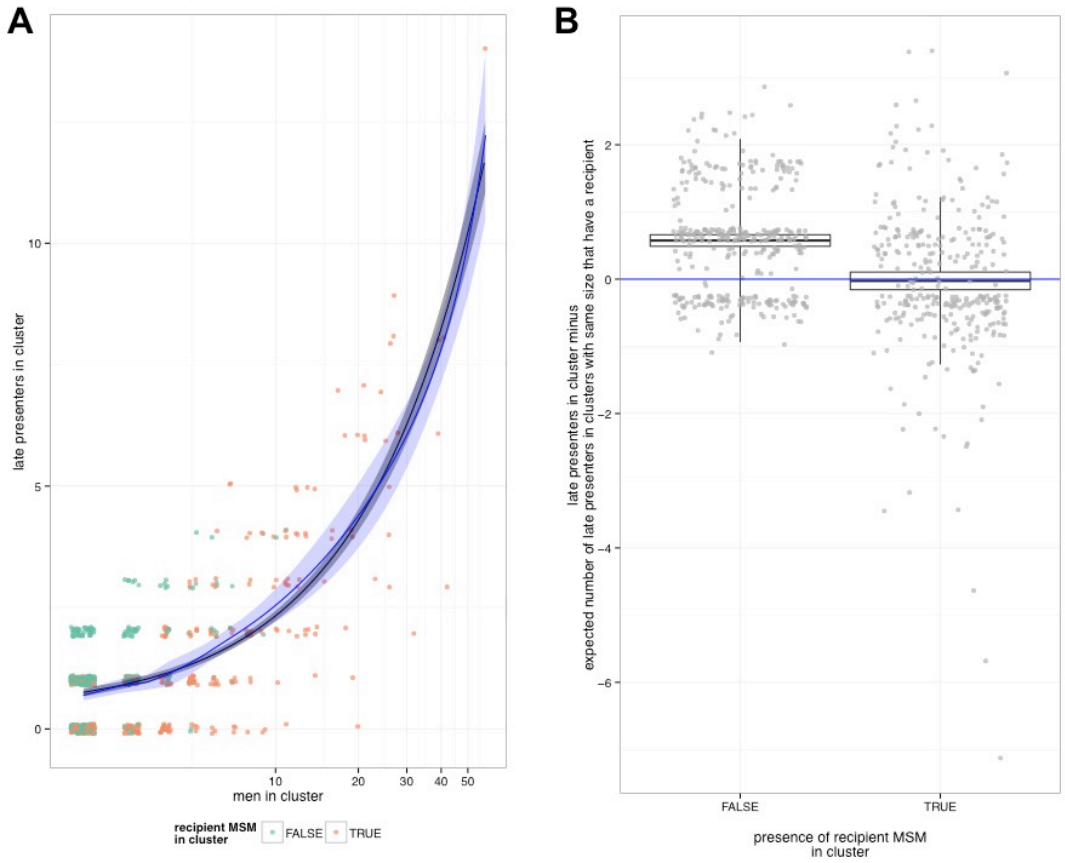


Figure S 23 Differences in transmission networks with and without a recipient MSM. To investigate if transmitters to MSM in recent infection at diagnosis might differ from typical transmitters, we considered all identified viral phylogenetic clusters with and without a recipient MSM. We then sought to evaluate if the number of late presenters (defined as men with a first CD4 count below 350 cells/ml within 12 months after diagnosis and before ART start) was enriched amongst clusters with no recipient MSM. **(A)** The number of late presenters increases linearly with cluster size (blue: loess smooth, black: regression model with linear dependence on cluster size). Adjusting for cluster size, we then fitted the following Poisson model with identity link

$$n_i \sim Poi(\mu_i)$$
$$\mu_i = \beta_0 + \beta_1 r_i + \beta_2 s_i,$$

where

n_i	number of late presenters in cluster i
μ_i	mean number of late presenters in cluster i
r_i	1 if cluster i has a recipient MSM and 0 otherwise
s_i	size of cluster i,

to estimate the contrast β_1 between the average number of late presenters in clusters with and without a recipient. β_1 was significantly smaller than zero after differences in cluster size were adjusted for (p=2e-14). **(B)** To visualize, we calculated the adjusted number of late presenters in a cluster as the number of late presenters minus the expected number of late presenters in clusters with a recipient MSM ($r_i = 1$) under the fitted Poisson model. Dots indicate the adjusted number of late presenters for all viral phylogenetic transmission clusters. Boxplots indicate the mean and two standard deviations from the mean. Viral phylogenetic transmission clusters without a recipient MSM are clearly enriched in late presenters.

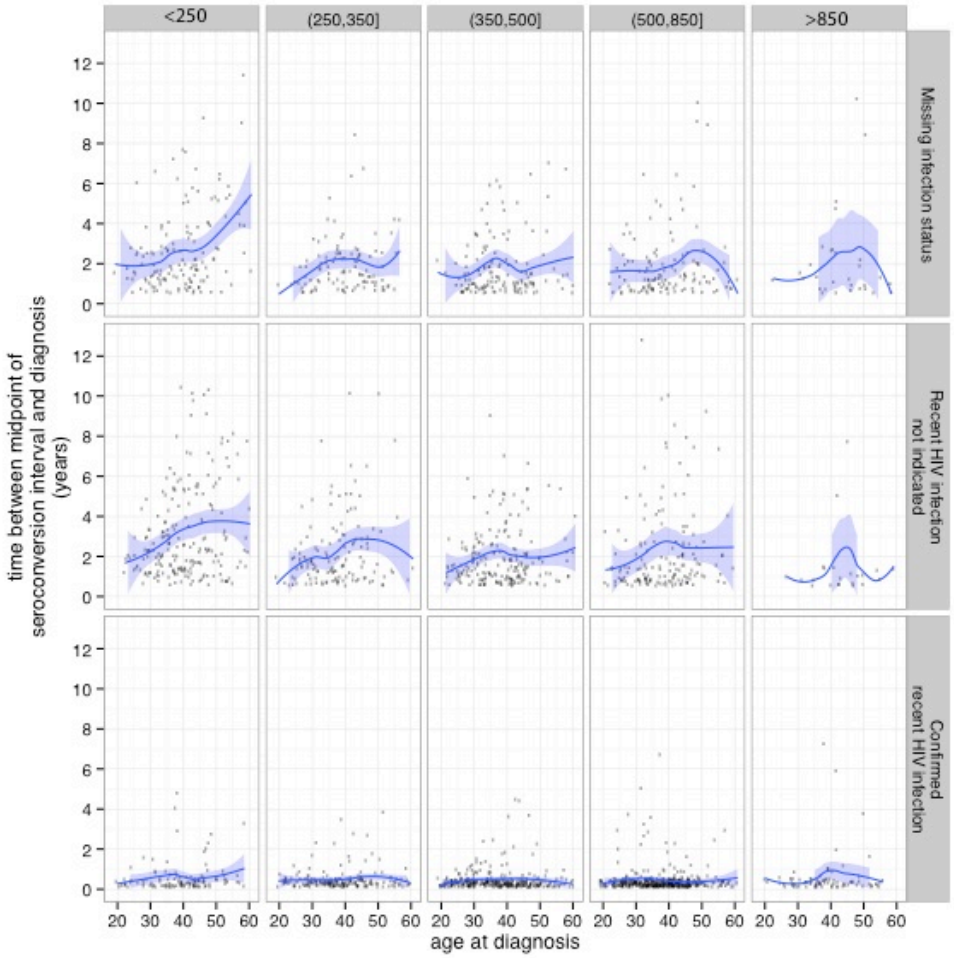


Figure S 24 Exploratory local polynomial regression fits to the time to diagnosis of MSM with a last negative test in the ATHENA cohort. Data from MSM with a last negative test (dots) is shown on top of the mean local polynomial regression fit (blue line) and the corresponding 95% quantile ranges (light blue region).

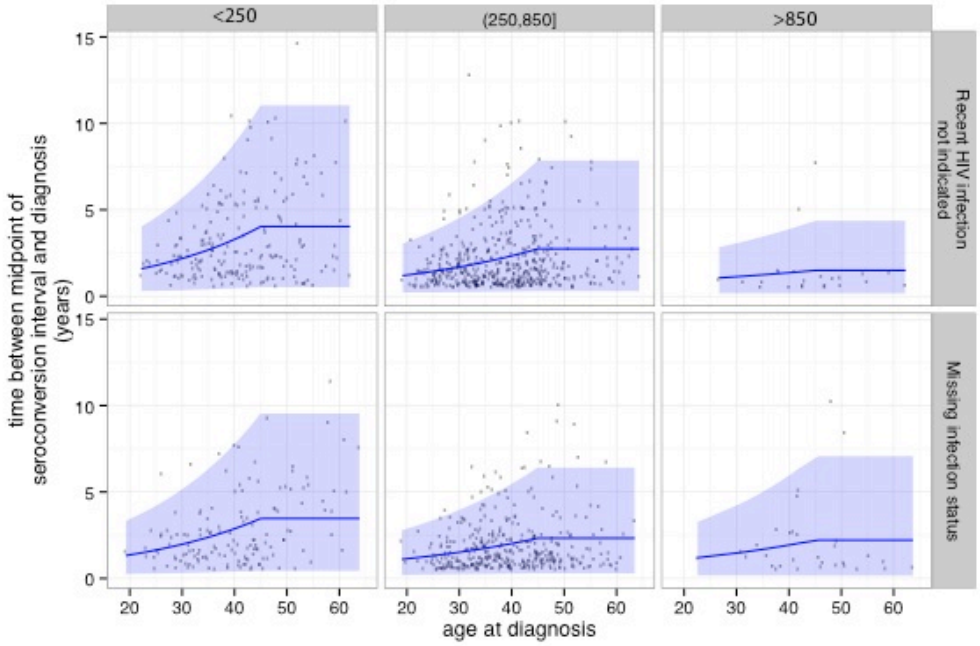


Figure S 25 Multivariable Gamma regression model fitted to the time between the midpoint of the seroconversion interval and diagnosis of MSM with a last negative test in the ATHENA cohort. Data from MSM with a last negative test (dots) are shown on top of the mean Gamma regression fit (blue line) and the corresponding 95% quantile ranges (light blue region).

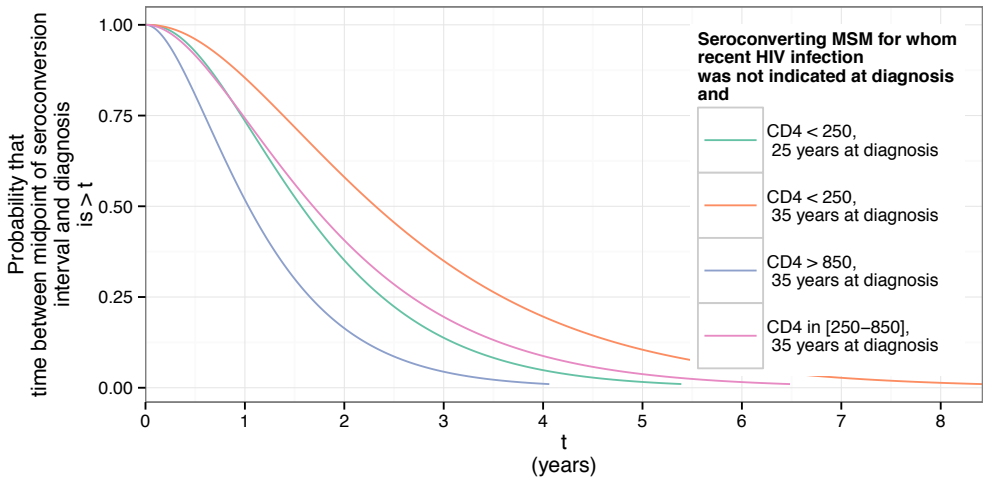


Figure S 26 Estimated probability that the time between the midpoint of the seroconversion interval and diagnosis among MSM with a last negative test is larger than t years.

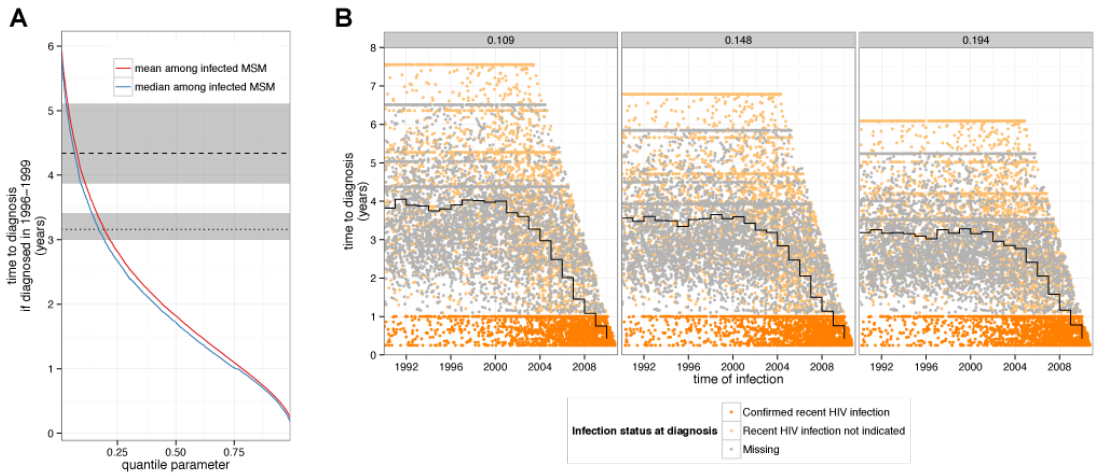


Figure S 27 Time to diagnosis estimates. (A) Average estimated time to diagnosis for MSM infected between 1996-1999 under the regression model as a function of the quantile parameter. The maximum-likelihood estimate and the 95% confidence interval derived from two mathematical modelling studies are highlighted in black/grey. (B) Predicted time to diagnosis for all MSM in the ATHENA cohort with estimated date of infection between 1990-2011. The three subplots show the higher, central and lower estimates of time to diagnosis that correspond to the calibrated quantile parameters 0.109, 0.148, 0.194.

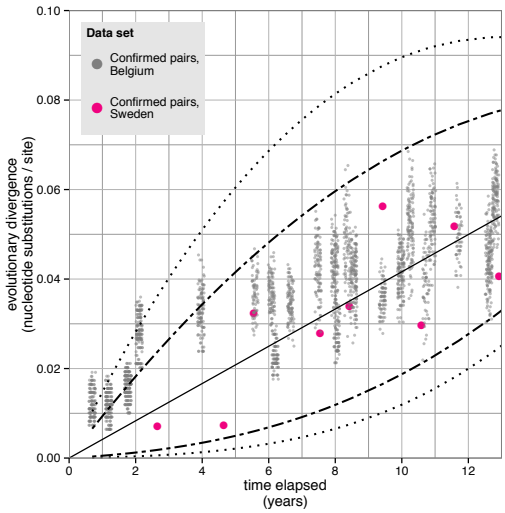


Figure S 28 Genetic distance among sequence pairs from transmitter-recipient pairs in the Belgium and Swedish transmission chains. The 80% and 95% interquartile range under the fitted probabilistic molecular clock model are shown with dashed and dotted lines.

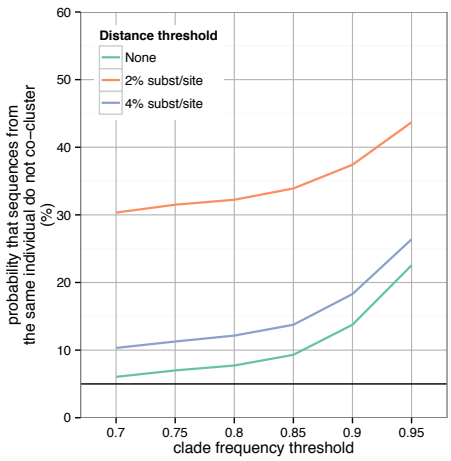


Figure S 29 Approximate type-I error of the phylogenetic clustering criterion as a function of the clade frequency threshold. Approximate type-I error in excluding sequences from the same individual if no genetic distance threshold is used (green), a 2% substitutions / site distance threshold is used (orange), and a 4% substitutions / site genetic distance threshold is used (purple).

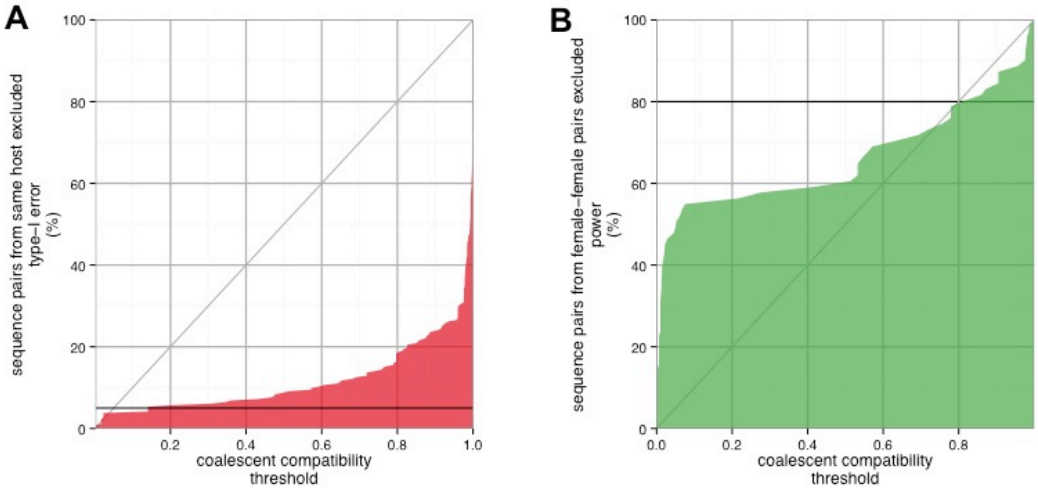


Figure S 30 Type-I error and power of the coalescence compatibility test. (A) Approximate type-I error in excluding co-clustering sequences from the same individual. (B) Power in excluding co-clustering female-female pairs.

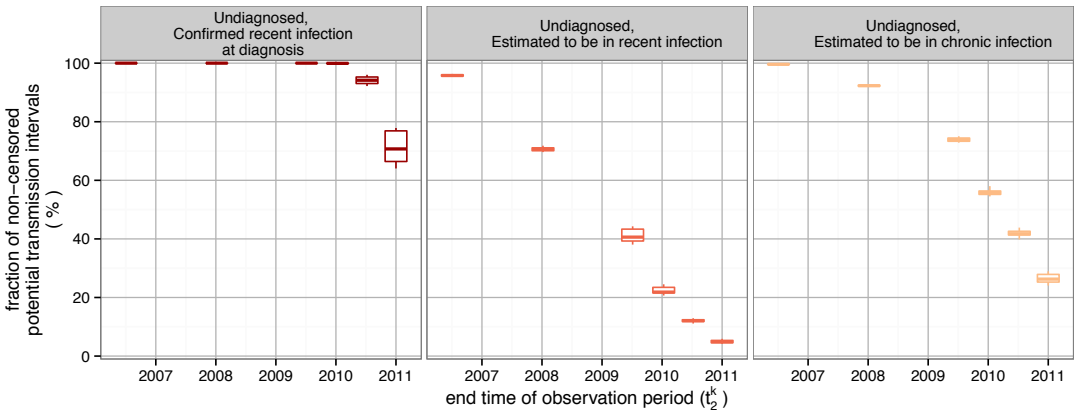


Figure S 31 Estimated fraction of non-censored potential transmission intervals. The fraction of non-censored potential transmission intervals was estimated for six time intervals, $[t_1, t_2]$= 1996/07-2006/06, 2006/07-2007/12, 2008/01-2009/06, 2009/07-2009/12, 2010/01-2010/06, 2010/07-2010/12. The fraction is plotted at the end time of each time interval.

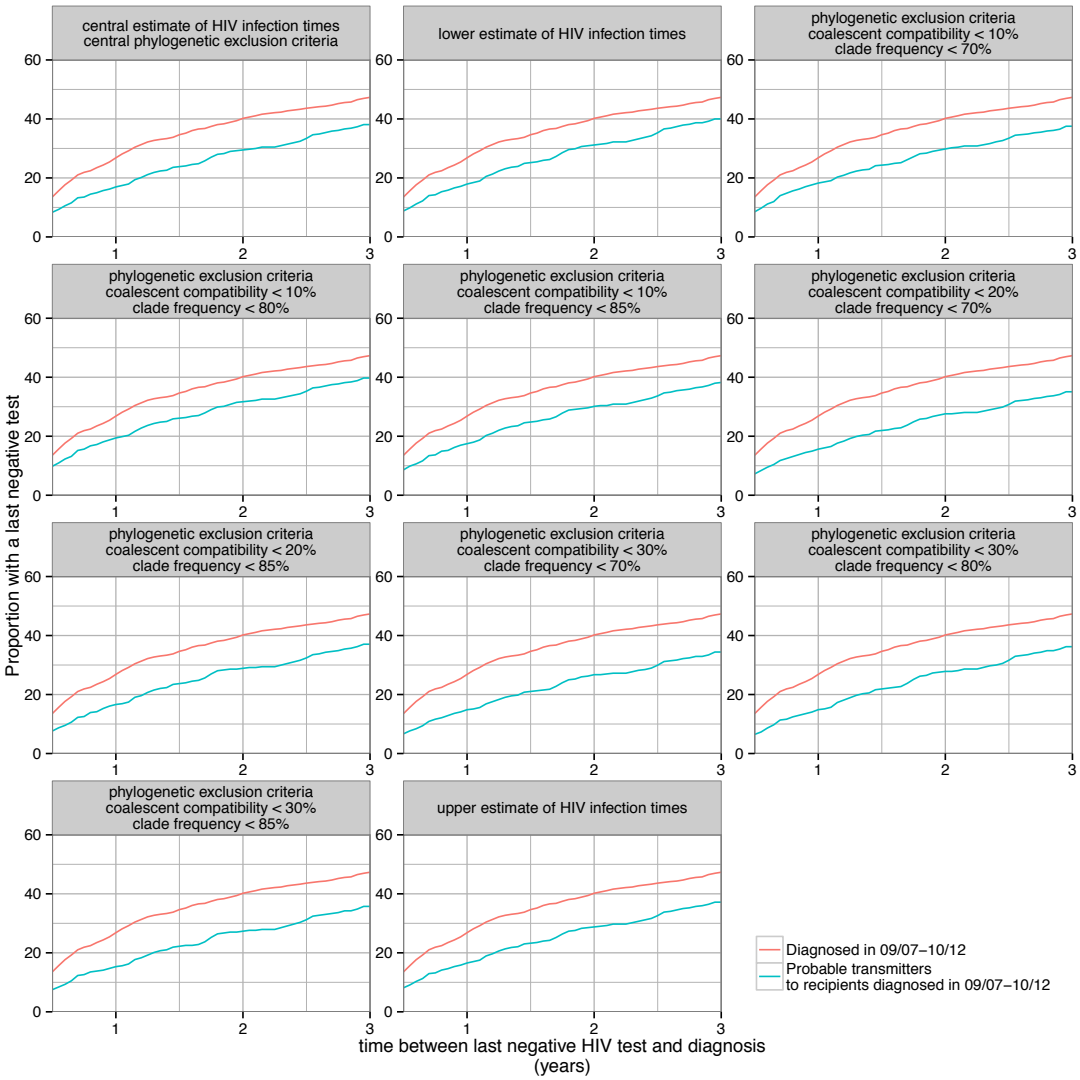


Figure S 32 Time between last negative test and diagnosis amongst MSM diagnosed in 2009/07-2010/12 and probable transmitters of recipients diagnosed in 2009/07-2010/12.

Table S1 Clinical and viral sequence data used in this study

Sample sizes					
Registered MSM by March 2013	11,863				
MSM with confirmed recent infection	1,794				
Viral load measurements	265,853				
CD4 measurements	284,151				
Coverage of linked clinical data					
Time of diagnosis of recipient MSM of potential transmitters	Unknown recency status at diagnosis (%)	No CD4 measurement between diagnosis and ART start (%)	No viral load measurement after ART start (%)		
96/07-06/06	51	8.6	8.4		
06/07-07/12	48	8.0	8.2		
08/01-09/06	46	7.8	8.0		
09/07-10/12	45	7.6	8.0		
Frequency of linked clinical data					
	CD4 measurements between diagnosis and ART start of potential transmitters (number / year)				
	2.5%	25%	Median	75%	97.5%
96/07-06/06	0.16	1.38	2.75	3.93	6.71
06/07-07/12	0.18	1.61	2.92	3.99	6.73
08/01-09/06	0.19	1.67	2.97	4.01	6.74
09/07-10/12	0.19	1.68	2.97	4.01	6.74
	Viral load measurements after ART start of potential transmitters (number/year)				
	2.5%	25%	Median	75%	97.5%
96/07-06/06	0.72	2.13	2.70	3.29	5.00
06/07-07/12	0.76	2.11	2.66	3.23	4.54
08/01-09/06	0.75	2.10	2.64	3.21	4.50
09/07-10/12	0.76	2.09	2.62	3.19	4.41
Partial polymerase HIV-1 subtype B sequences					
	n				
available Dutch sequences	8,748				
enriched with 10 most similar sequences in the Los Alamos Sequence Database (http://www.hiv.lanl.gov/)	9,474				
Data set used in analysis, after exclusion of potentially recombinant sequences (identified with 3SEQ), and exclusion of sequences with length <= 250 nucleotides	9,054				
Charateristics of Dutch HIV-1 subtype B sequences used in analysis					
	n				
Dutch sequences	8,328				
Sequences sampled after ART start	3,693				
Patients sampled	6,231				
	5%	25%	Mean	75%	95%
Length of ATHENA sequences (nt)	1175	1235	1256	1274	1600
Time from diagnosis to sampling of an individual's first sequence (years)	0	0.03	2.8	4.1	13.5
Individuals with at least one sequence	MSM	Heterosexual	Drug user	Other	Unknown
Male	4767	518	123	82	174
Female	0	455	57	42	14

Table S2 Potential transmitters and potential transmission pairs to the recipient MSM

	Time of diagnosis of recipient MSM	**Total**					
		Central estimate of infection time ¶		Lower estimate of infection time ¶		Upper estimate of infection time ¶	
		All (n)	With a sequence (n)	All (n)	With a sequence (n)	All (n)	With a sequence (n)
Recipient MSM	Overall	1794	1045	1794	1045	1794	1045
	96/07-06/06	695	368	695	368	695	368
	06/07-07/12	323	216	323	216	323	216
	08/01-09/06	405	233	405	233	405	233
	09/07-10/12	371	228	371	228	371	228
Potential transmitters*	Overall	12193	5585	12189	5585	12193	5585
	96/07-06/06	9687	4322	9537	4239	9816	4376
	06/07-07/12	10179	4750	10093	4718	10272	4779
	08/01-09/06	10962	5148	10935	5142	10989	5161
	09/07-10/12	11419	5329	11415	5329	11419	5329
Potential transmission pairs*	Overall	9722349	4428060	9617162	4378332	9824133	4475601
	96/07-06/06	2712072	1158948	2649412	1125357	2777347	1193496
	06/07-07/12	2075669	961286	2052721	951409	2096432	969131
	08/01-09/06	2427787	1134094	2412127	1128919	2440400	1138425
	09/07-10/12	2506821	1173732	2502902	1172647	2509954	1174549

* Potential transmitters and potential transmission pairs were counted for recipient MSM with a sequence for computational reasons. ¶ The infection time of potential transmitters was estimated from their age at diagnosis, recency of HIV infection at diagnosis, and first CD4 count within 12 months of diagnosis. The central estimate is based on $\alpha = 0.148$ in the estimation procedure described in the supplementary online material; the lower estimate is based on $\alpha = 0.194$, and the upper estimate is based on $\alpha = 0.109$.

Table S3 Identified phylogenetically probable transmitters and phylogenetically probable transmission pairs to the recipient MSM.

	Time of diagnosis of recipient MSM	Total									
		Central exclusion criteria⊙ Clade freq <80% Coal comp <20%		Lower-I exclusion criteria⊙ Clade freq <80% Coal comp <30%		Upper-I exclusion criteria⊙ Clade freq <80% Coal comp <10%		Lower-II exclusion criteria⊙ Clade freq <85% Coal comp <20%		Upper-II exclusion criteria⊙ Clade freq <70% Coal comp <20%	
		(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)	(n)	(%)
Recipient MSM with a phylo-genetically probable transmitter	Overall	617	59.14 *	594	56.84 *	638	61.05 *	564	53.97 *	656	62.78 *
	96/07-06/06	165	44.84 *	150	40.76 *	172	46.74 *	146	39.67 *	179	48.64 *
	06/07-07/12	144	66.67 *	143	66.2 *	146	67.59 *	134	62.04 *	150	69.44 *
	08/01-09/06	152	65.24 *	149	63.95 *	159	68.24 *	138	59.23 *	160	68.67 *
	09/07-10/12	157	68.86 *	152	66.67 *	161	70.61 *	146	64.04 *	167	73.25 *
Phylo-genetically probable transmitters	Overall	903	16.17 §	841	15.06 §	981	17.56 §	823	14.74 §	975	17.46 §
	96/07-06/06	268	6.32 §	237	5.59 §	308	7.27 §	240	5.55 §	308	7.13 §
	06/07-07/12	331	7.02 §	302	6.4 §	362	7.67 §	307	6.46 §	359	7.56 §
	08/01-09/06	348	6.77 §	322	6.26 §	400	7.78 §	323	6.27 §	391	7.6 §
	09/07-10/12	407	7.64 §	370	6.94 §	448	8.41 §	367	6.89 §	442	8.29 §
Phylo-genetically probable transmission pairs	Overall	2343	0.05 ¶	2059	0.05 ¶	2698	0.06 ¶	2097	0.05 ¶	2718	0.06 ¶
	96/07-06/06	401	0.04 ¶	353	0.03 ¶	477	0.04 ¶	380	0.03 ¶	498	0.04 ¶
	06/07-07/12	506	0.05 ¶	446	0.05 ¶	569	0.06 ¶	488	0.05 ¶	579	0.06 ¶
	08/01-09/06	731	0.06 ¶	636	0.06 ¶	842	0.07 ¶	642	0.06 ¶	819	0.07 ¶
	09/07-10/12	705	0.06 ¶	624	0.05 ¶	810	0.07 ¶	587	0.05 ¶	822	0.07 ¶

⊙See supplementary online material for a description of sensitivity analyses. * Proportion among all recipient MSM with a sequence, §proportion among potential transmitters with a sequence, ¶proportion among potential transmission pairs with sequences from both individuals based on central estimates of infection times.

Table S4 Demographic and clinic characteristics of the 3,025 MSM with a last negative test, that were used to fit the multivariable regression model

Time of diagnosis	**<=06/06**	**06/07-07/12**	**08/01-09/06**	**09/07-10/12**		
	931	609	721	764		
Age at diagnosis	**<=25**	**26-35**	**36-45**	**46-55**	**>55**	
	242	944	1114	545	180	
CD4 count at diagnosis	**No CD4 measurement to date**	**CD4 measurement after 1 year of diagnosis**	**First CD4 measurement after ART start**	**≤250**	**251-850**	**>850**
	10	93	52	471	2169	230
Infection status at diagnosis	**Confirmed recent HIV infection**	**Recent HIV infection not indicated**	**Missing**			
	1369	722	934			