

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Ragonnet-Cronin, M; Hué, S; Hodcroft, EB; Tostevin, A; Dunn, D; Fawcett, T; Pozniak, A; Brown, AE; Delpech, V; Brown, AJL; +1 more... UK HIV Drug Resistance Database; (2018) Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *The lancet HIV*, 5 (6). e309-e316. ISSN 2405-4704 DOI: [https://doi.org/10.1016/S2352-3018\(18\)30062-6](https://doi.org/10.1016/S2352-3018(18)30062-6)

Downloaded from: <http://researchonline.lshtm.ac.uk/4648184/>

DOI: [https://doi.org/10.1016/S2352-3018\(18\)30062-6](https://doi.org/10.1016/S2352-3018(18)30062-6)

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

<https://researchonline.lshtm.ac.uk>

1 Nondisclosed MSM in UK HIV
2 transmission networks: phylogenetic
3 analysis of surveillance data

Manon Ragonnet-Cronin, Stéphane Hué, Emma B. Hodcroft, Anna Tostevin, David
Dunn, Tracy Fawcett, Anton Pozniak, Alison E. Brown, Valerie Delpech and Andrew J.
Leigh Brown on behalf of the UK HIV Drug Resistance Database

4 [Addresses](#)

5 Manon Ragonnet-Cronin, PhD
6 University of California, San Diego, Department of Medicine
7 220 Dickinson street, San Diego, CA 92103, USA
8

9 Stéphane Hué, PhD
10 London School of Hygiene & Tropical medicine, Department of Infectious Disease Epidemiology
11 Keppel Street, London WC1E 7HT, UK
12

13 Emma B. Hodcroft, PhD
14 University of Edinburgh, Institute of Evolutionary Biology,
15 Ashworth Laboratories, West Mains Road, Edinburgh EH9 3JT, UK
16

17 Anna Tostevin, MSc
18 University College London, Institute for Global Health
19 Mortimer Market Centre, Capper Street, London WC1E 6JB, UK
20

21 Professor David Dunn, PhD
22 University College London, Institute for Global Health
23 Mortimer Market Centre, Capper Street, London WC1E 6JB, UK
24

25 Tracy Fawcett, MSc
26 Virology, Old Medical School, Leeds General Infirmary, Leeds LS1 3EX, UK
27

28 Anton Pozniak, MD
29 Chelsea and Westminster Hospital, 369 Fulham Road, London, SW10 9NH, UK
30

31 Alison E. Brown, PhD
32 Public Health England, Wellington House, 133-155 Waterloo Road, London SE1 8UG, UK
33

34 Valerie Delpech, PhD
35 Public Health England, Wellington House, 133-155 Waterloo Road, London SE1 8UG, UK
36

37 Professor Andrew J. Leigh Brown, PhD
38 University of Edinburgh, Institute of Evolutionary Biology,
39 Ashworth Laboratories, West Mains Road, Edinburgh EH9 3JT, UK
40

1 Research in context

2 Evidence before this study

3 A search on PubMed with the terms “genetic”, “cluster”, “HIV”, “epidemiology” and “network” in
4 January 2017 returned 25 articles. We focused our attention on analyses of densely sampled
5 epidemics in Europe and North America. The present UK epidemic among men who have sex with
6 men (MSM) arose through the introduction of subtype B viruses in the early 1980s. In contrast, in
7 heterosexuals, the majority of viruses are non-B subtypes. Within MSM transmission networks,
8 there is a subgroup of men who self-report as heterosexual but whose HIV genetic sequences link
9 only to MSM. This group is very poorly understood: they are inaccessible to classical epidemiological
10 approaches and were only revealed in the UK recently, in a single molecular epidemiology study.

11 Added value of this study

12 We focused on this group of self-reported heterosexual men who link only to MSM adopting a novel
13 network based method which allowed us investigate transmission across multiple subtypes.
14 Analysing viral sequences from over 50,000 individuals, we identified men who self-report as
15 heterosexual in MSM HIV transmission networks (potential nondisclosed MSM), examined their
16 position and compared them to the MSM with whom they clustered. We found that these men
17 tended to link to each other and were not involved in rapid chains of MSM transmission. They were
18 more likely to be non-White than the MSM they clustered with. We found more examples of
19 heterosexual men linking heterosexuals and MSM than heterosexual women in that position,
20 although the total number of such chains was small.

21 Implications of all the available evidence

22 The tendency of nondisclosed MSM in our network to link to each other indicates that they may
23 preferentially seek out other nondisclosed MSM partners for sex. This result aligns with conclusions
24 from small scale interview-based studies, which found that they attend different venues to MSM.
25 The current study is the first to characterise the behaviour of nondisclosed MSM at a national scale.
26 Our results suggest that nondisclosed MSM may be underestimating their risk and should be
27 encouraged to test more frequently with better targeted prevention messages. However, it should
28 be emphasised that the extent of bridging between the heterosexual and MSM epidemics appears to
29 be limited.

30 Summary

31 Background

32 Patients who do not disclose their sexuality, for example men who do not disclose same-sex
33 behaviour, are extremely difficult to characterise through traditional epidemiological approaches.
34 Using a recently developed method to detect large networks of viral sequences from time-resolved
35 trees, we have localised these men in UK transmission networks, gaining critical insights into the
36 behaviour of this group.

37 Methods

38 We obtained HIV *pol* sequences from the UK HIV Drug Resistance Database (UKRDB), a central
39 repository for resistance tests performed as part of routine clinical care throughout the UK.
40 Sequence data are linked to demographic and clinical data held by the UK Collaborative HIV Cohort study
41 and the national HIV/AIDS Reporting System database. Initially we reconstructed maximum likelihood
42 phylogenies from sequences from over 50,000 individuals in the UK. Sequences were selected for
43 time-resolved analysis in BEAST if they were clustered with at least one other sequence at a genetic

1 distance $\leq 4.5\%$ with support $\geq 90\%$. We used time-resolved phylogenies to create networks by
2 linking together nodes if sequences shared a common ancestor within the previous 5 years. We
3 identified potential nondisclosed MSM (pnMSM) as self-reported heterosexual men who clustered
4 only with men. We measured the network position of pnMSM, including betweenness (a measure of
5 connectedness and importance) and assortativity ((the propensity for nodes sharing attributes to
6 link).

7 Findings

8 14,405 individuals were represented in the network, including 8,452 MSM, 1743 heterosexual
9 women and 1341 male heterosexuals. 249 pnMSM were identified (18.6% of all clustered
10 heterosexual men). pnMSM were more likely to be Black-African ($p < 0.0001$) and were slightly older
11 (39.00 vs. 36.38 $p = 0.002$) than the MSM they clustered with. Betweenness centrality was lower for
12 pnMSM than for MSM (1.31 vs 2.24, $p = 0.002$), indicating that they were in peripheral positions in
13 MSM clusters. Assortativity by risk group was higher than expected (0.037 vs - 0.037, $p = 0.01$)
14 signifying that pnMSM linked to each other. We found that self-reported male heterosexuals were
15 more likely than heterosexual women to link MSM and heterosexuals (Fisher's exact test; $p = 0.0004$;
16 OR 2.24) but the number of such transmission chains was small (only 54 in total compared to 32 in
17 women).

18 Interpretation

19 We have shown that pnMSM are a subgroup distinct from both MSM and from male heterosexuals.
20 Male heterosexuals are the group most likely to be diagnosed in late stage disease, and nondisclosed
21 MSM may put female partners at higher risk. Thus, pnMSM require specific consideration to ensure
22 they are included in public health interventions.

23 Funding

24 NIH GM110749 (MIDAS).

25 Introduction

26 Men who have sex with men (MSM) are the group at highest risk for HIV infection, with higher HIV
27 prevalence than the general population in every world region (1). In the UK, MSM account for half of
28 those living with HIV despite increased testing and uptake of treatment (2).

29 Among MSM, those who do not disclose their same-sex behaviour for fear of stigmatisation,
30 rejection or prejudice (3) are more likely to exhibit risky sexual behaviour, for example not using
31 condoms (4), and are more likely to have sex with women (5). Insights into the behaviour of
32 nondisclosed MSM are difficult to obtain but central to our understanding of HIV transmission.

33 Phylogenetic analysis of viral sequences has provided a valuable route to elucidating transmission
34 dynamics (6). Phylogenies can be combined with epidemiological, demographic and behavioural
35 data to elucidate spread of HIV through time, geographical regions and different risk groups (7).
36 Phylogenies have been used to investigate the role of age, disease stage, viral load and treatment on
37 transmission (8).

38 In the UK, analysis of subtype B viruses demonstrated the existence of six independent transmission
39 chains among MSM following separate introductions in the 1980s (9). MSM were much more likely
40 than heterosexuals to fall within groups of closely linked infections, or clusters (6, 10). Within
41 clusters, 25% of transmissions happen within six months of infection among MSM (6), compared to
42 2% in heterosexuals (10). In a recent analysis we used phylogenetics to quantify the extent of
43 nondisclosure of sex with men among heterosexual men (11). Eleven percent of heterosexual men

1 clustered only with MSM compared to only 5% of heterosexual women. This discrepancy suggests
2 around 6% (1-11%) of men who report only heterosexual behaviour became infected through sex
3 with men. The difference was most significant among Black African men, where up to 21% may have
4 been infected by men.

5 The structure of contact networks provides information on the process that led to their formation
6 (12). The transmission network is a subset of the contact network whose structure is shaped by the
7 underlying contact network and by the transmission characteristics of the pathogen (13). Thus, an
8 isolated node in an HIV transmission network may represent an individual who had no sexual
9 contacts after infection or one who had many sexual contacts but never transmitted the virus.
10 Nonetheless, aggregate differences in network characteristics between groups can reveal
11 differences in behaviour: MSM clusters are larger than heterosexual clusters (6, 10), indicating on
12 average more sexual partners and more onward transmission events.

13 Here we have further exploited our network reconstruction methodology (14) in order to describe
14 the network positions of nondisclosed MSM for the first time. We hypothesised that their position
15 within network clusters would differ from that of the MSM with whom they cluster, revealing
16 differences in sexual or transmission behaviour.

17 Methods

18 1. Study population

19 We obtained 63,065 aligned HIV *pol* sequences, each representing a distinct HIV positive individual,
20 from the UK HIV Drug Resistance Database (UKRDB; <http://www.hivrdb.org>; sequences up to mid-
21 2013). The UKRDB was established in 2001 as a central repository for resistance tests performed as
22 part of routine clinical care throughout the UK. Clinical centres that contribute to the UKRDB are
23 listed in the Appendix, p5. Sequences covering protease and 900 bases of reverse transcriptase are
24 received from laboratories. Sequences are aligned by the database managers at UKRDB through the
25 Stanford University Genotypic Resistance Interpretation Algorithm (hivdb.stanford.edu). Sequence
26 data are linked to demographic and clinical patient data held by the UK Collaborative HIV Cohort
27 study (15) and the national HIV/AIDS Reporting System database held at Public Health England (16).
28 After linking sequences to age, sex, risk group, ethnicity, avidity score and stage of infection, the
29 data were anonymised. Infections are classified as recent (<155 days) if their avidity is <80%. UKRDB
30 data collection and processing is explained on the UKRDB website ([http://www.hivrdb.org.uk/data-
31 collection-and-processing](http://www.hivrdb.org.uk/data-collection-and-processing)). We assigned subtypes using SCUEAL (17). The three most common
32 subtypes, B (49.5%), C (24.9%), and A1 (4.5%), were analysed, comprising 50,025 individuals
33 diagnosed between 1997 and 2013, for each of whom we used the earliest available sequence. The
34 London Multicentre Research Ethics Committee gave ethical approval for non-consented studies of
35 anonymised data (MREC/01/2/10).

36 In parallel, we retrieved over 130,000 *pol* (HXB2 positions 2253-3870, minimum length 900bp)
37 sequences of HIV subtypes A1, B, and C, from the Los Alamos National Laboratory (LANL) database
38 (January 2014). To limit the size of alignments, we selected the ten closest non-UK sequences to
39 each UK sequence. All sequences were stripped of 44 sites associated with drug resistance (18).

40 2. Phylogenetic analysis

41 For each subtype, we reconstructed phylogenies in RaxMLv8.2 with 100 bootstrap replicates (19).
42 We considered 100 bootstraps to be sufficient because the maximum likelihood phylogeny was the
43 first step towards deciding whether to include sequences in a time-resolved analysis. We identified
44 clusters in phylogenies using Cluster Picker v1.2.3 if they were supported by a bootstrap $\geq 90\%$ and

1 had a maximum genetic distance $\leq 4.5\%$ (6), with gaps ignored in pairwise comparisons and
2 ambiguities counted as mismatches. The phylogenetic relationships between clustered sequences
3 were then time-resolved in BEASTv1.8 (20).

4 We sorted previously identified clusters into pools of up to 300 sequences so that all sequences from
5 the same cluster were analysed together (14), but limiting the size of datasets. Seventy-three pools
6 were created altogether. For each pool, we generated summary statistics, including the number of
7 sequences, the number of (RaxML-identified) clusters and the breadth of years covered (the
8 difference between the oldest and the youngest sequence). We tested settings in two pools each of
9 subtype B and subtype C sequences with maximum breadth of sequence dates (12 years). We
10 compared clock (strict and relaxed) and demographic models (constant, exponential, logistic,
11 Bayesian skyline and Bayesian skyride) in the test pools. We used an SRD06 substitution model as
12 codon models give the best fit to HIV *pol* sequence data (21). A normally distributed molecular clock
13 rate prior with a mean of 2.55×10^{-3} was used (9). The best model was selected as the skyride with
14 an uncorrelated lognormal clock by means of its Bayes Factor (the ratio of the marginal likelihoods of
15 two models), as estimated through path sampling. Chains were run for 100,000,000 generations in
16 duplicate, checked for convergence in Tracer v1.6 and combined. We generated maximum clade
17 credibility trees in Tree Annotator v1.8.4.

18 We reconstructed networks from phylogenies as previously described (14) using R scripts (available
19 at github.com/hxnx-sam/Rcode_utils/). In brief, each sequence is represented by a node in the
20 network and nodes are linked together (clustered) if their time to most recent common ancestor is
21 ≤ 5 years in the time-resolved maximum clade credibility tree. In contrast to clusters defined
22 phylogenetically, nodes in network clusters are not all connected to each other, and clusters can
23 contain sequences which are not monophyletic in the tree. Different depths of network
24 reconstruction were evaluated in a previous analysis (14), and here we selected five years as a cut-
25 off for consistency with the National Survey of Sexual Attitudes and Lifestyles (NATSAL) (22).

26 3. Procedures

27 We constructed a multivariate logistic regression model with the outcome being either clustered in
28 the network (1) or not (0). Subtype, sex, risk group and ethnicity were included as predictor
29 variables, and we included terms for the interactions between variables.

30 We conducted network analysis using the *igraph* v1.1 (23) and *network* v1.13 (24) packages in R
31 v3.4.1. Networks for all subtypes were amalgamated and nodes were linked to demographic data.
32 While the transmission dynamics of B and non-B subtypes are undoubtedly different in the UK (6,
33 10), subtype was maintained as a variable associated with each node so that results could be
34 interpreted by subtype.

35 We classified self-reported heterosexual men who clustered only with MSM and with no women as
36 potential nondisclosed MSM (pnMSM; see results) and we compared their positions within networks
37 to that of the MSM with whom they clustered based on their centrality and linkage (Figure 1).
38 Centrality measures (degree, betweenness centrality and eigenvector centrality) quantify how
39 important a node is in keeping a network connected. Degree is the number of edges connected to a
40 node. Betweenness centrality is the number of shortest paths linking every pair of nodes that go
41 through the node examined. Eigenvector centrality weighs the centrality of neighbouring nodes so
42 that a node with high eigenvector centrality may not have high degree itself but is central to the
43 structure of the network. Betweenness and eigenvector centrality measure how connected a node
44 is, and how important that node is to the network. Eigenvector centrality is more conservative than

1 degree and betweenness and is less affected by network and cluster size. Low centrality indicates
2 that a node is peripheral to a network (Figure 1A).

3 We evaluated the propensity for nodes to link to other nodes sharing their characteristics in order to
4 test whether pnMSM were more likely to link to each other than to MSM (Figure 1B). This was done
5 by calculating network assortativity as available in the *network* package (24). Assortativity varies
6 between -1 (completely disassortative) and 1 (completely assortative) and is influenced by the ratio
7 of node labels (here, MSM and HET/pnMSM). To ensure both assortative and disassortative links
8 were possible in each cluster, we analysed only clusters comprising at least two MSM and two male
9 self-reported heterosexuals. To generate a null distribution of expected assortativity given the
10 relative representation of MSM and pnMSM, we randomised risk group 10,000 times across the sub-
11 network and recalculated assortativity each time.

12 Finally, we tested whether pnMSM might link the MSM and heterosexual epidemics (Figure 1C) by
13 counting the number of relationships they brokered between MSM and heterosexual women and
14 comparing it to the number of relationships brokered by heterosexual women (Figure 1D). We
15 performed a Gould-Fernandez brokerage analysis available in R (25) to count occurrences in the
16 network of heterosexual men linking MSM and female heterosexuals, and compared this to the
17 incidence of females linking MSM and heterosexual men (Figure 1D). This analysis was performed on
18 the entire network as it required clusters in which females were also present.

19 4. Data sharing

20 As submission of the entire national cohort to public databases would permit transmission network
21 identification and risk breaching patient confidentiality, we have followed earlier practice (6, 10) and
22 submitted a random sample of 10% of each subtype to GenBank (accession numbers available in the
23 online Appendix p4). The UKRDB Steering Committee encourages analysis proposals. Instructions for
24 applying are available at <http://www.hivrd.org.uk/>.

25 5. Role of the funding source

26 The funders of the study had no role in study design, data collection, data analysis, data
27 interpretation, or writing of the report. The corresponding author had full access to all the data in
28 the study and had final responsibility for the decision to submit for publication.

29 Results

30 The earliest sequence analysed originated from an individual sampled in January 1997, the most
31 recent sequence analysed here originated from an individual diagnosed and sampled in July 2013
32 (Supplementary Figure 3, Appendix p2). Of 50,025 sequences analysed, 21,053 (42.1%) sequences
33 linked to at least one other at a genetic distance of 4.5% in the RaxML phylogeny. At a time-depth of
34 five years, 14,405 patients shared an ancestor within 5 years and were therefore represented in the
35 resulting network (28.8%, Table 1, Figure 2). Of 4,836 clusters, 3,344 (69.1%) comprised only two
36 sequences, and maximum cluster size was 87 (Supplementary Figure 1, Appendix p1).

37 Based on the logistic regression, MSM were more likely to cluster than heterosexuals (OR: 2.6,
38 $p < 0.0001$, Table 1), subtype B sequences were more likely to cluster than subtypes A1 and C (OR:
39 2.8, $p < 0.0001$) and white individuals were more clustered than Black Africans (OR: 1.9, $p = 0.0002$).
40 Some interactions between predictors were significant but effects were small (Supplementary Table
41 1, Appendix p3). Subtype B, MSM risk group and white ethnicity were independently predictive of
42 clustering even after their interactions were taken into account.

1 We analysed all clusters that contained one or more self-reported heterosexual men and one or
2 more MSM but no women. We describe these men as potential nondisclosed MSM (pnMSM). There
3 were 223 clusters comprising 955 MSM and 249 self-identified heterosexual men (Figure 3). These
4 pnMSM represented 18.6% of a total of 1341 clustered self-reported heterosexual men. As a
5 comparison, 131/1743 (7.5%) of clustered heterosexual women clustered only with MSM. In the
6 present dataset, 22,296/25,492 (87.5%) MSM were infected with subtype B, while for other risk
7 groups, only 9,353/37,671 (24.8%) were infected with subtype B (Fisher's exact t-test, $p < 0.0001$).

8 We compared pnMSM to the MSM with whom they clustered: pnMSM were more likely to be Black
9 African, less likely to be infected with subtype B and were slightly older than the MSM they clustered
10 with (Table 2). There was no difference in the proportion of recent infections between the two
11 groups.

12 We wanted to test whether the position of pnMSM and MSM in the network differed because a
13 difference in transmission network characteristics is indicative of a difference in underlying
14 transmission dynamics. In clusters comprising MSM and pnMSM, betweenness (1.31 vs 2.24) and
15 eigenvector centrality (0.50 vs 0.52) were significantly lower for pnMSM than for the MSM they
16 clustered with (paired Wilcoxon signed rank test, $p = 0.002$ and $p = 0.039$, respectively; Table 3),
17 indicating that pnMSM were in peripheral positions in clusters. However, mean degree was not
18 different between MSM and pnMSM (paired Wilcoxon signed rank test, 3.02 vs 2.83, $p = 0.07$) and so
19 pnMSM were not less connected overall than MSM.

20 We then investigated whether pnMSM might be more likely to link to each other in clusters,
21 indicating that pnMSM might be more likely to partner with and be infected by other pnMSM. There
22 were 26 clusters containing at least two MSM and at least two self-reported heterosexual men,
23 comprising 310 MSM and 87 heterosexual men. Among the 26 clusters in the true network, the
24 assortativity coefficient was 0.037, and an assortativity coefficient higher than this was seen in only
25 110/10,000 of the permutations ($p = 0.011$; simulation mean: -0.037; simulation 95% CI -0.004 - -
26 0.034; Supplementary Figure 2, Appendix p2), indicating that pnMSM linked to each other.

27 Finally, we tested whether pnMSM might link the MSM and heterosexual epidemics. We found that
28 self-reported heterosexual men were more likely to broker relationships between MSM and
29 heterosexuals (Fisher's exact test, $p = 0.0004$; OR 2.24 95%CI 1.41-3.61) than heterosexual women
30 but numbers were small for both categories. Of 1341 self-reported heterosexual men represented
31 by nodes in the network, 54 (4.0%) were located in between an MSM and a heterosexual of the
32 opposite gender. This was true for only 32/1711 (1.9%) of heterosexual women.

33 Discussion

34 We identified potential nondisclosed MSM in the UK epidemic and describe their position in
35 transmission networks. In contrast to previous phylogenetic analyses, which were limited to a single
36 subtype at a time (6, 10, 14), the network-based approach allowed for analysis of multiple subtypes
37 concurrently, leading to a substantial benefit in power. We found that pnMSM were more likely to
38 be in peripheral positions in clusters and to link to each other. pnMSM were more likely to link MSM
39 and heterosexuals of the opposite gender than were heterosexual women. Taken together, these
40 differences in the reconstructed transmission network suggest differences in the transmission
41 dynamics of pnMSM as compared to both MSM and heterosexual men.

42 A limitation of this study was that the reconstruction method creates a network with more links than
43 the true transmission network because the order of transmissions is difficult to resolve when
44 successive infections occur close together in time. Sexual networks in which individuals became

1 infected within a short time will appear as fully connected clusters. For this reason, degree here is
2 not solely related to the number of onward transmissions, but is also a measure of the speed of
3 transmission within a group. In this context, the equivalent degree but peripheral positions of
4 pnMSM suggest that they are involved in closely linked MSM sexual networks but transmit less
5 between these sexual networks, and probably transmit less overall. These differences might reflect
6 differences in the contact networks of pnMSM or differences in transmission. One explanation might
7 be that they have a lower partner exchange rate, but another possibility is that they use condoms
8 more consistently or have sex with lower risk partners. Linking to each other indicates that they may
9 not find sexual partners in the same venues as MSM and/or seek out other nondisclosed MSM for
10 sex. This finding is supported by interviews, with nondisclosed MSM in the USA (26) and men who
11 have sex with men and women (MSMW) in the UK (22) reporting avoiding venues considered too
12 public or openly gay, and is important because prevention messages displayed in these venues will
13 not reach them. MSMW in NATSAL were less likely than MSM to visit sexual health clinics and to
14 receive prevention there, despite similar rates of sexually transmitted infections, and nearly 50% had
15 never been to a gay club (22). The Gay Men's Sex Survey 2014 similarly found that bisexual and non-
16 gay identified men were 2.5 times less likely to have ever been tested for HIV (27).

17 If nondisclosed MSM have sex with women, they may put female partners at risk. We found that
18 self-reported heterosexual men more frequently linked MSM to heterosexual women than
19 heterosexual women linked MSM to heterosexual men. This finding is surprising as men who report
20 both male and female partners are classified as MSM in the database, so we expect linkage between
21 MSM and women. However, the number of such potential transmission chains was small, only 54 in
22 total (4% of all heterosexual men clustered in the network), in agreement with a model quantifying
23 the extent of bisexual bridging in large cities in the USA (28). In NATSAL data, MSMW reported
24 unprotected sex with men and women, but lower rates of unprotected sex with men than those
25 observed among MSM (22), which could explain why this sexual behaviour does not lead to more
26 transmissions.

27 The absence of women in clusters formed the basis of our procedure for identifying pnMSM. Our
28 inclusion of clusters containing multiple self-reported heterosexuals enabled us to test whether
29 pnMSM linked to each other. We did not analyse clusters involving only heterosexual men ($n > 200$)
30 because we cannot determine whether women are missing. Even among the clusters examined, we
31 cannot exclude the possibility that female intermediaries are missing. The higher frequency of non-B
32 subtypes, most common among heterosexuals, among pnMSM, could indicate that a proportion of
33 them were in fact infected heterosexually. However, given that MSM and women are both
34 diagnosed at much higher rates than heterosexual men (2), it is likely that our estimates of
35 parameters relating to pnMSM remain conservative. If the additional > 200 clusters containing only
36 heterosexual men and > 2500 clusters comprising only MSM had been included in the assortativity
37 analysis, assortativity based on self-reported risk group would have been stronger but even with our
38 conservative approach it was significant.

39 Based on the proportion of individuals diagnosed with recent infections, pnMSM were not
40 diagnosed later than MSM. In contrast, heterosexual men in the UK tend to be diagnosed later than
41 MSM (2) thus pnMSM may be testing more regularly than heterosexual men. However, the number
42 of individuals with recency test results was small, which may limit our ability to detect any
43 difference. We previously found that pnMSM were more likely to be Black African than other
44 heterosexuals (11) and here found that they were more likely to be Black African than the MSM they
45 cluster with. These results are in agreement with findings from interviews with Black gay men in the
46 UK, who are less likely to disclose their sexuality than White gay men (29).

1 Our results reinforce the need for prevention campaigns appropriate for nondisclosed MSM. They
2 will not be reached by messages displayed in traditional gay venues if they do not frequent them and
3 as a consequence may underestimate their own risk. Interventions that target risk behaviours
4 without focusing on disclosure will be more effective; however, encouraging patients to build
5 trusting relationships with their healthcare providers, in which they feel able to disclose their sexual
6 behaviour, could be a valuable focus for Public Health. In parallel, our findings support the
7 deployment of real-time HIV phylogenetic analyses for targeted public health responses (30). In this
8 case, identification of pnMSM based on their position in networks could help tailor partner services
9 interviews towards them, for example. Further investigation into the behaviour of nondisclosed
10 MSM in the UK, in collaboration with social scientists, should be carried out to better tailor health
11 promotion towards this group.

1 **Tables**

2 *Table 1: Risk group, sex, ethnicity and subtype of sequences clustered in the network as*
 3 *compared to the rest of the UK HIV database (UKdb).*

		UKDB	NODES IN NETWORK	OR (95% CI)	P
n		50025	14405 (28.8%)		
Risk group	HET	17072	3084 (18.1%)	Ref	
	MSM	23128	8452 (36.5%)	2.6 (1.6-4.3)	<0.0001
	Other/ NA	9825	2869 (28.8%)	0.9 (0.4-1.8)	
Sex	F	11461	1881 (16.4%)	Ref	
	M	31360	10305 (32.9%)	1.1 (0.8-1.4)	
Ethnicity	Black African	12136	1701 (14.0%)	Ref	
	White	25092	8745 (34.9%)	1.9 (1.3-2.6)	0.0002
	Other/ NA	12797	3959 (31.0%)	2.2 (1.5-3.2)	<0.0001
Subtype	A1	2512	405 (16.1%)	Ref	
	B	31649	11326 (35.8%)	2.8 (2.1-3.8)	<0.0001
	C	15864	2674 (16.9%)	1.0 (0.8-1.2)	

4 UKdb UK HIV Database, HET heterosexual, MSM men who have sex with men, NA not available, F
 5 female, M male, CI confidence Interval.

6 *Table 2: Ethnicity, HIV subtype, recency of infection and age of men who have sex with men*
 7 *(MSM) and potential nondisclosed MSM (pnMSM).*

		MSM	pnMSM	p
n		955	249	
Ethnicity	Black African	20 (2.1%)	42 (16.9%)	<0.0001 [#]
	White	813 (85.1%)	147 (59.0%)	
	Other	114 (11.9%)	58 (23.3%)	
	NA	8(0.8%)	2(0.8%)	
Subtype	A1	9 (1.0%)	4 (1.6%)	0.006 [#]
	B	922 (96.5%)	229 (92.0%)	
	C	24 (2.5%)	16 (6.4%)	
Recency	Recent	236 (83.7%)	53 (84.1%)	1 [#]
	Established	46 (16.3%)	10 (15.9%)	
Age	(mean +/- sd)	36.38 +/- 10.46	39.00 +/- 11.96	0.002 ⁺

8 [#] fisher's exact test, ⁺ two sample t-test, NA not available

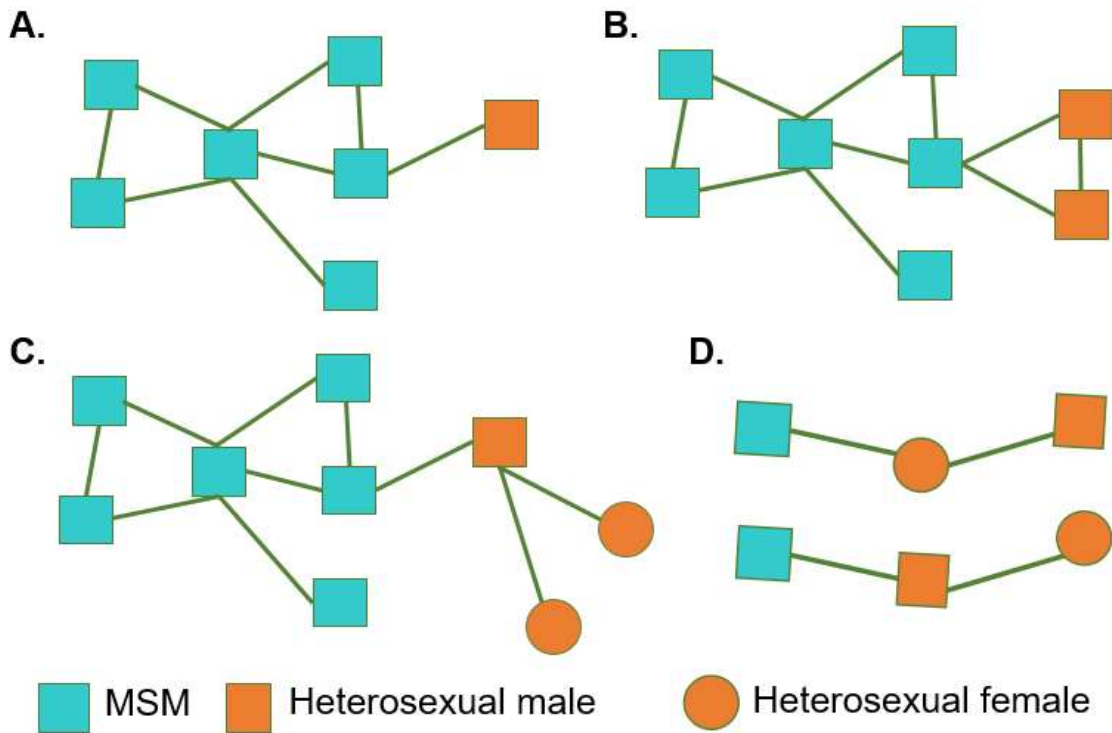
9 *Table 3: Centrality indicators for MSM and potential nondisclosed MSM (pnMSM). P values*
 10 *were calculated using a paired Wilcoxon signed rank test.*

	MSM (MEAN, 95%CI)	PNMSM (MEAN, 95%CI)	P-VALUE
BETWEENNESS CENTRALITY	2.24 (0.98 – 3.51)	1.31 (0.48-2.15)	0.002*
EIGENVECTOR CENTRALITY	0.52 (0.50 - 0.55)	0.50 (0.47 - 0.53)	0.039*
DEGREE	3.02 (2.65 – 3.40)	2.83 (2.46 – 3.20)	0.073

11 Note: statistical significance was calculated using a non-parametric paired Wilcoxon signed rank test

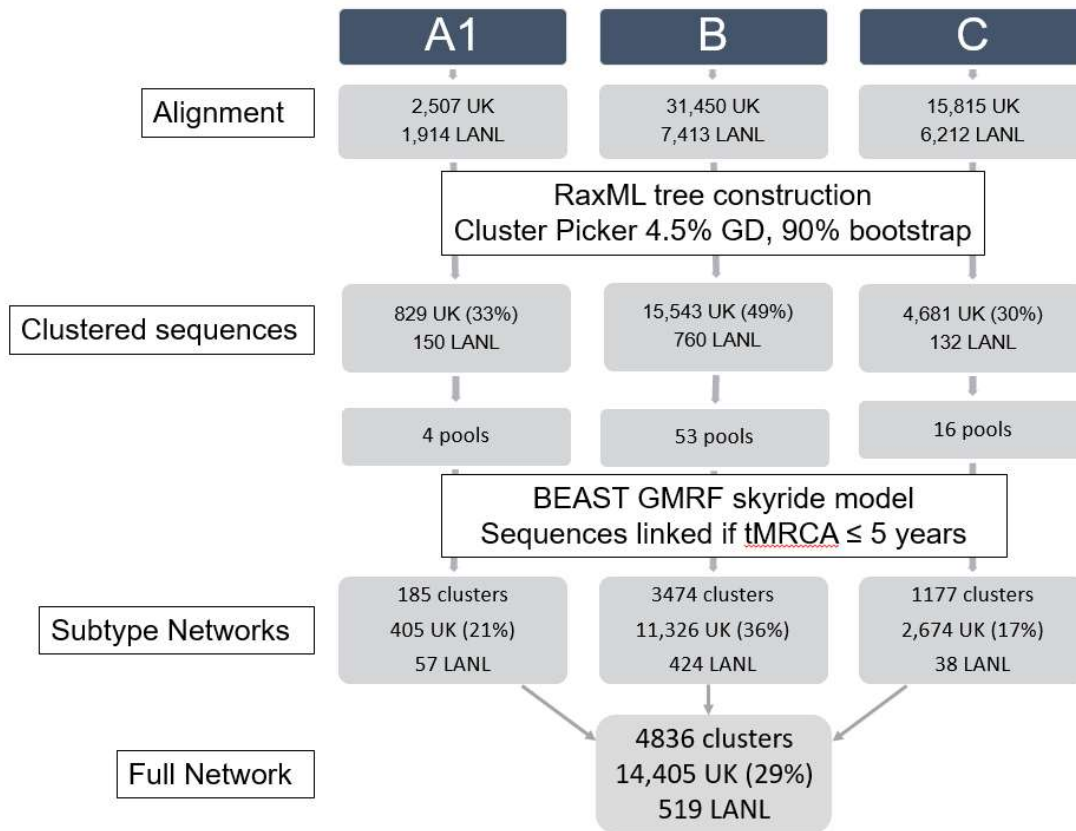
1 **Figures**

2 Figure 1: Cartoon depicting the characteristic positions within men who have sex with men (MSM)
3 clusters of potential nondisclosed MSM (pnMSM). Nodes are coloured by self-reported risk group. A.
4 Based on betweenness and eigenvector, pnMSM tended to be on the periphery of clusters. B. Based
5 on assortativity, pnMSM linked to each other in MSM clusters. C. Based on a brokerage analysis,
6 heterosexual men were more likely to link MSM and heterosexual women than heterosexual
7 females were to link MSM and self-reported heterosexual men. D. The brokerage analysis counts the
8 number of occurrences of each of the two configurations in the network.



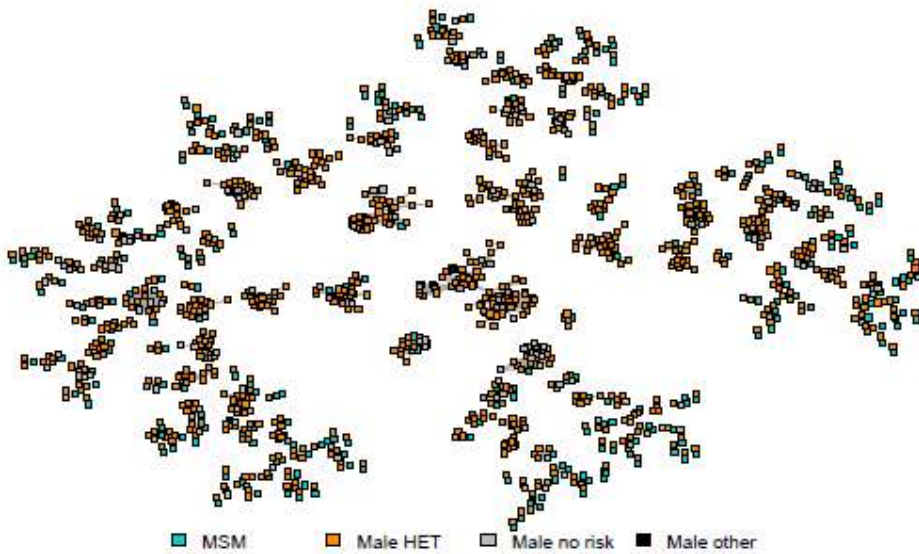
9

10 Figure 2: Data processing flow chart showing the number of sequences at each stage of analysis.
11 Only clusters comprising at least one UK sequence were counted in the RaxML tree. Percentages
12 were calculated with the total number of sequences from that subtype as the denominator. GD:
13 genetic distance, tMRCA: time to most recent common ancestor.



1

2 Figure 3: Clusters containing men who have sex with men (MSM) and heterosexual men. Clusters
 3 were selected if they contained no women and at least one heterosexual man and one MSM. Nodes
 4 are coloured by risk group with heterosexuals in orange and MSM in blue. Nodes in grey are those
 5 for which risk is unknown. All nodes represent men.



6

1 Acknowledgments

2 The authors would like to thank Samantha J. Lycett and Joel O. Wertheim for advice relating to data
3 analysis. The list of UK HIV Drug Resistance Database Steering committee members is available in the
4 Appendix p5.

5 Author contributions

6 MRC, SH and ALB devised the study. TF, AP collected patient data and contributed to UK RDB and
7 provided insights into patient behaviour. AEB and VD provided demographic data for all UK patients
8 and provided insights into epidemiological trends. AT and DD collated and processed the viral
9 sequences. EH wrote code and gave advice throughout the project. MRC conducted all analyses.
10 MRC and ALB wrote the manuscript. All authors reviewed and accepted the final manuscript.

11 Conflicts of interest

12 MRC is currently receiving funding from Gilead for work on Hepatitis C transmission through a UCSD
13 grant, but was not at the time this research was done. AP has received honoraria for advisory boards
14 and symposia from ViiV, Gilead, Merck and Janssen and has acted a consultant for Janssen within the
15 past 4 years. All other authors report no competing interests.

16 Role of funding source

17 The funders had no role in study design, data collection and analysis, decision to publish, or
18 preparation of the manuscript.

19 Ethics committee approval

20 The London Multicentre Research Ethics Committee gave ethical approval (MREC/01/2/10).

21 References

- 22 1. Beyrer C, Baral SD, van GF, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global
23 epidemiology of HIV infection in men who have sex with men. *Lancet*. 2012;380(9839):367-77.
- 24 2. Skingsley A, Yin Z, Kirwan.P., Chau C, Conti S, Presanis A, et al. HIV in the United Kingdom –
25 Situation Report 2015: data to end 2014. Public Health England, London; 2015 11/1/2015.
- 26 3. Schrimshaw EW, Downing MJ, Jr., Cohn DJ. Reasons for Non-Disclosure of Sexual Orientation
27 Among Behaviorally Bisexual Men: Non-Disclosure as Stigma Management. *Arch Sex Behav*. 2016.
- 28 4. Henry E, Marcellin F, Yomb Y, Fugon L, Nemande S, Gueboguo C, et al. Factors associated
29 with unprotected anal intercourse among men who have sex with men in Douala, Cameroon. *Sex*
30 *Transm Infect*. 2010;86(2):136-40.
- 31 5. Ndiaye HD, Toure-Kane C, Vidal N, Niama FR, Niang-Diallo PA, Dieye T, et al. Surprisingly high
32 prevalence of subtype C and specific HIV-1 subtype/CRF distribution in men having sex with men in
33 Senegal. *J Acquir Immune Defic Syndr*. 2009;52(2):249-52.
- 34 6. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of
35 HIV revealed by molecular phylodynamics. *PLoS Med*. 2008;5(3):e50.
- 36 7. Rambaut A. Estimating the rate of molecular evolution: incorporating non-contemporaneous
37 sequences into maximum likelihood phylogenies. *Bioinformatics*. 2000;16(4):395-9.
- 38 8. Fisher M, Pao D, Brown AE, Sudarshi D, Gill ON, Cane P, et al. Determinants of HIV-1
39 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic
40 approach. *AIDS*. 2010;24(11):1739-47.
- 41 9. Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-
42 1 transmission within defined risk groups. *Proc Natl Acad Sci U S A*. 2005;102(12):4425-9.

1 10. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ. Molecular
2 phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog.*
3 2009;5(9):e1000590.

4 11. Hue S, Brown AE, Ragonnet-Cronin M, Lycett SJ, Dunn DT, Fearnhill E, et al. Phylogenetic
5 analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS.*
6 2014;28(13):1967-75.

7 12. Hamilton DT, Handcock MS, Morris M. Degree distributions in sexual networks: a framework
8 for evaluating evidence. *Sex Transm Dis.* 2008;35(1):30-40.

9 13. Robinson K, Fyson N, Cohen T, Fraser C, Colijn C. How the dynamics and structure of sexual
10 contact networks shape pathogen phylogenies. *PLoS Comput Biol.* 2013;9(6):e1003105.

11 14. Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT. Transmission network
12 parameters estimated from HIV sequences for a nationwide epidemic. *J Infect Dis.*
13 2011;204(9):1463-9.

14 15. UK Collaborative HIV Cohort Steering Committee. The creation of a large UK-based
15 multicentre cohort of HIV-infected individuals: The UK Collaborative HIV Cohort (UK CHIC) Study. *HIV*
16 *Med.* 2004;5(2):115-24.

17 16. Public Health England. National HIV/AIDS Reporting System (HARS). 2008.

18 17. Kosakovsky Pond SL, Posada D, Stawiski E, Chappey C, Poon AF, Hughes G, et al. An
19 evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype
20 prediction in HIV-1. *PLoS Comput Biol.* 2009;5(11):e1000581.

21 18. Johnson VA, Calvez V, Gunthard HF, Paredes R, Pillay D, Shafer RW, et al. Update of the drug
22 resistance mutations in HIV-1: March 2013. *Top Antivir Med.* 2013;21(1):6-14.

23 19. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
24 thousands of taxa and mixed models. *Bioinformatics.* 2006;22(21):2688-90.

25 20. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC*
26 *Evol Biol.* 2007;7:214.

27 21. Shapiro B, Rambaut A, Drummond AJ. Choosing appropriate substitution models for the
28 phylogenetic analysis of protein-coding sequences. *Mol Biol Evol.* 2006;23(1):7-9.

29 22. Mercer CH, Hart GJ, Johnson AM, Cassell JA. Behaviourally bisexual men as a bridge
30 population for HIV and sexually transmitted infections? Evidence from a national probability survey.
31 *Int J STD AIDS.* 2009;20(2):87-94.

32 23. Csardi G, Nepusz T. The igraph software package for complex network research.
33 *InterJournal.* 2006;Complex Systems:1695.

34 24. Newman ME. Mixing patterns in networks. *Phys Rev E Stat Nonlin Soft Matter Phys.*
35 2003;67(2 Pt 2):026126.

36 25. Gould RV, Fernandez RM. Structures of Mediation: A Formal Approach to Brokerage in
37 Transaction Networks. *Sociological Methodology.* 1989;19:89-126.

38 26. Schrimshaw EW, Downing MJ, Jr., Siegel K. Sexual venue selection and strategies for
39 concealment of same-sex behavior among non-disclosing men who have sex with men and women. *J*
40 *Homosex.* 2013;60(1):120-45.

41 27. Witzel TC, Melendez-Torres GJ, Hickson F, Weatherburn P. HIV testing history and
42 preferences for future tests among gay men, bisexual men and other MSM in England: results from a
43 cross-sectional study. *BMJ Open.* 2016;6(9):e011372.

44 28. Kahn JG, Gurvey J, Pollack LM, Binson D, Catania JA. How many HIV infections cross the
45 bisexual bridge? An estimate from the United States. *AIDS.* 1997;11(8):1031-7.

46 29. McKeown E, Nelson S, Anderson J, Low N, Elford J. Disclosure, discrimination and desire:
47 experiences of Black and South Asian gay men in Britain. *Cult Health Sex.* 2010;12(7):843-56.

48 30. Poon AF, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, et al. Near real-time monitoring of
49 HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV.*
50 2016;3(5):e231-e8.

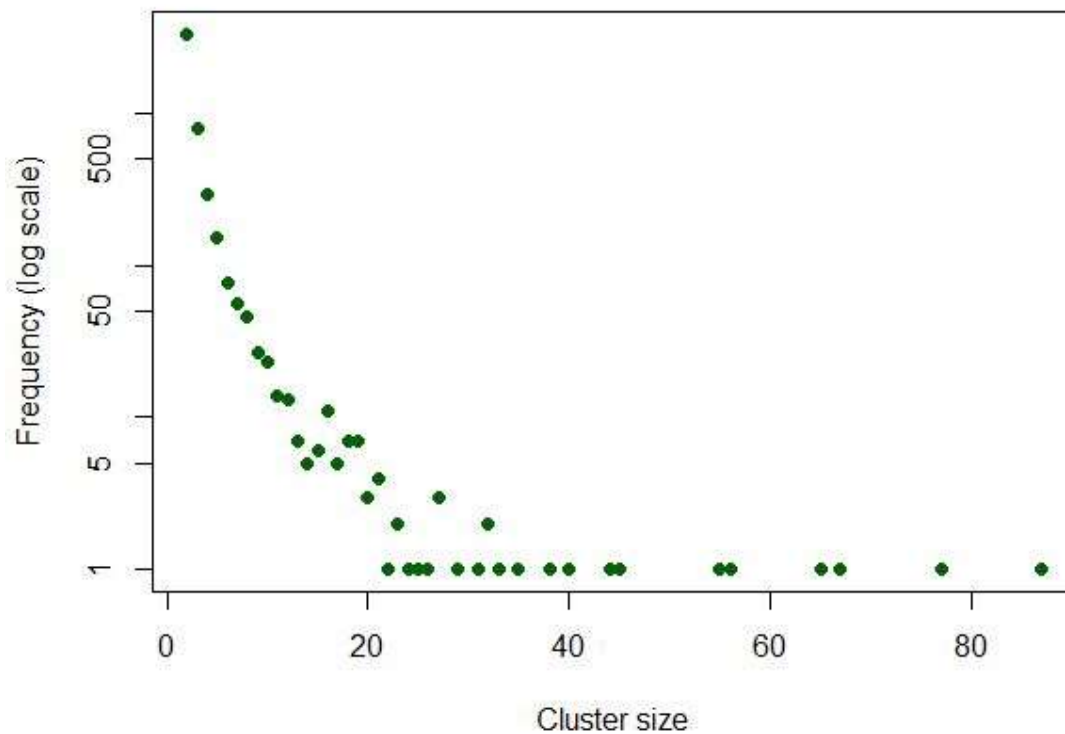
51

1 The position of nondisclosed MSM in UK
2 HIV transmission networks: phylogenetic
3 analysis of a national cohort

Manon Ragonnet-Cronin, Stéphane Hué, Emma B. Hodcroft, Anna Tostevin, David
Dunn, Tracy Fawcett, Anton Pozniak, Alison E. Brown, Valerie Delpech and Andrew J.
Leigh Brown on behalf of the UK HIV Drug Resistance Database

4 [Appendix](#)

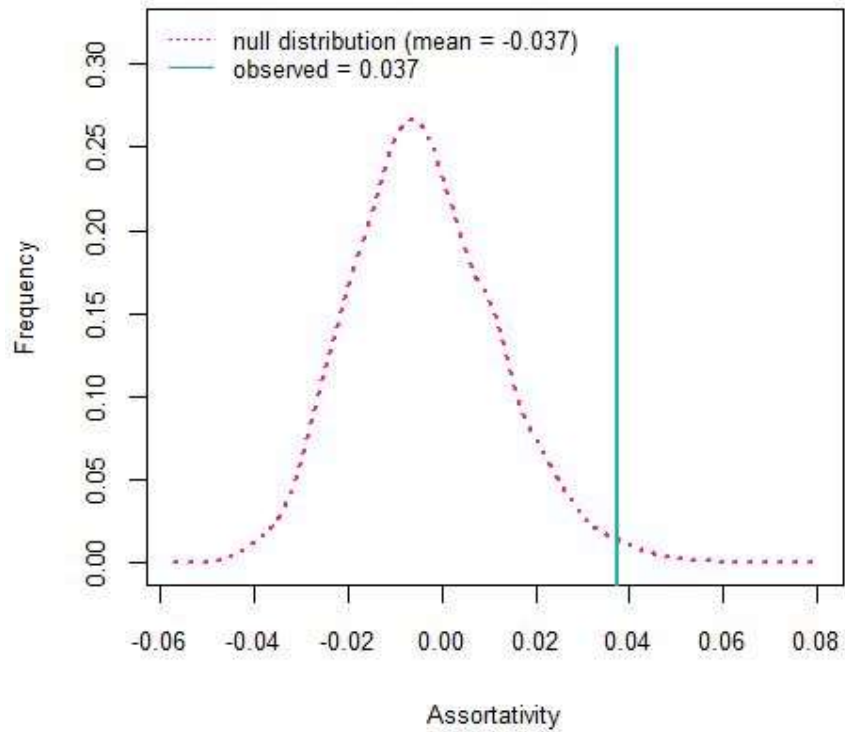
5 [Supplementary Figures](#)



6

7 *Supplementary Figure 1: Cluster size distribution of the UK database HIV transmission*
8 *network.*

9

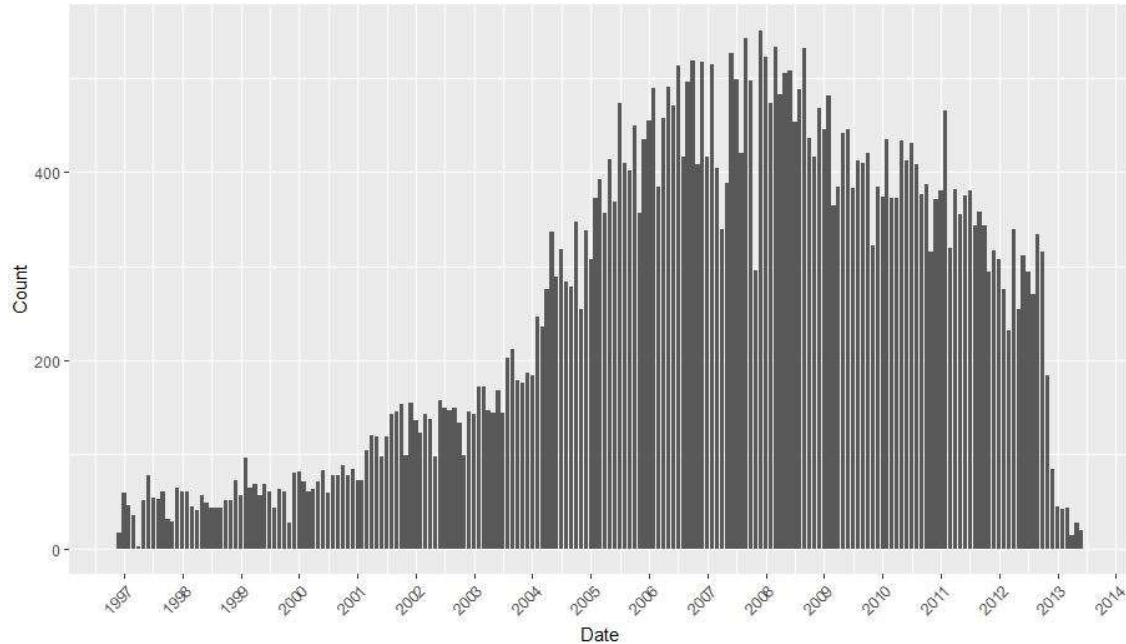


1

2 *Supplementary Figure 2: Assortativity by self-reported risk group was calculated among the*
3 *clusters (in turquoise). In parallel, risk group labels were randomised 10,000 times within*
4 *clusters and assortativity was re-estimated in each network to generate a null distribution*
5 *given risk group frequencies (in pink).*

6

1 [Supplementary Figure 3](#)



2
3 *Supplementary Figure 3: Sample dates for the 50,025 individuals included in the study,*
4 *broken down by month.*

5

6 [Supplementary Tables](#)

7 *Supplementary Table 1: Correlates of clustering and interaction between predictors in a*
8 *multivariate logistic regression model*

		OR	95%CI (lower)	95%CI (upper)	p
Subtype	B	2.8384***	2.0962	3.8429	1.47E-11
	C	0.9928	0.8152	1.2173	0.9438
Sex	Male	1.0846	0.8124	1.4405	0.5779
Risk group	PWID	0.9273	0.2297	3.0538	0.9073
	MSM	2.6300***	1.6119	4.2709	0.0001
	Other/NA	0.875	0.3895	1.7543	0.7253
Ethnicity	Other/NA	2.1835***	1.4882	3.1634	4.78E-05
	White	1.8811***	1.344	2.6122	0.0002
Subtype/ Sex	B/Male	1.4795*	1.0724	2.0476	0.0175
	C/Male	1.2552	0.9364	1.6898	0.1309
Subtype/ Risk group	B/PWID	0.4349**	0.2399	0.8135	0.0073
	C/PWID	1.123	0.556	2.311	0.7489
	B/MSM	0.4136***	0.2651	0.6455	0.0001
	C/MSM	0.7259	0.4541	1.1611	0.1805
	B/Other	1.4702	0.6787	3.5008	0.3528
	C/Other	0.9131	0.4335	2.1142	0.8201
Subtype/ Ethnicity	B/Other	0.4709***	0.3076	0.7277	0.0006
	C/Other	0.8508	0.577	1.2679	0.4204
	B/White	0.6342*	0.4331	0.9332	0.0200

	C/White	1.3528	0.9599	1.9185	0.0869
Sex/ Risk group	M/PWID	1.0662	0.756	1.5142	0.7172
	M/MSM	NA	NA	NA	NA
	M/Other	0.8306	0.5627	1.2304	0.3515
Sex/ Ethnicity	M/Other	0.7657	0.6016	0.974	0.0299
	M/White	0.678***	0.5438	0.8448	0.0005
Risk group/ Ethnicity	PWID/Other	1.5572	0.5203	5.7936	0.4610
	MSM/Other	1.0848	0.7804	1.5116	0.6294
	Other/Other	0.9384	0.5385	1.6213	0.8206
	PWID/White	1.2976	0.4697	4.5893	0.6458
	MSM/White	1.0646	0.784	1.4494	0.6895
	Other/White	0.9772	0.5745	1.6595	0.9319

- 1 Reference groups are not shown: subtype A1, female sex, heterosexuals, Black-African ethnicity.
- 2 PWID people who inject drugs, MSM men who have sex with men, NA not available, CI confidence
- 3 Interval. ***p<0.001, **p<0.01, *p<0.05

4

1 [Genbank accession codes](#)
2 KU498303 - KU499301, EU236439 - EU236538, [JN100661](#) - [JN101948](#), KX661407 - KX663330,
3 Q462040, Q462050, Q462058, Q462067, Q462088, Q462094, Q462095, Q462096, Q462098,
4 Q462100, Q462107, Q462109, Q462129, Q462141, Q462144, Q462156, Q462157, Q462159,
5 Q462164, Q462166, Q462172, Q462175, Q462183, Q462184, Q462216, Q462226, Q462231,
6 Q462234, Q462260, Q462264, Q462267, Q462272, Q462300, Q462333, Q462335, Q462353,
7 Q462403, Q462413, Q462416, Q462417, Q462446, Q462463, Q462464, Q462501, Q462526,
8 Q462532, Q462027, Q462029, Q462030, Q462031, Q462032, Q462033, Q462034, Q462036,
9 Q462042, Q462046, Q462047, Q462049, Q462054, Q462057, Q462059, Q462062, Q462064,
10 Q462070, Q462071, Q462072, Q462074, Q462075, Q462077, Q462081, Q462082, Q462083,
11 Q462085, Q462086, Q462087, Q462089, Q462090, Q462091, Q462093, Q462097, Q462101,
12 Q462102, Q462103, Q462111, Q462112, Q462113, Q462116, Q462118, Q462120, Q462121,
13 Q462123, Q462124, Q462126, Q462128, Q462134, Q462137, Q462139, Q462143, Q462147,
14 Q462148, Q462149, Q462150, Q462152, Q462153, Q462154, Q462155, Q462158, Q462161,
15 Q462163, Q462167, Q462171, Q462173, Q462176, Q462178, Q462179, Q462182, Q462185,
16 Q462186, Q462187, Q462189, Q462195, Q462196, Q462197, Q462198, Q462199, Q462200,
17 Q462208, Q462209, Q462212, Q462213, Q462219, Q462222, Q462227, Q462228, Q462229,
18 Q462230, Q462235, Q462239, Q462240, Q462241, Q462242, Q462244, Q462245, Q462247,
19 Q462248, Q462254, Q462262, Q462265, Q462266, Q462269, Q462273, Q462274, Q462278,
20 Q462280, Q462282, Q462283, Q462285, Q462290, Q462291, Q462294, Q462295, Q462299,
21 Q462301, Q462302, Q462305, Q462306, Q462308, Q462309, Q462311, Q462312, Q462314,
22 Q462315, Q462316, Q462319, Q462320, Q462321, Q462322, Q462325, Q462326, Q462327,
23 Q462328, Q462329, Q462330, Q462331, Q462332, Q462334, Q462337, Q462338, Q462341,
24 Q462342, Q462343, Q462344, Q462346, Q462347, Q462348, Q462349, Q462350, Q462351,
25 Q462352, Q462356, Q462357, Q462358, Q462359, Q462363, Q462364, Q462366, Q462369,
26 Q462371, Q462372, Q462374, Q462375, Q462376, Q462377, Q462382, Q462383, Q462387,
27 Q462392, Q462393, Q462394, Q462395, Q462396, Q462397, Q462398, Q462399, Q462400,
28 Q462401, Q462404, Q462406, Q462407, Q462411, Q462412, Q462414, Q462415, Q462418,
29 Q462419, Q462423, Q462424, Q462425, Q462426, Q462428, Q462441, Q462443, Q462444,
30 Q462447, Q462448, Q462450, Q462454, Q462455, Q462457, Q462458, Q462460, Q462461,
31 Q462466, Q462467, Q462468, Q462473, Q462474, Q462476, Q462477, Q462481, Q462482,
32 Q462483, Q462484, Q462485, Q462486, Q462487, Q462488, Q462489, Q462491, Q462496,
33 Q462497, Q462498, Q462500, Q462502, Q462503, Q462504, Q462505, Q462506, Q462507,
34 Q462508, Q462509, Q462510, Q462511, Q462514, Q462515, Q462516, Q462518, Q462521,
35 Q462522, Q462523, Q462524, Q462527, Q462529, Q462530, Q462531

36

1 [UK HIV Drug Resistance Database Steering committee members](#)

2 **Steering Committee:** David Asboe, Anton Pozniak (Chelsea & Westminster Hospital, London); Patricia
3 Cane (Public Health England, Porton Down); David Chadwick (South Tees Hospitals NHS Trust,
4 Middlesbrough); Duncan Churchill (Brighton and Sussex University Hospitals NHS Trust); Duncan Clark
5 (Barts Health NHS Trust, London); Simon Collins (HIV i-Base, London); Valerie Delpuch (National
6 Infection Service, Public Health England); Samuel Douthwaite (Guy's and St. Thomas' NHS Foundation
7 Trust, London); David Dunn, Esther Fearnhill, Kholoud Porter, Anna Tostevin, Oliver Stirrup (Institute
8 for Global Health, UCL); Christophe Fraser (University of Oxford); Anna Maria Geretti (Institute of
9 Infection and Global Health, University of Liverpool); Rory Gunson (Gartnavel General Hospital,
10 Glasgow); Antony Hale (Leeds Teaching Hospitals NHS Trust); Stéphane Hué (London School of
11 Hygiene and Tropical Medicine); Steve Kaye (Imperial College, London); Linda Lazarus (Expert Advisory
12 Group on AIDS Secretariat, Public Health England); Andrew Leigh-Brown (University of Edinburgh);
13 Tamyo Mbisa (National Infection Service, Public Health England); Nicola Mackie (Imperial NHS Trust,
14 London); Samuel Moses (King's College Hospital, London); Chloe Orkin (Barts Health NHS Trust,
15 London); Eleni Nastouli, Deenan Pillay, Andrew Phillips, Caroline Sabin (University College London,
16 London); Erasmus Smit (Public Health England, Birmingham Heartlands Hospital); Kate Templeton
17 (Royal Infirmary of Edinburgh); Peter Tilston (Manchester Royal Infirmary); Ian Williams (Mortimer
18 Market Centre, London); Hongyi Zhang (Addenbrooke's Hospital, Cambridge).

19

20 **Coordinating Centre:** Institute for Global Health, UCL (David Dunn, Keith Fairbrother, Esther Fearnhill,
21 Kholoud Porter, Anna Tostevin, Oliver Stirrup)

22

23 **Centres contributing data:** Clinical Microbiology and Public Health Laboratory, Addenbrooke's
24 Hospital, Cambridge (Jane Greatorex); Guy's and St Thomas' NHS Foundation Trust, London (Siobhan
25 O'Shea, Jane Mullen); PHE – Public Health Laboratory, Birmingham Heartlands Hospital, Birmingham
26 (Erasmus Smit); Antiviral Unit, National Infection Service, Public Health England, London (Tamyo
27 Mbisa); Imperial College Health NHS Trust, London (Alison Cox); King's College Hospital, London
28 (Richard Tandy); Medical Microbiology Laboratory, Leeds Teaching Hospitals NHS Trust (Tracy
29 Fawcett); Specialist Virology Centre, Liverpool (Mark Hopkins); Department of Clinical Virology,
30 Manchester Royal Infirmary, Manchester (Peter Tilston); Department of Virology, Royal Free
31 Hospital, London (Clare Booth, Ana Garcia-Diaz); Edinburgh Specialist Virology Centre, Royal
32 Infirmary of Edinburgh (Lynne Renwick); Department of Infection & Tropical Medicine, Royal Victoria
33 Infirmary, Newcastle (Matthias L Schmid, Brendan Payne); South Tees Hospitals NHS Trust,
34 Middlesbrough (David Chadwick); Department of Virology, Barts Health NHS Trust, London
35 (Jonathan Hubb); Molecular Diagnostic Unit, Imperial College, London (Steve Kaye); University
36 College London Hospitals (Stuart Kirk); West of Scotland Specialist Virology Laboratory, Gartnavel,
37 Glasgow (Rory Gunson, Amanda Bradley-Stewart).

38 **Funding:** This work is currently supported by the UK Medical Research Council (Award Number
39 164587).

40

41

42