Sources of HIV infection among men having sex with men and implications for prevention

Oliver Ratmann,1* Ard van Sighem,2 Daniela Bezemer,2 Alexandra Gavryushkina,3 Suzanne Jurriaans,4 Annemarie Wensing,5 Frank de Wolf,1 Peter Reiss,2,6 Christophe Fraser,1 ATHENA observational cohort

New HIV diagnoses among men having sex with men (MSM) have not decreased appreciably in most countries, even though care and prevention services have been scaled up substantially in the past 20 years. To maximize the impact of prevention strategies, it is crucial to quantify the sources of transmission at the population level. We used viral sequence and clinical patient data from one of Europe’s nationwide cohort studies to estimate probable sources of transmission for 617 recently infected MSM. Seventy-one percent of transmissions were from undiagnosed men, 6% from men who had initiated antiretroviral therapy (ART), 1% from men with no contact to care for at least 18 months, and 43% from those in their first year of infection. The lack of substantial reductions in incidence among Dutch MSM is not a result of ineffective ART provision or inadequate retention in care. In counterfactual modeling scenarios, 19% of these past cases could have been averted with current annual testing coverage and immediate ART to those testing positive. Sixty-six percent of these cases could have been averted with available antiretrovirals (immediate ART provided to all MSM testing positive, and preexposure antiretroviral prophylaxis taken by half of all who test negative for HIV), but only if half of all men at risk of transmission had tested annually. With increasing sequence coverage, molecular epidemiological analyses can be a key tool to direct HIV prevention strategies to the predominant sources of infection, and help send HIV epidemics among MSM into a decisive decline.

INTRODUCTION

Combination antiretroviral therapy (ART) transformed HIV from a deadly to a lifelong disease and is also one of the most effective strategies for preventing onward infections (1, 2). However, among men having sex with men (MSM), the substantial scale-up of ART in the past 20 years has not resulted in appreciable reductions of new HIV infections and diagnoses (Table 1) (3). Building on successful behavioral and biomedical HIV prevention strategies (4), further interventions exist that could be used to reduce the number of HIV infections among MSM. The 2016 World Health Organization (WHO) guidelines now recommend ART initiation regardless of CD4 cell count after diagnosis (immediate ART), as well as provision of antiretrovirals as preexposure prophylaxis (PrEP) to those at substantial risk of infection (5). Future prevention programs could focus on one or both recommended interventions, as well as on increased routine HIV testing and diagnosis (6); RNA testing to detect MSM with acute infection, at which time they are thought to be most infectious (7), and improved adherence and linkage support to assist patients with attaining and sustaining undetectable viral loads while on ART (8). The potential impact of any of these interventions, and specifically those recommended by the WHO, relies crucially on how many HIV transmissions originate from different stages in the entire HIV infection and care continuum, ranging from undiagnosed acute infection through treated infection and loss to follow-up. This has been challenging to measure directly through classical epidemiological approaches.

Here, we use the viral phylogenetic relationship between partial HIV-1 subtype B polymerase sequences to reconstruct past, probable transmission events in the Netherlands (Fig. 1). These sequences were routinely collected for drug resistance testing of HIV-infected patients who are in care (9). Among sampled MSM, 94% were of subtype B. Then, we use clinical records to determine the staging of probable transmission events within the infection and care continuum (Fig. 2A and Table 2). This enabled us to estimate the population-level proportion of transmissions among the reconstructed transmission events that are attributable to the 14 stages of the infection and care continuum in Fig. 2A. Transmissions could be attributed to stages before diagnosis because HIV sequences, always collected after diagnosis, diverge fast enough to indicate past transmission events (10). Similarly, transmissions could also be attributed to men with no contact to care for at least 18 months. Finally, using these estimates, we quantified the potential impact of currently not implemented prevention programs in the Dutch MSM population, had they been used in the last 3 years. In particular, we evaluate if the revised 2016 WHO guidelines on immediate ART and PrEP could have substantially altered the course of the Dutch HIV epidemic among MSM.

Understanding which interventions should be prioritized for the Dutch MSM epidemic is an important case study. First, the number of new MSM infections in the Netherlands has not decreased appreciably (9) despite comprehensive linkage and retention in care, substantial ART scale-up free of charge, and frequent follow-up to maintain viral control of the vast majority of those on ART (Table 1). Second, similar epidemic trends are reported from other countries with an overall equally comprehensive cascade of care (Table 1), casting more general doubts on the population-level impact of current prevention strategies targeting MSM epidemics (11). Third, nearly all HIV-infected MSM in care are enrolled in the nationwide Aids Therapy Evaluation in the Netherlands cohort (ATHENA) since early 1996 (9). HIV care is monitored comprehensively at high frequency (clinical visits, treatment histories, comorbidities...
Phylogenetically probable transmission events

Genetic sequences of the virus alone cannot prove epidemiological linkage (13). However, most of the potential transmission pairs could be ruled out as implausible based on the phylogenetic relationship of the viral sequences. The viral phylogeny among the Dutch sequences and their closest matches in the Los Alamos HIV sequence database (www.hiv.lanl.gov/) were reconstructed with maximum-likelihood methods, and reliable subtrees were identified (see Material and Methods). Potential transmitters whose sequences did not occur in the same reliable subtree as those of the recipient MSM were excluded (stage C in Fig. 1) (10), as were potential transmitters whose sequences were incompatible with a direct HIV transmission event (stage D in Fig. 1) (14). Direct transmission could be excluded in 99.96% of all potential transmission pairs. We identified 903 phylogenetically probable transmitters to 617 recipient MSM in 2343 pairs. Our analyses are based on this open observational cohort of past, phylogenetically reconstructed transmission events.

To guide and interpret this exclusion analysis, we evaluated patterns of viral divergence between sequences isolated from epidemiologically confirmed transmission pairs (14, 15) and pairings of Dutch MSM that could not have infected each other (see Material and Methods). On the basis of these pairs, the above exclusion criteria were highly specific (true transmitters to recipients are not excluded, >90%), whereas sensitivity was low (incorrect transmission

---

**RESEARCH ARTICLE**

**RESULTS**

Potential transmissions to MSM that were confirmed to have recent infection at time of diagnosis

By 2013, 11,863 HIV-infected MSM were registered and still in care in the Netherlands. To estimate their sources of transmission and then the impact of prevention programs, we focused on transmissions to MSM that were recently infected at time of diagnosis (stage A in Fig. 1). Between July 1996 and December 2010, 1794 MSM were confirmed to have been infected at most 12 months before diagnosis. Types of evidence were a previous negative HIV test (76%), laboratory diagnosis (7%), or clinical diagnosis of acute infection (17%). For 1045 (58%) of these, a sequence was available. To these recipient MSM, we considered as potential transmitters all HIV-infected men whose course of infection overlapped with the infection window of the recipient (stage A in Fig. 1). Using this approach, we could resolve the timing and direction of potential transmission events (12). Of all 12,207 potential transmitters, 5593 (46%) had a viral sequence and formed ~4.4 million potential transmission pairs with sequences available for both individuals (stage B in Fig. 1).

---

**Table 1. HIV incidence trends and care for infected MSM in the Netherlands and other countries.**

<table>
<thead>
<tr>
<th>Country</th>
<th>Uninfected MSM testing annually</th>
<th>Diagnosed MSM receiving ART</th>
<th>Treated MSM with suppressed viral load</th>
<th>MSM retained in care</th>
<th>HIV incidence among MSM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year</td>
<td>%</td>
<td>Year</td>
<td>%</td>
<td>Median CD4 count at ART initiation (cells/ml)</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>2003</td>
<td>??</td>
<td>2003</td>
<td>79</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>38.4*</td>
<td>2013</td>
<td>90</td>
<td>382</td>
</tr>
<tr>
<td>Australia</td>
<td>2013</td>
<td>61.1†</td>
<td>2013</td>
<td>75 †</td>
<td>379 †</td>
</tr>
<tr>
<td>British Columbia</td>
<td>2009</td>
<td>51 ††</td>
<td>2014</td>
<td>85 ††</td>
<td>411 ††</td>
</tr>
<tr>
<td>Switzerland</td>
<td>2010</td>
<td>39.3‡</td>
<td>2014</td>
<td>86 ‡</td>
<td>402 ‡</td>
</tr>
<tr>
<td>UK</td>
<td>2010</td>
<td>36.4 ‡</td>
<td>2013</td>
<td>86 ‡‡</td>
<td>420 ‡‡</td>
</tr>
</tbody>
</table>

pairs could not always be excluded, ~60%). This indicates that the actual transmitter is almost certainly among the phylogenetically reconstructed, probable transmitters, provided he was sequenced. From the known sequence coverage alone, we expected that about half of all 1045 recipient MSM with a sequence had their actual transmitter sampled, further suggesting that the actual transmitter is among the phylogenetically reconstructed, probable transmitters for the large majority of the reconstructed 617 transmission events.

The clinical and demographic characteristics of the selected 617 recipient MSM were typical of all 1794 MSM that were confirmed to have recent infection at time of diagnosis (Table 3).

Characterization of individual transmission events by stage in the HIV infection and care continuum

Using clinical records, we then enumerated all stages in the HIV infection and care continuum during which the 617 transmission events could have occurred. Probable transmitters progressed in stage over time and overlapped with infection windows in 13,169 time-resolved, 6-week-long transmission intervals (Fig. 2B). Censoring and sequence sampling biases were identified for each stage by comparing men with and without a sequence, and were adjusted in line with previous work (16). Reflecting targeted sequence collection, intervals were not missing at random (Fig. 2C and fig. S9). Each interval (0.7 to 1.6%) from men with no contact to care for at least 18 months. An estimated 43% (37 to 46%) of the 617 recipient MSM were infected by men undergoing their first year of infection.

Impact of prevention strategies

Figure 4 describes the counterfactual prevention scenarios for which we calculated the proportion of transmissions in the cohort that could have been averted between mid-2008 to December 2010, had we intervened to redistribute the identified, probable transmitters to less infectious infection/care stages. Young MSM are at particularly high risk of infection (17, 18). We therefore considered—along the revised 2016 WHO guidelines (5)—rollout of immediate ART to all infected MSM and PrEP to half of all MSM age 30 years or younger who test negative: at most, 30% (22 to 39%) of infections could have been averted without increased annual testing. Immediate ART alone could have averted 19% (13 to 26%) of these cases at current testing levels. In practice, low adherence is associated with decreasing effectiveness of PrEP (19). We assumed an 86% efficacy of PrEP as reported in the recent Ipergay and PROUD trials (20, 21). Figure S12 reports the impact of lower efficacy values. Figure S13 reports the impact of lower or higher PrEP coverage. Next, we considered increased annual testing. Only 17% of identified probable transmitters had a last negative test in the year before diagnosis compared to 27% of diagnosed MSM between mid-2008 and December 2010 and 38% of uninfected MSM in 2013 (Table 1). If half of all transmitters was associated with a phylogenetic transmission probability on the basis of the genetic distance between sequences from the transmitter and recipient and the time elapsed since the putative transmission interval and the sampling dates of both individuals (see Materials and Methods and fig. S10). For each recipient, the probability that transmission occurred from one of the 14 stages then depends on the number of his probable transmitters in that stage and the transmission probabilities associated with each of the corresponding transmission intervals (see Materials and Methods).

Sources of HIV transmission

The population-level proportions of HIV transmissions attributable to the 14 infection/care stages were obtained by summing individual-level transmission probabilities by stage across all recipients and are shown in Table 4. Figure 3 compares the proportion of transmissions from each stage to the population-level proportion of infected men in these stages. Between July 1996 and December 2010, an estimated 71% (66 to 73%) of all 617 transmission events originated from undiagnosed men, 22% (21 to 26%) from diagnosed but not yet treated men, 6% (5 to 8%) from men who initiated ART, and 1%...
had tested annually, immediate ART and PrEP to half of all MSM age 30 years or younger who test negative could have averted 45% (34 to 56%) of infections. Comprehensive rollout of PrEP to half of all men testing negative irrespective of their age would have substantially boosted the combination intervention: 66% (50 to 78%) of infections could have been averted.

**DISCUSSION**

HIV epidemics among MSM have, unlike other settings (22), not declined appreciably with substantial improvements to care and ART scale-up (Table 1). We characterized 617 past transmission events among MSM in the Netherlands based on phylogenetic and clinical data, estimated their sources throughout the infection and care continuum, and quantified the impact that biomedical prevention programs could have had in averting the reconstructed transmission events. Analyzing this transmission cohort, we aim to inform the design of future prevention interventions beyond high levels of ART coverage and the numerous successful behavioral interventions that are already in place (9).

A potential limitation of this study is that transmitters to MSM with recent infection at diagnosis may differ from typical transmitters. On average, fewer men diagnosed late with a CD4 count below 350 cells/ml occurred in phylogenetic transmission clusters with a recipient MSM compared to those without (fig. S23). This may imply that, overall,
the proportion of transmissions from undiagnosed men with chronic infection is higher and, consequently, that the impact that immediate ART could have had is lower than our estimates. Conversely, the impact of increased annual testing and PrEP could be larger than reported, if men diagnosed late are not more difficult to reach than the average transmitter in our cohort. Further, this study focuses on the sources and prevention of in-country transmissions: 97% of the recipient MSM reported that infection was likely acquired in the Netherlands, compared to 86% of diagnosed MSM. The contribution of cross-border transmissions may increase as the response is strengthened (23), an effect that we did not consider. Phylogenetic uncertainty and the phylogenetic exclusion criteria had little impact on our findings (figs. S14 to S22).

**Fig. 2. Phylogenetically probable transmission intervals, linked to stages in the infection and care continuum.** (A) Left: Each recipient could have been infected during his infection window from multiple probable transmitters. For each transmitter, the transmission window was split into 6-week-long probable transmission intervals. Infection/care stages were assigned to these intervals on the basis of clinical data to reflect progression of the transmitters through the infection/care continuum. Right: Relationship between the 14 infection/care stages as defined in Table 2. Transmitters progress unidirectionally, except for stages after first viral suppression, or when individuals reenter care (as indicated by arrows). (B) For each stage, the total number of observed transmission intervals to recipient MSM during their infection windows is shown by date of diagnosis of the recipients. Overall, the number of probable transmission intervals per recipient increases with time, reflecting the increasing number of infected men in care. Transmitters are increasingly less likely to have been diagnosed by 2013, resulting in a decreasing number of undiagnosed transmission intervals toward the present. (C) In addition to censoring, diagnosed transmitters may not have a sequence sampled. Comparing men with and without a sequence in the near-complete population cohort, we could adjust for these biases. The total number of expected missing transmission intervals to recipients is shown, along with 95% bootstrap confidence intervals. Observed and expected missing transmission intervals were associated with phylogenetic transmission probabilities, which sum to 1 per recipient.
Table 3. Characteristics of the recipient MSM with identified sources of transmission. IQR, interquartile range.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Recipient MSM with a phylogenetically probable transmitter (n = 617)</th>
<th>Recipient MSM with or without a sequence (n = 1794)</th>
<th>Diagnosed MSM (n = 7978)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence for infection in the past year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous negative test in the past year (%)</td>
<td>77</td>
<td>76</td>
<td>17</td>
</tr>
<tr>
<td>Laboratory diagnosis (%)</td>
<td>8</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Clinical diagnosis of acute infection (%)</td>
<td>15</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Age at diagnosis (years; mean and IQR)</td>
<td>36.8 (29.5–42.9)</td>
<td>37.2 (29.9–43.5)</td>
<td>38.7 (31.3–45.1)</td>
</tr>
<tr>
<td>First CD4 count within 12 months of diagnosis and before ART start (cells/ml; mean and IQR)</td>
<td>505 (350–630)</td>
<td>534 (360–670)</td>
<td>402 (200–560)</td>
</tr>
<tr>
<td>Viral load count within 12 months of diagnosis (log10 RNA; mean and IQR)</td>
<td>4.9 (4.4–5.5)</td>
<td>4.8 (4.3–5.4)</td>
<td>4.7 (4.3–5.3)</td>
</tr>
<tr>
<td>In care in the Amsterdam metropolitan area (%)</td>
<td>45.1</td>
<td>43.5</td>
<td>43.6</td>
</tr>
<tr>
<td>Last negative test within 12 months before diagnosis (%)</td>
<td>77.0</td>
<td>76.1</td>
<td>17.1</td>
</tr>
<tr>
<td>Self-reported in country infection* (%)</td>
<td>96.9</td>
<td>91.9</td>
<td>88.5</td>
</tr>
</tbody>
</table>

*Of those self-reporting a country of origin.

Table 4. Proportion of transmissions by stage in the HIV infection and care continuum.

<table>
<thead>
<tr>
<th>Infection/care stage of transmitter</th>
<th>% of transmissions by time of diagnosis of recipient MSM (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undiagnosed (total)</td>
<td>70.9 (65.8–72.5)</td>
</tr>
<tr>
<td>Confirmed recent infection at diagnosis</td>
<td>15.5 (11.9–17.4)</td>
</tr>
<tr>
<td>Estimated to have recent infection</td>
<td>25.1 (19.4–28.1)</td>
</tr>
<tr>
<td>Estimated to have chronic infection</td>
<td>30.3 (28–34)</td>
</tr>
<tr>
<td>Diagnosed (total)</td>
<td>22.4 (20.7–26.2)</td>
</tr>
<tr>
<td>Diagnosed &lt;3 months, recent infection at diagnosis</td>
<td>2.9 (2.2–4.1)</td>
</tr>
<tr>
<td>No CD4 measured</td>
<td>1.6 (1.2–2.4)</td>
</tr>
<tr>
<td>CD4 &gt;500</td>
<td>8.3 (7–10.3)</td>
</tr>
<tr>
<td>CD4 in 350–500</td>
<td>6.4 (5.4–7.9)</td>
</tr>
<tr>
<td>CD4 &lt;350</td>
<td>3.4 (2.5–4.3)</td>
</tr>
<tr>
<td>ART initiated (total)</td>
<td>5.7 (5.2–7.8)</td>
</tr>
<tr>
<td>Before first viral suppression</td>
<td>1.8 (1.6–2.7)</td>
</tr>
<tr>
<td>After first viral suppression</td>
<td></td>
</tr>
<tr>
<td>No viral load measured</td>
<td>0.5 (0.3–1)</td>
</tr>
<tr>
<td>No viral suppression</td>
<td>1.4 (0.9–2.1)</td>
</tr>
<tr>
<td>Viral suppression, one observation</td>
<td>0.4 (0.3–0.8)</td>
</tr>
<tr>
<td>Viral suppression, ≥2 observations</td>
<td>1.6 (1.1–2.5)</td>
</tr>
<tr>
<td>Not in contact</td>
<td>1 (0.7–1.6)</td>
</tr>
<tr>
<td>Recent infection (total)</td>
<td>43.5 (36.6–46)</td>
</tr>
</tbody>
</table>
Fig. 3. Proportion of transmissions by stage in the infection and care continuum versus proportion of these stages among infected men. (A) Relative frequency of infection/care stages in the population, among potential transmitters that overlap with the infection windows of recipient MSM and could have, in principle, transmitted to one of the recipient MSM (stage A in Fig. 1; color codes as in Fig. 2). (B) Proportion of the 617 transmission events attributable to each infection/care stage (bar: 95% bootstrap confidence interval).

Fig. 4. Impact of biomedical interventions among MSM in the Netherlands. Estimated proportion of transmissions that could have been averted in 08/07 – 10/12 (A) No improvements to annual testing coverage, (B) Annual testing coverage of probable transmitters 30%, (C) Annual testing coverage of probable transmitters 50%, (D) Annual testing coverage of probable transmitters 70%.

potential caveat to the robustness of our findings is that only half of all potential transmitters had a viral sequence sampled. Although population-level sampling biases were adjusted, we must acknowledge that the actual transmitter may not have been sampled for all recipients. Improving sequence sampling coverage at time of diagnosis is needed to facilitate phylogenetic prevention analyses (24).

The identified sources of transmission imply, first, that viral suppression induced by ART is highly effective in preventing transmissions in this population (Fig. 3). The relative risk of HIV transmission from men after ART initiation varies by stage but is always estimated well below 1 when compared to diagnosed, untreated men with a CD4 count above 500 cells/ml, and is in particular 0.04 (0.02 to 0.1) for men with viral suppression (fig. S11).

Second, very few transmissions are attributable to temporary or permanent loss to follow-up, which must be considered in the context of high linkage and retention to care in the Netherlands: few diagnosed MSM had subsequently no contact to care for at least 18 months (8.2%) and most reentered care within 5 years (69%) (9). In contrast, several studies indicate that more than half of all transmissions among MSM in the United States originate from men that were not retained in care (25–27). The estimated impact of particular prevention strategies in Fig. 4 is limited to settings with a similar epidemic profile and care cascade as the Netherlands (Table 1).

Third, not more than an estimated 20% of infections in the cohort could have been averted between mid-2008 and December 2010 with immediate ART after diagnosis. Given the remarkable expansion of ART coverage in the Netherlands in the past (9), the prevention potential of immediate ART is now limited. Nonetheless, starting ART at a cell count above 500 cells/ml leads to improved clinical outcomes and remains a priority (28).

Fourth, and similar to other locations (24, 29), almost half of all infections in our transmission cohort originated from men in their first year of infection. Frequent early transmission limits the overall impact of annual
testing plus immediate ART to those testing positive (Fig. 4), thus implying that prevention services to uninfected MSM must be strengthened. The substantial, estimated impact that PrEP would have had in averting transmissions in our cohort (Fig. 4) supports making PrEP available to MSM testing negative as in the United States (30). Recent PrEP demonstration projects (31, 32) indicate that existing barriers such as low awareness (33) and a lack of experience among providers (34) can be addressed. Concerns regarding the toxicity of PrEP, increasing sexual risk behavior, and emerging drug resistance have to date not been substantiated since PrEP was made available in the United States (35). In the context of PrEP-experienced prevention services, high discontinuation rates after PrEP initiation appear to be the greatest challenge to maintain protection from infection (31).

Fifth, without substantial increases in current annual testing coverage, ART and PrEP offered along the revised 2016 WHO guidelines could not have prevented more than a third of all infections in our transmission cohort. Because phylogenetically probable transmitters tend to test much less frequently than the average diagnosed MSM, substantial barriers likely exist in reaching men at high risk of onward transmission, and further work is needed to characterize these (36). Strategies such as self-testing (37), community-based testing (38), and more provider-initiated routine testing in general practices and at medical admissions raised annual testing coverage in pilot projects (39) and need to be expanded alongside biomedical interventions.

Sixth, this study indicates that substantial reductions in HIV incidence among MSM could have been realized with a combination approach that includes—critically—increased annual testing, with uptake of PrEP by young MSM testing negative and provision of immediate ART to those testing positive. This finding is primarily based on the impact of increased annual testing and the higher efficacy of PrEP reported in two recent randomized controlled trials (20, 21), and updates previous studies that estimate more limited benefits (4, 40, 41). Beyond age at testing, other characteristics not available to this study may also indicate high infection risk (42) and thereby identify groups of MSM to which PrEP should be made available as a priority too. Provision of PrEP to all men testing negative is not affordable at current drug prices in high-income countries (40). The magnitude of the predicted impact of test-and-PrEP-and-treat for all (Fig. 4) could set an aspirational target for the fight against HIV among MSM.

The lack of substantial reductions in incidence among Dutch MSM is not a result of ineffective ART provision or inadequate retention in care. New HIV infections among MSM are challenging to prevent because of frequent early transmission and continued low testing uptake of men at risk of transmission. Counterfactual prevention scenarios on phylogenetically reconstructed, past transmission events to MSM with recent infection at diagnosis predict that increased annual testing and uptake of PrEP by men at high risk of infection have a key role to send the HIV epidemic among MSM into a decisive decline.

MATERIALS AND METHODS

Study design

We conducted a retrospective viral phylogenetic transmission and prevention study that focuses on transmissions to MSM that were confirmed to have recent HIV infection at time of diagnosis in the Netherlands (Fig. 1). The prespecified objectives were to, first, reconstruct past, phylogenetically probable transmission events to these recipient MSM; second, to estimate the proportion of transmissions originating throughout the infection and care continuum based on the reconstructed transmission events; and, third, to estimate the proportion of infections that could have been averted through reallocating past, probable transmitters to less infectious stages in counterfactual modeling scenarios.

The ATHENA national observational HIV cohort includes anonymized data of all HIV-infected patients followed longitudinally in the 27 HIV treatment centers in the Netherlands since 1996, except 1.5% who opt-out (9). ATHENA patients are informed of data collection by their treating physician and can refuse further collection of clinical data according to an opt-out procedure. Patients who were diagnosed between 1981 and 1995 were included in the cohort when they were still alive in 1996 (9). Demographic, clinical, and viral sequence data were collected at entry and follow-up visits as described previously (9). By March 2013, viral sequence data had been systematically entered until December 2010. Therefore, recipients were enrolled between early 1996 and December 2010. Potential transmitters were enrolled until database closure in March 2013. Table S1 characterizes the demographic, clinical, and viral sequence data that were used in this study. The resolution of the infection/care stages in Table 2 was adjusted to ensure adequate sample sizes. The number of probable transmission intervals after first viral suppression was too small to enable further stratification by treatment class. This study was reviewed and approved by the HIV Monitoring Institutional Data Access and Ethics Committee, and reported along STROKE-ID guidelines.

Viral sequences of different subtypes (n = 355 from MSM), with less than 250 nucleotides (n = 368) or indication for intra-subtype recombination (n = 52), were removed before analysis. Primary drug resistance mutations were masked in each sequence (43). Demographic and clinical data were checked for consistency along patient timelines and to lie within appropriate ranges. Outliers were reported to the ATHENA quality control team and manually updated.

Recently infected, recipient MSM and infection windows

We enrolled as recipients all MSM for whom a narrow infection window could be identified. MSM had evidence of infection within 12 months before diagnosis if either 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 was reported. Figure S1 shows enrollment progress over time. Infection windows were at most 12 months or shorter if indicated by a last negative HIV antibody test (fig. S2).

Potential transmitters to recipient MSM

We enrolled as potential transmitters all registered infected men that overlapped with infection windows of recipients and thus could have in principle infected a recipient. This definition required estimation of putative infection times. Calculations are based on a method by Rice and colleagues (44) (see the Supplementary Materials). Estimated infection times are associated with substantial uncertainty, and sensitivity analyses were conducted for lower and upper 95% estimates. Table S2 characterizes the potential transmitters to all recipients. Further analysis was restricted to potential transmission pairs with sequences from both individuals (stage B in Fig. 1).

Viral phylogenetic exclusion analysis to construct the transmission cohort

The viral phylogeny was reconstructed under the GTR nucleotide substitution model with maximum-likelihood methods (45) and is
shown in fig. S3. Five hundred bootstrap trees were created to quantify uncertainty in tree reconstruction (10). Genetic distances between sequences from transmitter-recipient pairs were highly variable (fig. S4), which was accounted for in all analyses. To guide our choice of exclusion criteria, we considered, first, epidemiologically confirmed transmission pairs from previously published transmission chains in Belgium and Sweden (15, 46). The Belgium transmission chain was subsequently oversampled (14), providing 2807 sequence pairs from confirmed transmitters and recipients without multidrug resistance. Further, we considered 4117 pairs of sequences from the same Dutch patient and 201,605 pairs between Dutch patients who died before the last negative antibody test of another patient. These pairs were used to quantify patterns of viral evolutionary diversification that can be expected among confirmed linked and unlinked pairs, and to develop exclusion criteria with high specificity (see the Supplementary Materials). The Swedish pairs were used for validation purposes. All potential transmitters that were not excluded were considered phylogenetically probable and are characterized in table S4.

Relative pairwise transmission probabilities
Among the 2807 confirmed transmission pairs (14), the genetic distance between sequences from the transmitter and the recipient was strongly associated with the time elapsed between both sampling dates and the midpoint of the established infection window (fig. S5). We fitted a probabilistic molecular clock model to these data to describe the relative probability of observing a given genetic distance between sequences from a transmission pair that diverged for a specified amount of time from each other. The fitted model was then used to express the relative probability that a phylogenetically identified transmitter was the actual transmitter to a recipient (fig. S5).

Matching of clinical data to associate infection/care stages with transmission intervals
Sources of transmission were not defined in terms of individuals but by the 14 stages in the infection and care continuum in Table 2 (stage F in Fig. 1). Stages were allocated to transmission intervals on the basis of available clinical data (table S1). The duration of transmission intervals was set to 6 weeks to accommodate abrupt changes in infection/care stages.

Adjusting for censoring and sequence sampling biases
Toward the present, an increasing fraction of potential transmitters may not have been diagnosed by the time of database closure. Potential transmitters with recent infection at time of diagnosis must, by definition, have been diagnosed within 12 months after the putative transmission interval. Therefore, the extent of right censoring differs between stages. To adjust for right censoring, we counted when potential transmitters in a particular infection/care stage became diagnosed in relation to the time of diagnosis of their recipient (fig. S6). This enabled us to estimate the proportion of censored intervals for a hypothetical database closure time in the past (fig. S6). We then extrapolated these estimates to the actual database closure time with a bootstrap algorithm (see the Supplementary Materials). To quantify sequence sampling biases, we compared men with and without a sequence in the near-complete population cohort (fig. S7). A negative binomial missing data model was then used to adjust for the number of missing transmission intervals. Adjustments accounted for censoring; increasing sampling frequency with duration in care; high sampling frequency of men returning to care, men participating in particular substudies, and men with indication of drug resistance; as well as increasing sampling frequency with calendar time (fig. S7).

Epidemiological transmission analysis
Each interval was associated with a phylogenetic transmission probability (stage F in Fig. 1). The relative pairwise transmission probabilities (fig. S5) were equally apportioned to all observed intervals of the same transmitter-recipient pair. Stage-specific data such as viral load were not used to determine these probabilities to avoid circularity in the attribution of transmissions to infection/care stages. Then, the transmission probability in an observed interval \( \tau \) from transmitter \( i \) to recipient \( j \) was calculated by

\[
p_{ij \tau} = \omega_{ij \tau} / \sum_{k \neq j} \omega_{ik \tau} + \sum_{z} m_{z}(z) \omega_{jz},
\]

where \( \omega_{ij \tau} \) is the relative transmission probability in interval \( \tau \), and the denominator sums over all observed, competing intervals as well as expected missing intervals \( m_{z}(z) \) in stage \( z \) to recipient \( j \). For missing intervals, relative transmission probabilities were imputed and set to the median \( \omega_{ij \tau} \) of all observed intervals \( s \) in stage \( z \), denoted by \( \omega_{s} \). For a missing transmission interval \( \nu \) in stage \( x \) to recipient \( j \), we calculated

\[
p_{j \nu} = \omega_{j \nu} / \sum_{k \neq j} \omega_{ik \nu} + \sum_{z} m_{z}(z) \omega_{jz},
\]

where \( \omega_{j \nu} \) is equal to 0.5 if \( i \) and \( j \) are each other’s phylogenetically probable transmitters, otherwise 1; \( \omega_{ij \tau} \) are the relative pairwise transmission probabilities calculated by

\[
\omega_{ij \tau} = \omega_{ij} / \tau_{ij},
\]

where \( \omega_{ij} \) is the number of recipients with date of diagnosis in \([x, \tau_{ij}]\) and \( \tau_{ij} \) is the number of transmission intervals between transmitter \( i \) and recipient \( j \). These probabilities sum to 1 per recipient. If all transmitters are sampled, we obtain \( p_{ij \tau} = \omega_{ij \tau}/\sum_{s} \omega_{is \tau} \). If some transmitters are not sampled, the first part of the denominator, \( \sum_{s} \omega_{is \tau} \), is smaller and adjusted by the second part of the denominator. The number of expected missing intervals \( m_{z}(z) \) differs by stage and adjusts for stage-specific censoring and sampling biases.

The proportion of transmissions originating from the 14 infection/care stages was obtained by summing the corresponding individual-level transmission probabilities (fig. S8). Precisely, the proportion of transmissions from stage \( x \) to recipients diagnosed in \([t_{1}, t_{2}]\) was calculated by

\[
P^{T}(x, t_{1}, t_{2}) = \sum_{j \in R(t_{1}, t_{2})} \frac{p_{j}(x)}{\sum_{z \in R(t_{1}, t_{2})} p_{j}(z)} = \frac{1}{J} \sum_{j \in R(t_{1}, t_{2})} p_{j}(x),
\]

where \( R(t_{1}, t_{2}) \) is the set of recipients with date of diagnosis in \([t_{1}, t_{2}]\), \( J \) is the number of recipients with date of diagnosis in \([t_{1}, t_{2}]\), and \( p_{j}(x) \) is the probability that recipient \( j \) was infected by a transmitter in stage \( x \). The probability \( p_{j}(x) \) is the sum

\[
p_{j}(x) = \sum_{i} \sum_{z \neq j} p_{ij}(x) + \sum_{w} m_{w}(x) p_{wj},
\]

where \( I_{j} \) are the observed, phylogenetically probable transmitters to recipient \( j \); \( V_{j}(x) \) is the set of observed transmission intervals between \( i \) and \( j \) in stage \( x \); and all other quantities are as defined above. The formula for \( P^{T}(x, t_{1}, t_{2}) \) can be intuitively interpreted as the average
probability that a recipient was infected by a transmitter in stage \( x \). Thus, the precision in the estimated \( P^H(x, t_1, t_2) \) depends primarily on the number of recipients. We identified substantial individual-level variation in the transmission probabilities \( p(x) \) (fig. S8), suggesting that a relatively large number of past transmission events are needed to reliably quantify sources of transmission.

**Epidemiological prevention analysis**

With the sources of transmission estimated, we compared the impact of prevention strategies in counterfactual scenarios that modeled the redistribution of phylogenetically identified transmitters to less infectious stages in the HIV infection and care continuum. This reduced the overall probability that any of the recipients would have been infected to less than 1. The proportion of infections that could have been averted in the period \([t_1, t_2]\) with a counterfactual prevention scenario \( H \)

\[
a(H) = 1 - \sum_{j \in A(t_1,t_2)} \sum_{x} p^H_j(x) \]

where \( p^H_j(x) \) is the probability that recipient \( j \) is infected by someone in stage \( x \) under the counterfactual prevention scenario \( H \). The individual-level prevention models are described in the Supplementary Materials.

**Statistical uncertainty**

Central estimates of \( P^H(x, t_1, t_2) \) and \( a(H) \) were obtained under central estimates of the genetic distances in fig. S4, the resulting phylogenetic transmission probabilities \( \omega_{ij} \), and the expected number of missing transmission intervals (Fig. 2C). Bootstrap sampling of the recipients, the empirical distribution of genetic distances, the number of missing transmission intervals under a negative binomial missing data model, and the counterfactual reallocation procedure of probable transmitters to less infectious infection/care stages were conducted to obtain nonparametric 95% confidence intervals. Confidence intervals were based on 1000 bootstrap replicates.

**SUPPLEMENTARY MATERIALS**

www.sciencetranslationalmedicine.org/cgi/content/full/8/320/320a2/DC1

Extended acknowledgments

Materials and Methods

Fig. S1. Number of identified recipient MSM by 3-month intervals.

Fig. S2. Duration of infection windows of recipient MSM.

Fig. S3. Snapshot of the reconstructed viral phylogeny.

Fig. S4. Uncertainty in the estimated genetic distance between sequences from the transmitter and recipient of potential transmission pairs.

Fig. S5. Genetic distance between sequence pairs from previously published, epidemiologically confirmed transmitter-recipient pairs, and sequence pairs from the phylogenetically probable transmission pair in this study.

Fig. S6. Right censoring at past, hypothetical database closure times.

Fig. S7. Sequence sampling probabilities by stage in the infection and care continuum.

Fig. S8. Individual-level variation in phylogenetically derived transmission probabilities by infection/care stages.

Fig. S9. Frequency of infection/care stages among phylogenetically probable transmitters.

Fig. S10. Phylogenetically derived transmission probabilities of observed transmission intervals.

Fig. S11. Transmission risk ratio from men after ART start compared to diagnosed untreated men with CD4 >500 cells/ml.

Fig. S12. Sensitivity analysis on the impact of PrEP with lower efficacy.

Fig. S13. Sensitivity analysis on the impact of lower or higher PrEP coverage.

Fig. S14. Impact of sampling and censoring adjustments on the estimated proportion of transmissions from stages in the infection and care continuum.

Fig. S15. Impact of phylogenetic transmission probabilities on the estimated proportion of transmissions from stages in the infection and care continuum.

**REFERENCES AND NOTES**


Acknowledgments: We thank the Imperial College High Performance Computing Service (www3.imperial.ac.uk/ict/services/hpc), three anonymous referees, the HIV treating physicians, and HIV nurse consultants and staff of the diagnostic laboratories and facilities in the HIV treatment centers, along with the data collecting and monitoring staff both within and outside the Stichting HIV Monitoring Foundation for their contributions to make this work possible (see Extended Acknowledgments in the Supplementary Materials). Funding: O.R. is supported by the Wellcome Trust (fellowship WR092311MF), and C.F. by the European Research Council (Advanced Grant PBDR-339251) and the Bill & Melinda Gates Foundation (PANGEA-HIV consortium). P.R. through his institution received independent scientific grant support from Bristol-Myers Squibb, ViiV Healthcare, Gilead Sciences Inc., and Merck & Co; served on a scientific advisory board for Gilead Sciences; and serves on a data safety monitoring committee for Janssen Pharmaceuticals Inc., for which his institution has received remuneration. The ATHENA observational cohort study is part of Stichting HIV Monitoring and supported by a grant from the Netherlands Ministry of Health, Welfare and Sport through its Centre for Infectious Disease Control–National Institute for Public Health and the Environment. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Author contributions: O.R., F.d.W., P.R., and C.F. conceived the study. O.R. and C.F. developed the methods, did the analysis, and reviewed all statistical aspects of the analysis. A.v.S., D.B., S.J., and A.W. provided the data used in the analysis. A.G. assisted in estimating the viral phylogeny. A.v.S., D.B., F.d.W., and P.R. gave advice on analysis and interpretation. O.R. and C.F. wrote the first draft. All authors reviewed and approved the final version. Competing interests: The authors declared that they have no competing interests. Data and materials availability: Data are available from the HIV Monitoring Institutional Data Access/Ethics Committee for researchers who meet the criteria for access to confidential data. Contact E-mail: secretariaat.shm@amc.uva.nl.

Submitted 5 August 2015
Accepted 3 December 2015
Published 6 January 2016
10.1126/scitranslmed.aad1863

Sources of HIV infection among men having sex with men and implications for prevention

Oliver Ratmann, Ard van Sighem, Daniela Bezemer, Alexandra Gavryushkina, Suzanne Jurriaans, Annemarie Wensing, Frank de Wolf, Peter Reiss, Christophe Fraser and ATHENA observational cohort

Sci Transl Med 8, 320ra2320ra2.
DOI: 10.1126/scitranslmed.aad1863

The ART of HIV prevention

Despite the relative success of antiretroviral therapy (ART) for individuals infected with HIV, the rate of new diagnoses has remained fairly constant in vulnerable population groups, particularly men having sex with men (MSM). Now, ART is also available in the United States to uninfected individuals to directly prevent infection with the virus. Ratmann et al. were able to reconstruct ~600 past transmission events among men having sex with men in the Netherlands, and examined probable sources of transmission. They found that the large majority of new infections is neither attributable to ineffective ART nor inadequate retention in care. Rather, many of these cases could have been averted with more comprehensive HIV testing and a broader use of ART that includes provision to uninfected men as well as starting ART as soon as possible among newly diagnosed men. These findings support making ART for pre-exposure prophylaxis available worldwide, and especially in countries with high retention in care and high ART coverage among infected MSM.