Supporting Material

High Sexual Risk Behavior Indicates Explosive HIV Transmission among Men who have Sex with Men in Vietnam

Authors: Linus Bengtsson[a,b], Xin Lu[a,b,c,d], Fredrik Liljeros[b,c], Hoang Huy Thanh[d], Anna Thorson[a]

[a]*Department of Public Health Sciences. Karolinska Institutet, Stockholm, Sweden;*
[b]*The Institute for Futures Studies, Stockholm, Sweden;*
[c]*Department of Sociology, Stockholm University, Stockholm, Sweden;*
[d]*College of Information Systems and Management, National University of Defense Technology, Changsha, China;*
[e]*Institute for the Study of Society, Economy and the Environment, Ha Noi, Vietnam.*

Corresponding author:
Linus Bengtsson
Dept. of Public Health Sciences
Karolinska Institute
171 77 Stockholm
Sweden
linus.bengtsson@ki.se

Supporting Material

1. Supporting figures

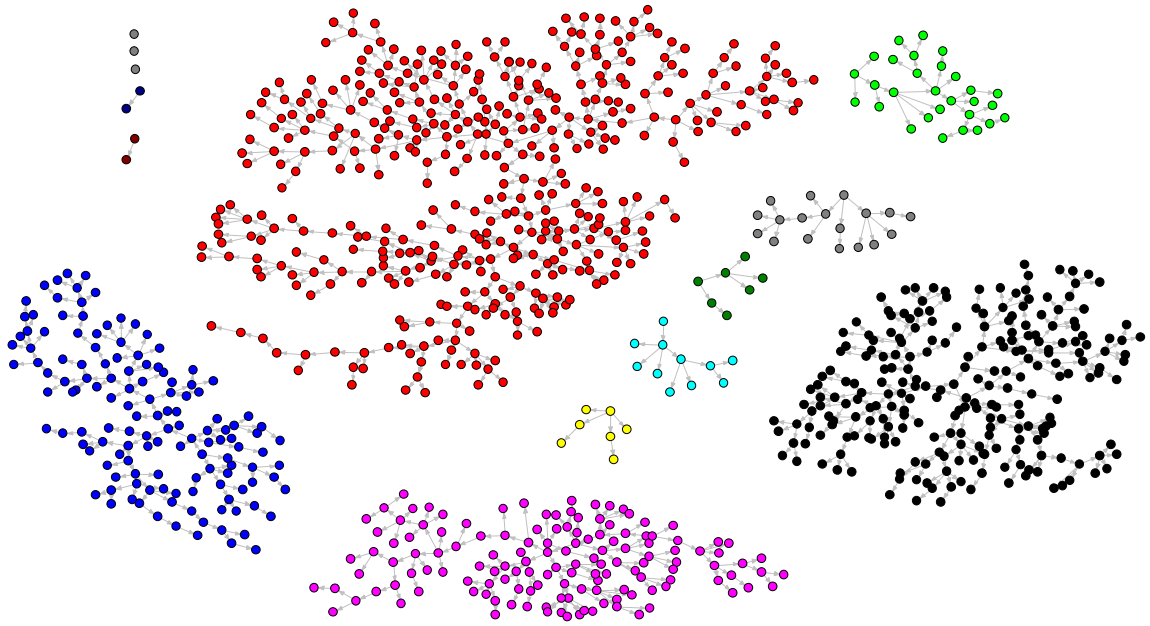

Figure S1: Recruitment trees. Each color represents a separate recruitment tree originating from a single seed. Three seeds did not generate further participant

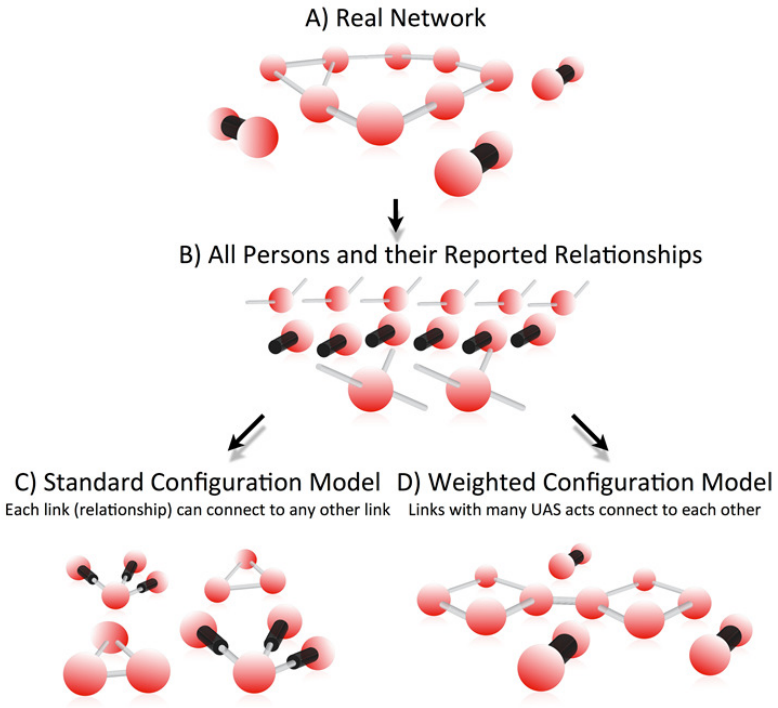

Figure S2: Recreating the sexual network based on individually reported partnerships. Red balls (persons) are connected to each other with thick black links (partnerships with many UAS acts) or by thin grey links (partnerships with few UAS acts); a) The real network contains six persons in stable partnerships and eight individuals with several one-night stands; b) Partner data gathered from the 14 respondents in the network; c) In the standard configuration model the network is re-created by randomly connecting links; d) The weighted configuration model takes advantage of the fact that both partners in any sexual partnership will have the same number of UAS acts and connects links with the same UAS frequency. In the example the method correctly recreates a network with three stable, monogamous couples, disconnected from the rest of the network.

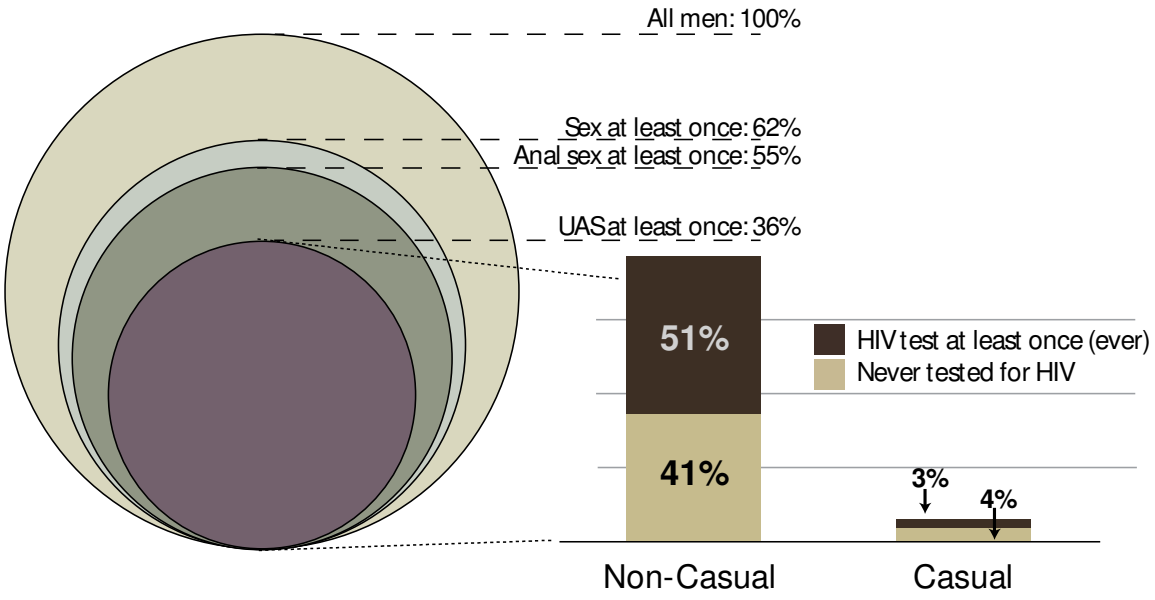

Figure S3: Proportions of all MSM practicing different sexual behaviors last three months. Sixty-two percent of all MSM reported having had sex and 36% of all MSM reported having had at least one unprotected anal sex act; b) Among the group of MSM who had unprotected anal sex during the last three months, a total of 2291 UAS acts took place (non-RDSII adjusted). The bar chart shows the proportion of these UAS acts that took place with non-casual (92%) and casual partners (7%). Within both categories separate numbers are shown for acts in which the respondent had ever tested for HIV vs. tested for HIV at least once E.g. four percent of all reported UAS acts were reported to have taken place with casual partners by respondents who had never tested for HIV. Due to rounding, the four groups in 2b sum to 99%.

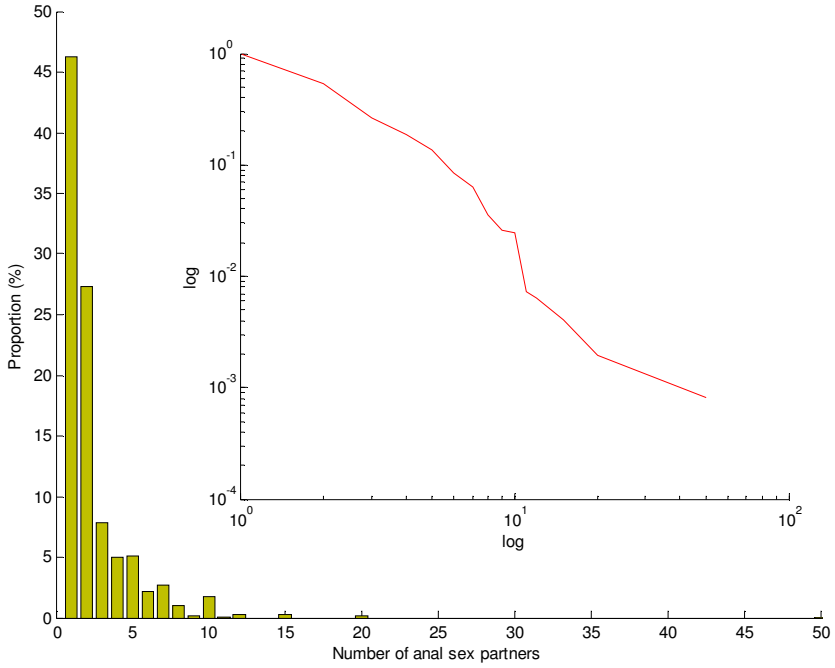

Figure S4: Distribution of anal sex partners during the past three months (with and without condom).

2. Supporting table

In addition to the checks in the main paper we checked all surveys in the cleaned sample for other signs of duplication or presence of non-valid data. We flagged surveys containing a repeated IP number, short completion times or other deviations (as described below). IP numbers are generally shared by all customers at an Internet café, which are common in Vietnam as many people do not own their own computer. Repeated IP-numbers could thus indicate repeated submission by the same person but also participation from an Internet café catering to MSM. We analyzed the sensitivity of the estimates to inclusion and exclusion of these flagged submissions. Specifically, we compared the RDS estimates generated from the cleaned sample with the RDS estimates generated from groups with progressively stricter inclusion criteria. These groups were generated as follows:

Full sample (not shown)
All seeds and non-seed respondents

Cleaned sample[1]
- All non-seed respondents with valid age (≥18 years);
- Exclusion of 4.3% of surveys, which had not provided answers to the socio-demographic items;
- Exclusion of submissions with an email, Yahoo! Chat ID or telephone number, which had previously been registered in the system;

Strict sample
As the cleaned sample, but additionally excluding submissions with an IP address that had previously been registered in the system;

Extreme:
As the strict sample but additionally excluding submissions with short completion times (< 3 minutes) and submissions stating no education (rare in Vietnam).

RDS estimates of socio-demographic variables and sexual partner preferences for the above groups with progressively stricter inclusion criteria are listed in Table S1. The differences between groups are generally small but noticeable differences are present for province and Internet-use. The main change between in these two estimates takes place when we remove participants with a repeated IP address (moving from the cleaned to the strict sample). A smaller proportion of the strict sample consists of persons with infrequent Internet-use and of people from outside the large cities. This is also what we would expect if we removed persons who accessed the survey from Internet cafés.
The average absolute differences in proportional estimates when comparing the cleaned sample to the two other groups, is three percent (maximum difference 21%), see Table S1.

[1] This is the cleaned sample discussed in the paper.

We checked selected analyzes (main paper: Table 2, Figure 1 and Figure 2), using the different subgroups and there were no important changes to the reported results.

		(cleaned sample n=857)				(strict n=639)				(extreme n=592)			
Category	Covariate	n	Raw Prop.	RDSII Est	CI-95%	n	Raw Prop.	RDSII Est	CI-95%	n	Raw Prop.	RDSII Est	CI-95%
Age	18-20	215	25%	25%	22-30%	179	28%	29%	26-36%	174	30%	31%	27-37%
(years)	21-22	288	34%	34%	29-37%	221	35%	33%	29-38%	206	35%	33%	28-38%
	23-25	202	24%	20%	17-23%	145	23%	21%	18-25%	139	23%	22%	18-26%
	26-30	103	12%	13%	10-16%	68	10%	12%	8-15%	54	9%	11%	7-14%
	>30	49	6%	8%	5-12%	26	4%	4%	2-6%	19	3%	4%	2-6%
	Don't want to answer	0				0				0			
Education	12 years or less	117	13%	19%	15-25%	74	11%	12%	9-15%	67	11%	12%	9-16%
(highest level	2-year college	63	7%	8%	6-11%	46	7%	8%	5-11%	41	7%	8%	5-11%
started)	3-year college	156	19%	17%	14-21%	121	19%	20%	16-24%	110	19%	20%	16-24%
	Bachelor level, univ.	482	57%	52%	47-57%	369	58%	57%	52-62%	350	60%	57%	51-62%
	Mast./doct. lev. univ.	34	4%	4%	2-5%	24	4%	4%	2-7%	19	3%	4%	2-6%
	Don't want to answer	5				5				5			
Income	0- less than 1	87	11%	13%	10-17%	58	10%	10%	7-12%	55	10%	10%	8-12%
(million VND	1- less than 3	335	41%	45%	40-50%	246	41%	39%	35-43%	238	43%	39%	35-45%
all sources)	3-less than 5	226	28%	25%	22-29%	169	28%	30%	25-35%	154	28%	29%	25-34%
	Above 5	164	20%	17%	14-20%	131	21%	22%	18-26%	112	19%	21%	17-26%
	Don't want to answer	45				35				33			
Province	HCMC	339	40%	32%	22-34%	292	46%	44%	35-50%	283	48%	45%	37-51%
	Ha Noi	256	30%	26%	18-29%	180	28%	26%	18-31%	165	28%	25%	16-31%
	Hai Phong	55	7%	8%	5-12%	37	6%	8%	4-11%	30	5%	7%	3-11%
	Khang Hoa	20	2%	3%	2-8%	11	2%	2%	1-6%	8	1%	2%	0-5%
	Can Tho	5	1%	1%	0-2%	3	0%	0%	0-1%	3	1%	0%	0-1%
	Da Nang	15	2%	1%	1-2%	9	1%	1%	0-1%	6	1%	1%	0-2%
	Other	160	19%	29%	27-38%	101	16%	19%	16-25%	91	16%	19%	16-26%
	Don't want to answer	7				6				6			
Internet use	1-7	64	8%	19%	21-35%	24	4%	4%	2-6%	22	4%	3%	2-6%
(days/month)	8-14	72	9%	12%	9-15%	36	6%	7%	5-10%	29	5%	7%	4-9%
	15-21	150	18%	16%	14-20%	104	17%	17%	13-20%	84	15%	15%	12-19%
	22-30	559	66%	53%	45-56%	467	74%	72%	67-76%	449	77%	75%	70-79%
	Don't want to answer	12				8				8			
Partner	Prefer only men	544	68%	66%	61-70%	411	68%	68%	62-72%	388	70%	68%	63-73%
Preference	Prefer men to women	181	23%	20%	17-23%	143	24%	24%	20-29%	125	23%	24%	20-28%
	Prefer women to men	20	2%	3%	2-5%	16	3%	4%	2-6%	13	2%	3%	1-6%
	Prefer only women	58	7%	12%	12-24%	30	5%	5%	3-9%	28	5%	5%	4-10%
	Don't want to answer	54				39				38			

Table S1. *Descriptive statistics of socio-demographic variables and sexual partner preferences for samples with progressively stricter inclusion criteria*

3. Sample equilibrium analysis

To get an overview of whether the sample reached equilibrium, we plot both the RDSII estimates along with the increased sample size for most variables surveyed in this study (see Figure S5). During the last 200 respondents, the changes of estimates for all variables are very small. Let $\Delta\hat{p}^{RDSII} = \left|\hat{p}_n^{RDSII} - \hat{p}_{n-200}^{RDSII}\right|$ be the absolute change of estimates for the variable, on average the difference is only 0.87% for proportional estimates with a maximum of 3.6%. For numeric estimates, the maximum difference is only 0.02.

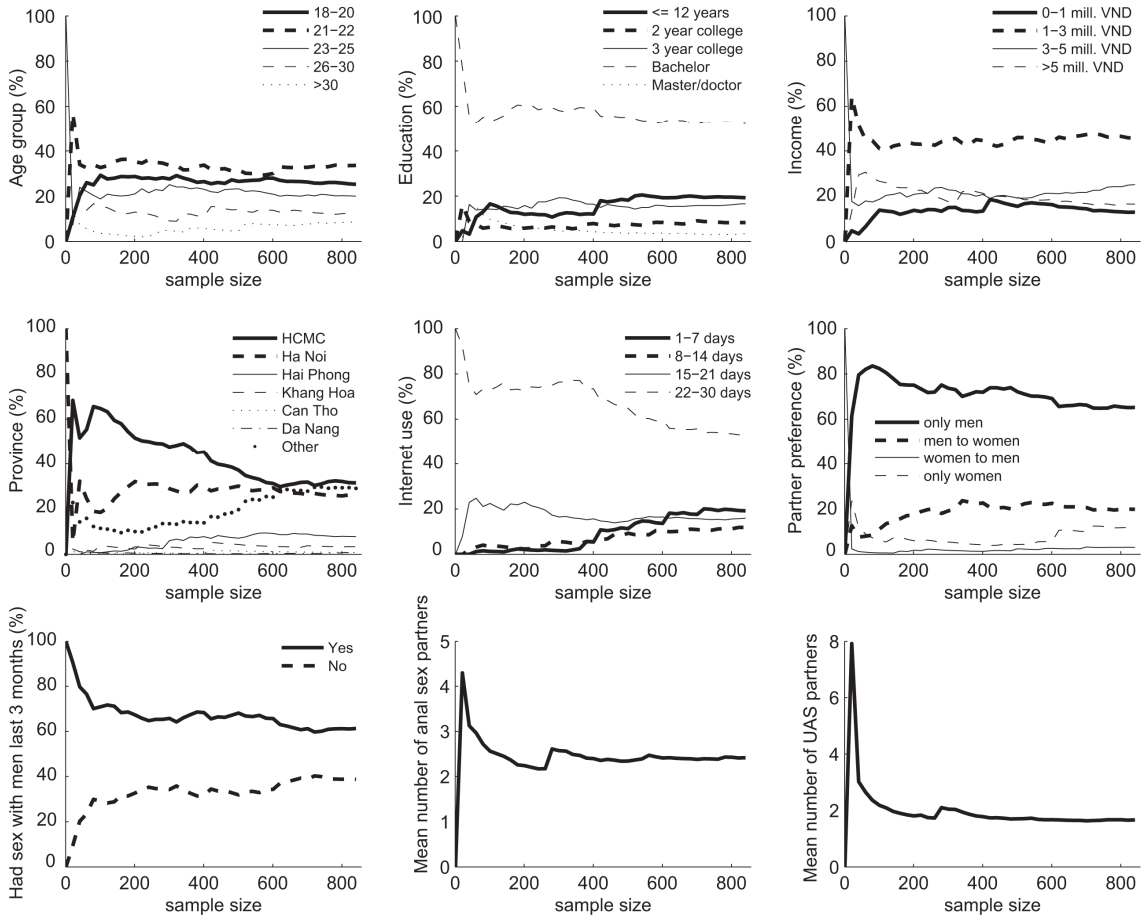

Figure S5: RDSII estimates when sample sizes increase.

4. Weighted configuration model for R0

The weighted configuration model by Britton et al [1] provides means for analysis infectious disease transmission on weighted networks. More specifically, given a degree distribution with a degree-weight dependency, their approach allows an analytical solution for the basic reproduction number, R0, which is defined as a function of the model parameters such that the probability for a large outbreak to occur with any initial infection in the network is positive if and only if R0>0.

Let $P(D = d) = p(d)$ be the probability distribution for nodes with degree d in the network, and let $P(W_{ij} = w \mid D_i = d) = q(w \mid d),\ j = 1,\ldots,d$ be the distribution of edge weights associated with a node of degree d. Assuming that the network is formed by connecting links with the same weight at random (see also Fig 1a-d, main paper), R0 is given by the largest eigenvalue of the mean offspring matrix from the theory of multi-type branching process:

$$M_2 = (m_{dk})_{d,k} \geq 2$$

where m_{dk} is the element of the mean offspring matrix, which is the expected number of *k*-nodes that an infected *d*-node infects:

$$m_{dk} = (d-1)\sum_{w} \pi(w) q(w \mid d)\, \tilde{p}_w(k)$$

where $\pi(w) = 1-(1-s)^w$ is the infection probability for a connection to an infected node with intensity w, and $\tilde{p}_w(k) = \dfrac{q(w \mid k) k p(k)}{\sum_j q(w \mid j)\, j p(j)}$

Note that the definition of R0 we used in this paper (denoted $R0^{3M}$), is slightly different from the standard epidemiological definition, which refer to the average number of new infections over an infected person's *entire infectious period*, while we in this paper use a *three-month period*.

Reference

1. Britton T, Deijfen M, Liljeros F. A Weighted Configuration Model and Inhomogeneous Epidemics. *Journal of Statistical Physics* 2011,**145**:1368-1384.