Supplementary materials: Heavy-tailed sexual contact networks and monkeypox epidemiology in the global outbreak, 2022

Akira Endo*, Hiroaki Murayama, Sam Abbott, Ruwan Ratnayake, Carl A. B. Pearson, W. John Edmunds, Elizabeth Fearon†, Sebastian Funk†

* Correspondence to: akira.endo@lshtm.ac.uk
† These co-last authors contributed equally to this work.

This PDF file includes:

Materials and methods

Data source

We used the distribution of the self-reported number of sexual partners over one year in three of the UK National Survey of Sexual Attitudes and Lifestyles (Natsal) datasets: Natsal-2 (1999–2000; 12,110 participants) (*26*), Natsal-3 (2010–2012; 15,162 participants) (*27*) and Natsal-COVID (2020; 6,654 participants) (*28*). We used the number of reported partners by men and women aged 18–44 who reported at least one sexual partner (same-sex and opposite-sex) over a year in each of the datasets. We defined men and women who reported at least one same-sex partnership over one year as sexually-active men who have sex with men (MSM) and women who have sex with women (WSW), respectively. We assumed that the reported opposite-sex partners in men and women are almost representative of the sexual partnership of heterosexual (HS) individuals. To maximise the sample size for the analysis (especially for MSM, who accounted for about 4% of the male respondents), we combined multiple datasets if the model fit supported the assumption that the samples are derived from the identical distribution (see *Additional results*), which resulted in 409, 290, 7,278 and 4,623 samples used in the estimation of partnership distributions for MSM, WSW, HS men and HS women, respectively.

Of these Natsal datasets, the Natsal-COVID dataset warranted caution: sexual partnership data in the Natsal-COVID dataset was obtained from 29 July to 10 August 2020, three months into the first lockdown in the UK. While a significant minority reported change in their sexual activity during the lockdown (*39*), the questionnaire asked the number of partners "in the last year" and the relative impact of the three months under lockdown on the responses to this question is unclear (some people might also have interpreted "the last year" as the entire 2019, which was then not under the impact of lockdown). The results using the Natsal-2/3 datasets (pre-COVID data) and the Natsal-COVID dataset (COVID wave 1 data) were compared in our sensitivity analysis (*Sensitivity analysis 3*).

The outbreak size as of 31 May 2022 of the monkeypox outbreak in non-endemic regions (globally) and that in the UK was collected from the Global.health public dashboard (*6*). We aggregated both the confirmed and suspected number of cases registered in the dashboard.

Ethical review

This study was approved by the London School of Hygiene & Tropical Medicine ethics committee (reference number: 27985).

Model and simulation procedure

We modelled the branching process of monkeypox transmission over MSM and non-MSM sexual contact networks as follows. Although we do not specify the population (MSM, HS men or HS women) for notational convenience in the following formulation, the same framework was applied to both groups. Let an integer D^{1Y} be a random variable representing the degree of an individual in a sexual partnership network over a year. We assumed that the degree distribution follows a truncated Weibull distribution (truncated at $x = 1$), i.e.:

$$P(D^{1Y} \geq x) \propto 1 - W(x; \alpha, \theta) = \exp\left(-\left(\frac{x}{\theta}\right)^{\alpha}\right), \qquad (x \geq 1) \tag{1}$$

where α and θ are the shape and scale parameters of the Weibull distribution whose cumulative distribution function is represented by W. We assumed that individuals reporting at least one partner over a year are potentially sexually active over the 21 days of infectious period of monkeypox (*30–32*). We estimated the parameters α and θ for same-sex/opposite-sex partnerships in men and women by fitting the Weibull distributions to the empirical degree distribution observed in Natsal datasets using Markov-chain Monte Carlo (MCMC) (see *Additional results* for details). The data was only publicly available for those with at least one sexual partner and thus p_0 was not estimated, which is a nuisance parameter and does not affect the discussion hereafter. Among the three subcategories of non-MSM (HS men, HS women and WSW), HS men had the most heavy-tailed distribution, followed by HS women and WSW. This suggests that WSW, who accounted for about 3% of female respondents, have a sexual contact network with the least transmission potential and thus minimally contribute to overall transmission among non-MSM. We therefore represented the non-MSM sexual network by the partnership distributions of HS men and women. In the following simulation, we used posterior median estimates from the MCMC samples: See *Additional results* section for the estimation procedure.

The distribution of the expected number of partners over 21 days among sexually active individuals was then modelled by a continuous truncated Weibull distribution obtained by scaling the distribution in Equation (1) by a factor of 21/365, i.e.:

$$D^{3W}(x) \propto \frac{d}{dx} W\left(x; \alpha, \frac{21}{365}\theta\right) H\left(x - \frac{21}{365}\right), \tag{2}$$

where $\frac{d}{dx} W\left(x; \alpha, \frac{21}{365}\theta\right)$ is the probability density function of the rescaled Weibull distribution and $H(\cdot)$ is the Heaviside step function specifying the rescaled truncation point of $x = 21/365$. Assuming a proportionality between the risk of infection and the number of partners, the distribution of the mean network degree over 21 days among sexually-associated monkeypox cases will follow a density:

$$D_s^{3W}(x) \propto x D^{3W}(x). \tag{3}$$

We simulated the branching process of an infection cluster originating from a given number of initial cases and estimated the probability that the final outbreak size C_F (excluding the initial cases) is equal to or greater than c, i.e. $P(C_F \geq c)$, by 100,000 simulations. Each case was assumed to transmit the virus to their sexual partners at a constant probability per partner (i.e. secondary attack risk; SAR) over their infectious period. The number of sexual partners of a case, n, was drawn from a Poisson distribution whose mean follows either $D^{3W}(n)$ (sexually active case with non-sexually-associated exposure), $D_s^{3W}(n)$ (cases with sexually-associated exposure) or a zero-inflated distribution $(1-p_A)\delta(n) + p_A D^{3W}(n)$ (random case from the general population with non-sexually-associated exposure; $\delta(n)$ is the Dirac delta function), depending on the type of the initial cases. The mixture weight p_A represents the proportion of sexually active individuals in the general population. We used $p_A = 0.02$ for MSM, $p_A = 0.86$ for HS men and $p_A = 0.89$ for HS women, respectively, referring to the proportion of individuals reporting at least one partner in the Natsal datasets. The number of secondary transmissions was then drawn from a binomial trial with a size n-1 (sexually-associated case, who was infected by one of their sexual partners) or n (non-sexually-associated case) with a probability equal to SAR. We iterated this transmission propagation process for each case over generations of the branching process (using a single partnership distribution for MSM throughout and alternating distributions between generations for non-MSM) until the cluster either becomes extinct or reaches the specified upper limit.

Reproduction number over a network

The basic reproduction number (R_0) over a sexual contact network can be computed using the mean excess degree (*40*), which is given for a truncated Weibull distribution with a lower limit l as:

$$\varepsilon(W) = \frac{\int_l^\infty \max(x-1,0)\, x \frac{d}{dx} W(x;\alpha,\theta) H(x-l)dx}{\int_l^\infty x \frac{d}{dx} W(x;\alpha,\theta) H(x-l)dx}$$

$$= \begin{cases} \dfrac{\theta\Gamma\left(1+\frac{2}{\alpha}, \left(\frac{1}{\theta}\right)^\alpha\right) - \Gamma\left(1+\frac{1}{\alpha}, \left(\frac{1}{\theta}\right)^\alpha\right)}{\Gamma\left(1+\frac{1}{\alpha}, \left(\frac{l}{\theta}\right)^\alpha\right)} & (0 \leq l < 1) \\ \dfrac{\theta\Gamma\left(1+\frac{2}{\alpha}, \left(\frac{l}{\theta}\right)^\alpha\right)}{\Gamma\left(1+\frac{1}{\alpha}, \left(\frac{l}{\theta}\right)^\alpha\right)} - 1 & (l \geq 1) \end{cases} \quad (4)$$

where $\Gamma(a, b)$ is the incomplete gamma function. The R_0's for the MSM and non-MSM sexual contact networks are then defined as

$$R_0 = \begin{cases} R_{MM} = \beta\varepsilon(W_{MSM}) & \text{(MSM)} \\ \sqrt{R_{FM}R_{MF}} = \beta\sqrt{\varepsilon(W_{HM})\varepsilon(W_{HF})} & \text{(non-MSM)} \end{cases} \quad (5)$$

where β is sexually-associated SAR and W_{MSM}, W_{HM}, and W_{HF} are the Weibull distributions corresponding to MSM, HS men and HS women, respectively. We refer to R_{FM} and R_{MF} as the one-way HS reproduction numbers, which represent the mean number of secondary female cases caused by a typical male case and secondary male cases caused by a typical female case, respectively.

Controlling the tail of a sexual partnership distribution

A truncated Weibull distribution can be compared with a Pareto distribution, which can be represented as a linear line in a log-log plot. Let y_P and y_W be the upper cumulative distribution functions of a Pareto and Weibull distribution both defined for $x \geq 1$. We then get

$$\begin{aligned} y_P &= \text{Pareto}(x;\kappa) = x^{-\kappa}, \\ y_W &= \frac{W(x;\alpha,\theta)}{W(1;\alpha,\theta)} = \exp\left(-\frac{x^\alpha - 1}{\theta^\alpha}\right). \end{aligned} \tag{6}$$

By a log-log transform where $X = \log(x)$ and $Y = \log(y)$, the above relationships can be rearranged as:

$$\begin{aligned} Y_P &= -\kappa X, \\ Y_W &= -\frac{\exp(\alpha X) - 1}{\theta^\alpha}. \end{aligned} \tag{7}$$

Note that for $\alpha X << 1$ we have the following approximation for Y_W:

$$Y_W \approx -\frac{\alpha}{\theta^\alpha} X, \tag{8}$$

which gives a Pareto approximation for the body (i.e. non-tail part) of a truncated Weibull distribution. When modelling the possible effect of interventions focusing on individuals with a large number of partners, we assumed that the Pareto-approximation parameter for the Weibull distribution $\kappa = \frac{\alpha}{\theta^\alpha}$ is kept constant while the tail length of the sexual partnership distribution is controlled by a change in α (i.e. the effective number of partners is modified by either reduced contact or reduced chance of transmission per contact). In this way, the tail of the Weibull distribution can be varied with minimal change in the body of the distribution. We estimated the R_0 varying the value of α, which reflects different levels of control on transmission among those with the highest number of partners. To provide better context, we represented the tail length of the distribution by the upper 1st percentile of those with at least one partner over the infectious period instead of α. The sensitivity of R_0 to different combinations of parameters α and κ are also explored in *Additional results*.

Next generation matrix for mixed modes of transmission over MSM and non-MSM networks

To account for the impact of non-sexually-associated transmission routes (e.g. non-sexual skin-to-skin contacts, droplets and fomites) on the overall transmission dynamics as well as the introduction of transmission between MSM and non-MSM sexual contact networks, we used the next generation matrix approach (*33*). Let i_k^t be the number of cases of type k in an infection generation t, where k represents: MSM cases with sexually-associated exposure (k = 1), heterosexual male (k = 2) and heterosexual female (k = 3) cases with sexually-associated exposure, non-MSM cases without sexually-associated exposure (both sexes combined; k = 4) and MSM cases without sexually-associated exposure (k = 5). We represented the reproduction process between a successive pair of infection generations using a next generation matrix K as:

$$\begin{bmatrix} i_1^{t+1} \\ i_2^{t+1} \\ i_3^{t+1} \\ i_4^{t+1} \\ i_5^{t+1} \end{bmatrix} = K \begin{bmatrix} i_1^{t} \\ i_2^{t} \\ i_3^{t} \\ i_4^{t} \\ i_5^{t} \end{bmatrix} = \begin{bmatrix} R_{MM}(1-q) & 0 & 0 & 0 & r_{MM}(1-q) \\ 0 & 0 & R_{MF} & r_{MF}/2 & 0 \\ R_{MM}q & R_{FM} & 0 & r_{FM}/2 & r_{MM}q \\ \rho(1-q') & \rho(1-q') & \rho(1-q') & \rho(1-q') & \rho(1-q') \\ \rho q' & \rho q' & \rho q' & \rho q' & \rho q' \end{bmatrix} \begin{bmatrix} i_1^{t} \\ i_2^{t} \\ i_3^{t} \\ i_4^{t} \\ i_5^{t} \end{bmatrix} \quad (9)$$

where r_{MM}, r_{FM} and r_{MF} correspond to the mean partnership degree of a random MSM, non-MSM male and female case (i.e. acquired via non-sexually-associated exposure), respectively, and ρ is the reproduction number for general non-sexually-associated routes. The average proportion of the number of female sexual partners that an MSM case would have among the total partners is given as q, which we refer to as the MSM to non-MSM mixing ratio. On the other hand, sexually-associated transmission from non-MSM cases to the MSM population was considered negligible compared to the magnitude of sexually-associated transmission among MSM. Similarly, the average proportion of non-sexually-associated contacts an individual has with MSM is given as q'. We assumed that non-sexually-associated transmissions are not strongly assortative by sexual orientation and that q' is thus small reflecting the proportion of MSM in the population. We also assumed that the risks of transmission via sexually-associated and other routes are mutually independent such that the number of partners among non-sexually-associated non-MSM cases follows the distribution among the general population.

When the generation interval is shared between sexually-associated and other routes of transmission, the relative magnitudes of i_k^t's tend to the dominant eigenvector of K. Let (i_1, i_2, i_3, i_4, i_5) be the eigenvector of K. Equation (9) suggests

$$\begin{aligned} &\rho(1-q')(i_1+i_2+i_3+i_4+i_5) = \lambda i_4, \\ &\rho q'(i_1+i_2+i_3+i_4+i_5) = \lambda i_5, \end{aligned} \quad (10)$$

where λ is the dominant eigenvalue of K. The asymptotic proportion of non-sexually-associated cases, $(i_4 + i_5)/(i_1 + i_2 + i_3 + i_4 + i_5)$, therefore corresponds to ρ/λ.

Let us assume $\frac{q'}{1-q'} \cdot \frac{r_{MM}}{R_{MM}} \ll 1$. Equations (9) and (10) then imply, if $i_1 \geq i_4$ (which holds in the range considered in this study),

$$\begin{aligned} &\lambda i_1 = (1-q)R_{MM}i_1 + (1-q)r_{MM}i_5 \\ &= (1-q)R_{MM}\left(i_1 + \frac{q'}{1-q'} \cdot \frac{r_{MM}}{R_{MM}} \cdot i_4\right) \approx (1-q)R_{MM}i_1, \end{aligned} \quad (11)$$

which indicates that the eigenvalues of K can be approximated by substituting its (1,5) entry with 0. It can be shown that the eigenvalues of such a matrix are $R_{\mathrm{MM}}(1-q)$ and the four eigenvalues of the (1,1) minor of K. Within the range of parameters considered in this study, $R_{\mathrm{MM}}(1-q)$ will be the dominant eigenvalue and in such instances we get

$$\frac{i_4 + i_5}{i_1 + i_2 + i_3 + i_4 + i_5} = \frac{\rho}{R_{MM}(1-q)}. \quad (12)$$

Note that when $q = 0$, the proportion of non-sexually-associated cases and the ratio between reproduction numbers are expected to be equal.

All the analysis was conducted either in Julia v.1.7.2 or R v. 4.0.2. Source codes are available on a GitHub repository: (https://github.com/akira-endo/monkeypox_heavytail).

Supplementary text: Additional results

Parameter estimation for Weibull distributions for sexual partnerships from the Natsal datasets

To parameterise the branching process model of transmission among the MSM and non-MSM populations, we fitted the Weibull distribution in Equation (1), truncated at $x = 1$, to the reported number of partners in Natsal datasets (who reported at least one partner over the previous year) using MCMC. Weibull distributions are usually characterised by two parameters: shape α and scale θ; however, to preclude a high correlation between parameters in posterior samples, we used an alternative parameterisation (α, κ), where $\kappa = \frac{\alpha}{\theta^{\alpha}}$ (see Equation (8))., We estimated the parameters separately for four different partnership distributions: same-sex and opposite-sex partnerships reported by men and women. For each of the partnership category, we used different subsets of the dataset that we assumed to be the most representative of the current sexual behaviours. These selections were informed by the model comparison described in the next section.

We employed a weakly-informed prior HalfNormal($\mu = 0$, $\sigma = 10$) in parameter estimation, except for the shape parameter α for the same-sex partnership reported by men, which had a limited sample size to fully inform this parameter. We set an informative prior for this parameter based on the gonococcal resistance to antimicrobials surveillance programme (GRASP) dataset: see the next section for details.

We used the Hamiltonian Monte Carlo algorithm with No-U-Turn-Sampler (via {rstan} package run in R v.4.0.2) and obtained 15,000 samples from five chains after discarding the first 2,000 samples as warm-up. The resulting MCMC samples showed an R-hat statistic of below 1.01 and an effective sample size of at least 2,000. Posterior median estimates and 95% credible intervals are shown in Table S1. Since the same-sex partnership reported by women had shorter and less heavy tail (indicated by large values of α and κ) than HS men or women, sustained transmission over the WSW sexual contact network is expected to be the least likely. Their estimates are shown in Table S1 for completeness but not used in our analysis. The observed and simulated distributions for MSM and HS partnerships are shown in Fig. S1.

Table S1. Maximum number of sexual partners among the MSM population in empirical and simulated datasets assuming a Pareto distribution.

Category of partners	Natsal dataset used	Sample size	Shape parameter (α)	Pareto-approximation parameter (κ)
Same sex, reported by men	2/3/COVID	409	0.10 (0.02–0.19)	0.77 (0.66–0.88)
Opposite sex, reported by men	2/3	7,278	0.01 (0.00–0.04)	1.68 (1.63–1.73)
Same sex, reported by women	2/3/COVID	380	0.01 (0.01–0.07)	2.08 (1.85–2.31)
Opposite sex, reported by women	3	4,623	0.01 (0.00–0.03)	2.14 (2.07–2.21)

Median estimates and 95% credible intervals are shown.

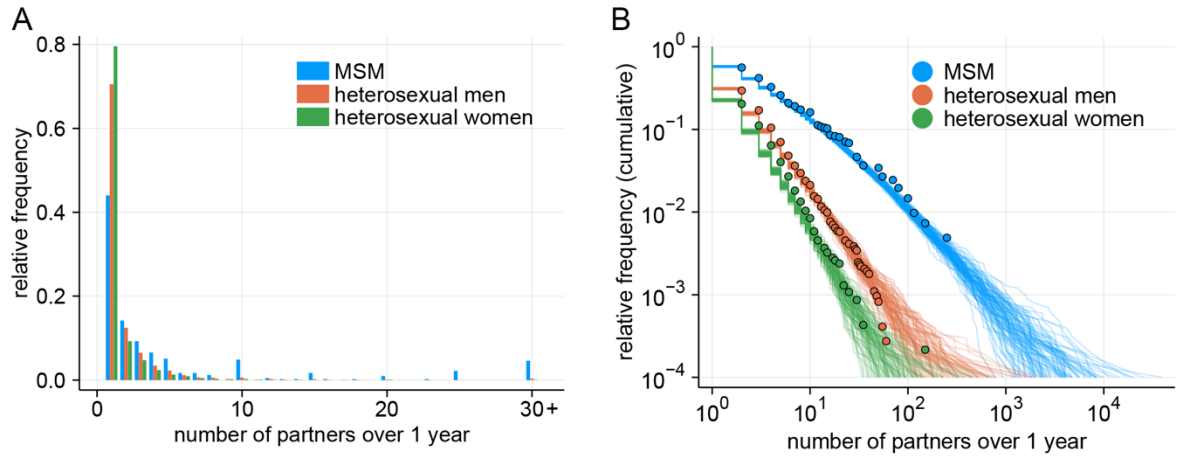

Fig. S1. Observed and simulated partnership distributions for MSM and HS populations. (A) Observed distributions for MSM and HS sexual partnerships over 1 year in the Natsal datasets (MSM: Natsal-2/3/COVID; HS men: Natsal-2/3; HS women: Natsal-3). (B) Comparison between observed and simulated partnership distributions. For each of the populations, 100 simulated datasets (each containing 10,000 samples) were drawn from the Weibull distributions parameterised by the posterior median estimates.

Prior specification for the MSM partnership estimation using GRASP data

Due to the limited sample size, the shape parameter α for the MSM sexual partnership distribution was not sufficiently informed by the Natsal data alone and required an informative prior from external source to realistically (but minimally) constrain the lower bound of its possible range. We used the fact that the reported maximum partners over three months was 700 among 691 samples from the GRASP data (Table S3) to define the informative prior. We fixed the Pareto-approximation parameter for the truncated Weibull distribution at $\kappa = 0.6$, the reported Pareto exponent parameter in (*24*), and obtained the likelihood of observing the maximum of 700 partners out of 691 samples:

$$L(\alpha) \propto w(700; \alpha, \kappa = 0.6)\left(\int_0^{700} w(x; \alpha, \kappa = 0.6)\, dx\right)^{690}, \qquad (13)$$

where $w(x; \alpha, \kappa)$ is the probability density function for the Weibull distribution. To limit the effect of the prior and prioritise the fit to the Natsal datasets rather than GRASP, we assigned a weight of 0.1: i.e. used $L(\alpha)^{0.1}$ as the prior. Note that the shape parameter of the Weibull distribution is scale-invariant and thus $L(\alpha)$ obtained from the 3-month partnership in GRASP data was assumed to directly inform α used for the 21-day partnership distribution in the main analysis.

Model comparison to check for temporal changes between Natsal datasets

In our parameter estimation, we aimed to combine multiple Natsal datasets (chosen from Natsal-2, Natsal-3 and Natsal-COVID, conducted in 1999–2000, 2010–2012 and 2020, respectively) to ensure sufficient sample sizes (in particular for MSM), if there is sufficient support to assume that samples are from the same distribution over multiple surveys. Specifically, for the four partnership distributions (displayed in Table S1), we compared three candidate assumptions with different combinations of shared parameters between datasets: the same parameters for all the three datasets ("All same"); different parameters for all the three datasets ("All different"); and the same parameters for Natsal-2 and 3 and different for Natsal-COVID ("Only 2020 different"). We used the Widely-applicable Bayesian Information Criterion (WBIC) (*41*) to compare models based on these assumptions. We considered a difference in WBIC of 2 or greater as positive support and 6 or greater as strong

support (*42*). The results suggested that the same-sex partnership data reported by men was most likely derived from the same distribution throughout the three Natsal datasets and that the opposite-sex partnership data reported by men likely had different distributions between Natsal-2/3 and Natsal-COVID. The opposite-sex partnership reported by women was strongly suggested to have followed different distributions between the three surveys. Assuming that the suggested deviation in the opposite-sex partnership distribution for both men and women reported in 2020 from the previous years may reflect the lockdown-associated behaviour changes, we used the all three Natsal datasets for MSM, Natsal-2 and 3 for HS men and Natsal-3 for HS women to estimate the partnership distributions in our main analysis. Results for alternative combinations of the datasets (pre-COVID and COVID wave 1 datasets) were also explored as part of our sensitivity analysis (*Sensitivity analysis 3*).

Table S2. Model selection to check for consistency between Natsal datasets.

Category of partners	WBIC values for different assumptions		
	All same	All different	Only 2020 different
Same sex, reported by men	**1919.0 (best)**	1932.2 (Δ 13.2)	1933.3 (Δ 14.3)
Opposite sex, reported by men	18873.5 (Δ 26.8)	18858.9 (Δ 12.2)	**18846.7 (best)**
Same sex, reported by women	**790.3 (best)**	797.3 (Δ 7.0)	792.2 (Δ 1.9)
Opposite sex, reported by women	17869,0 (Δ 32.3)	**17832.7 (best)**	17877.5 (Δ 44.8)

WBIC: Widely-applicable Bayesian Information Criterion. Δ: difference from the best model. The models with the best WBIC are highlighted in bold.

Maximum number of partners predicted by Pareto distributions fitted to empirical MSM sexual partnership data in previous studies

We simulated the Pareto-distributed number of sexual partners based on the estimates of exponent parameters reported in previous studies (*23*, *24*) and compared the maximum number of partners reported in the datasets and that in the simulated samples. Parameter estimates in those studies are presented as an exponent parameter γ, which corresponds to $\kappa + 1$ in our parameterisation in Equation (6). For each of the dataset, a set of samples with a size equal to that of the original dataset was repeatedly drawn 1,000 times and the quantiles of the maximum number of partners among each sample set are compared with the original (Table S3). The Pareto distribution based on the parameter estimate for each of the dataset tended to produce substantially larger values for the maximum number of sexual partners among the simulated samples than observed. The observed maximum values in the three datasets (Natsal-2, Natsal-3 and GRASP) were all around the lower 2.5th percentile of the Pareto-simulated maxima or below and smaller than the simulated medians by an order of magnitude or two. These results suggest that the Pareto distributions used to describe the sexual partnership distributions among MSM in the previous studies failed to capture the tail part of the empirical data and that a distribution with a more modest behaviour at the tail such as the Weibull distribution is expected to better characterise the observed sexual partnership distribution.

Table S3. Maximum number of sexual partners among the MSM population in empirical and simulated datasets assuming a Pareto distribution.

Dataset		Natsal-2	Natsal-3	GRASP
Study		Schneeberger et al., 2004 (*23*)	Whittles et al., 2019 (*24*)	Whittles et al., 2019 (*24*)
Year of survey		1999–2000	2010–2012	2004
Population included		General	General	Gonorrhoea-infected
Sample size (MSM, ≥ 1 partner)		138	188	691
Reporting window for partners		1 year	1 year	3 months
Exponent parameter (γ)		1.6	1.81	1.6
Maximum number of partners reported		**250**	**100**	**700 (650–750)***
Maximum number of partners in simulated samples (percentile)	0% (Min)	202	48	3,137
	2.5%	475	129	6,161
	25%	2,168	473	30,384
	50%	6,824	1098	105,193
	75%	28,841	3,480	3.8×10^5
	97.5%	2.2×10^6	60,271	4.5×10^7
	100% (Max)	5.2×10^9	1.1×10^7	8.9×10^9

* Based on reading of the figure and subject to uncertainty

Sensitivity analysis 1: outbreak size in the UK

In the main analysis, we discussed the outbreak of monkeypox over sexual network in the global context, assuming the sexual partnership distributions in the UK are also representative of behaviours in other countries. Here, we shifted our focus to the outbreak within the UK, where the estimated sexual partnership distributions would most likely apply, by using the number of reported cases in the UK as of 31 May 2022 (190 cases). The results were almost identical to the main analysis (Table S4), suggesting that our conclusions would plausibly apply to the domestic outbreak in the UK (and other countries with a similar local outbreak size and sexual partnership distributions comparable to those in the UK).

Table S4. Likelihood of an outbreak over MSM or non-MSM sexual contact network given different numbers and profiles of introduction events, using UK outbreak size data.

SAR	Likelihood of 190+ cases in MSM given 1 introduction event			Likelihood of 10000+ cases given multiple introductions	
	S-A event (MSM)	Non-S-A event (MSM)	Non-S-A event (Gen. pop.)	728 S-A events in MSM	1000 non-SA events in non-MSM
5%	10%	0.26%	0.007%	100%	< 0.001%
10%	20%	0.9%	0.02%	100%	< 0.001%
20%	32%	2.6%	0.05%	100%	< 0.001%
50%	45%	7.4%	0.15%	100%	< 0.001%

SAR: secondary attack risk; MSM: men who have sex with men; S-A: Sexually-associated; Gen. pop.: general population.

Sensitivity analysis 2: Variations in Weibull parameters

We assessed the sensitivity of R_0 over MSM and one-way HS reproduction numbers for HS men and women to variations in parameter estimates for the Weibull distribution fitted to the sexual partnership distributions (Fig. S2). The Pareto-approximated exponent κ was varied

within the credible intervals in Table S1 and the shape parameter α was varied between 0.001–1.0 and transformed into the upper 1st percentile of the number of partners (among those with non-zero partners) over 21 days. The results suggested that small changes in the shape of sexual partnership distributions can have a significant impact on the reproduction numbers.

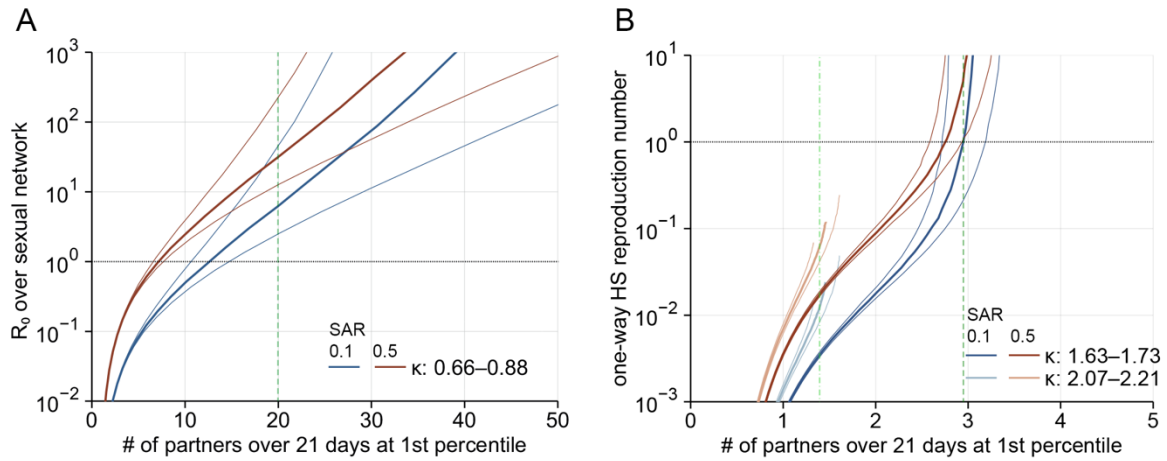

Fig. S2. Sensitivity to variations in Weibull parameters for MSM and HS partnership distributions. (A) Projected R_0 with the Weibull parameters varied within the 95% credible intervals for the sexual partnership distribution among MSM. The central thick lines correspond to the median estimates for κ and thinner lines the credible interval limits. The dotted horizontal lines denote the epidemic threshold ($R_0 = 1$). Dashed green line indicates the upper 1st percentile of the partnership distribution among those with non-zero partners corresponding to the median estimates. (B) Projected one-way HS reproduction numbers (R_{FM}: darker lines; R_{MF}: lighter lines) with the Weibull parameters varied within the credible intervals for the sexual partnership distributions among HS men and women. Dashed/dash-dotted green lines indicate the upper 1st percentiles for HS men (dashed darker green) and women (dash-dotted lighter green) corresponding to the median estimates. The open end of the curves corresponds to the lower limit of α = 0.001.

Sensitivity analysis 3: Infectious period and pre-COVID / COVID wave 1 datasets

We varied our assumption of 21 days for the infectious period of monkeypox as this estimate is based on the duration of illness from limited empirical data (*30–32*). The effective infectious period could also be shorter if individuals with symptoms refrain from having sexual contact. We used alternative assumptions of the infectious period (7, 14 and 28 days) and compared the R_0 and the relative reduction required for control (Figs. S3A and S3B). The results suggest that the R_0 for the non-MSM sexual contact network is below 1 for the entire range of SAR values if the infectious period is 7 or 14 days in this setting, while the R_0 exceeds 1 with an SAR of ~ 50% or higher if the infectious period is 28 days.

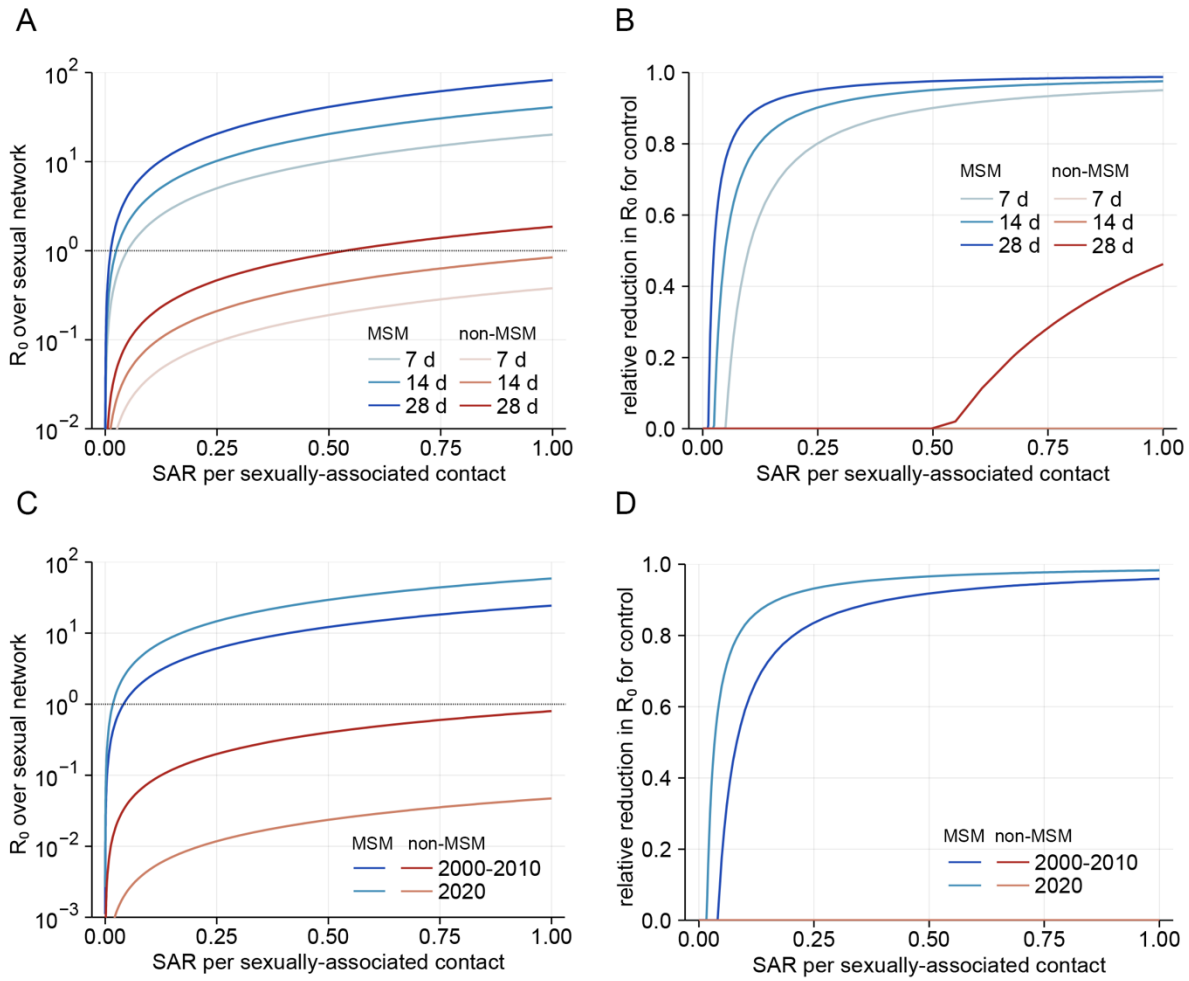

Fig. S3. R_0 of monkeypox over sexual contact networks and relative reduction required for control based on alternative infectious periods and datasets. (A) Projected R_0 over the MSM and non-MSM sexual contact networks and (B) relative reduction in R_0 required to bring the outbreak under control, assuming shorter and longer infectious periods (7 days, 14 days and 28 days) than the main analysis. (C) Projected R_0 over the MSM sexual contact networks and (D) relative reduction in R_0 required to bring the outbreak under control, based on alternative subsets of the Natsal data. The estimates from the pre-COVID datasets (Natsal-2/3, "2020-2010") and the COVID wave 1 data (Natsal-COVID, "2020") were used.

We also compared estimates obtained from alternative combinations of datasets, i.e. pre-COVID (Natsal-2 and 3, conducted in 1999–2000 and 2010–2012, respectively) and COVID wave 1 (Natsal-COVID conducted in summer 2020) datasets (Figs. S3C and S3D), to assess how having 3 months under lockdown prior to the Natsal-COVID survey might affect the estimated R_0 of monkeypox via possible deviations in the sexual partnership distributions. The model comparison supported that the parameters for MSM be shared between the three datasets and the parameters for non-MSM (HS men and women) be separated between Natsal-2 and/or 3 and Natsal-COVID (Table S2). However, it is of note that not selecting the Natsal-COVID dataset to characterise the current sexual contact patterns among non-MSM (which we assumed are not under the effect of COVID-related changes) in the main analysis is only an assumption. The results of the sensitivity analysis suggested that using either the Natsal-2/3 or Natsal-COVID datasets instead of the three datasets combined for MSM does not lead to a substantial qualitative change. On the other hand, using the Natsal-COVID dataset for the non-MSM sexual partnership distribution caused an over 10-fold reduction in

the estimated R_0 over the non-MSM sexual network, which probably reflects the reduction in the tail-part of the partnership distribution among non-MSM during the lockdown.

Sensitivity analysis 4: Degree assortativity

We assessed how degree assortativity, which we did not account for in the main analysis, may increase the reproduction number among a subset of HS population using a simple adjustment in our model. We defined a 'core group' as individuals with degrees that are above the specified quantile in the population ('threshold') and assumed that they come in contact exclusively within themselves. We estimated the R_0 within this exclusively-connected HS core group for different threshold values (Figure S4). The results suggested that R_0 may be above 1 in a high-degree subset of population if they form a densely connected network within themselves. Notably, even with an assumed SAR of 10%, R_0 may exceed 1 if individuals in the top 1% quantile form an exclusive core group.

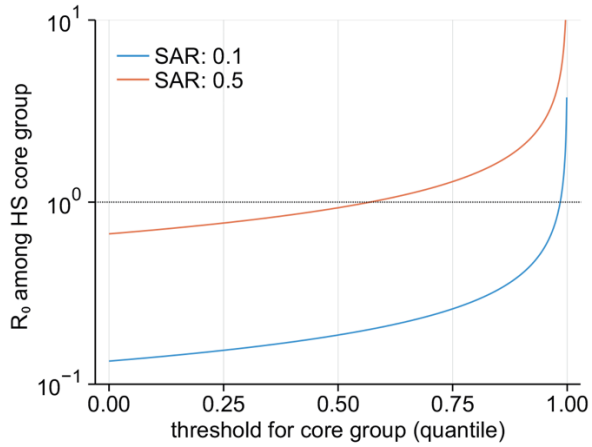

Fig. S4. R_0 of monkeypox within a heterosexual assortative core group. A heterosexual 'core group' was defined as a group of individuals whose degrees are above the specified quantile ('threshold for core group') and was assumed to contact exclusively within themselves.

Sensitivity analysis 5: Depletion of susceptibles

We used a simple model to consider the possible effect of depletion of susceptibles on the reproduction number in finite MSM populations (e.g. in a country). As individuals with higher degrees are likely to be infected in the earlier phase of an outbreak, the effect of depletion of susceptibles on the reproduction number is expected to be significant in heavy-tailed contact networks. We assumed that previously infected (immune) individuals retain their original partners but that they no longer become infected even if they are contacted by an infectious partner. We first sampled the degrees of 100,000 or 500,000 individuals from the MSM sexual partnership distribution. We then simulated the effective reproduction number over the progression of an outbreak in the following steps: (i) chose one of the currently susceptible individuals at a probability proportional to their degree; (ii) multiply the chosen individual's degree by the degree-weighted proportion susceptible in the population and assign that value as their 'effective degree'; (iii) remove the individual from the susceptible pool; (iv) repeat (i)–(iii) to obtain a sequence of the effective degree of the n-th case and define the effective reproduction number as the product of the effective degree and SAR. The 21-case moving average (i.e. from n – 10 to n + 10) of the effective reproduction number was computed to smooth out stochastic fluctuations. Figures S5A and S5B show the median and 95% quantiles of the effective reproduction numbers (computed from 1,000 simulated sequences) in MSM populations of size 100,000 and 500,000, respectively. The estimated effective reproduction number in the presence of depletion of susceptibles showed

a decline over the progression of an outbreak and suggest that the effective reproduction number may go below 1 before the cumulative case count reaches the classical herd immunity threshold: $1-1/R_0$. The growth rate of monkeypox cases in the UK was estimated to be near 0 by mid-July 2022, when a total of ~2,000 cases have been observed (*43*). Assuming that MSM account for 2% of the UK population according to the Natsal data, the MSM population size between age of 15 to 49, approximately corresponding to those above the age of sexual consent (16 years old) and not targeted by the past smallpox vaccine campaign (discontinued in 1971 in the UK (*44*)), is estimated to be about 500,000 (*45*). These figures are generally consistent with our simulation if SAR lies between 10 and 50%. However, a caveat must be noted that this analysis does not consider importation of cases from outside the population; around 30% of cases have reported travel history in the UK (*46*) and globally (*35*), suggesting that the assumption of closed populations may underestimate the outbreak sizes.

We also combined this analysis of depletion with consideration of assortativity, again by assuming that the 10% highest-degree MSM individual form the exclusive core group and estimated the change in the effective reproduction number among them (Figures S5C and S5D). While the initial reproduction number is higher, the cumulative case count at which the effective reproduction number is estimated to cross 1 is similar to that in the baseline scenario.

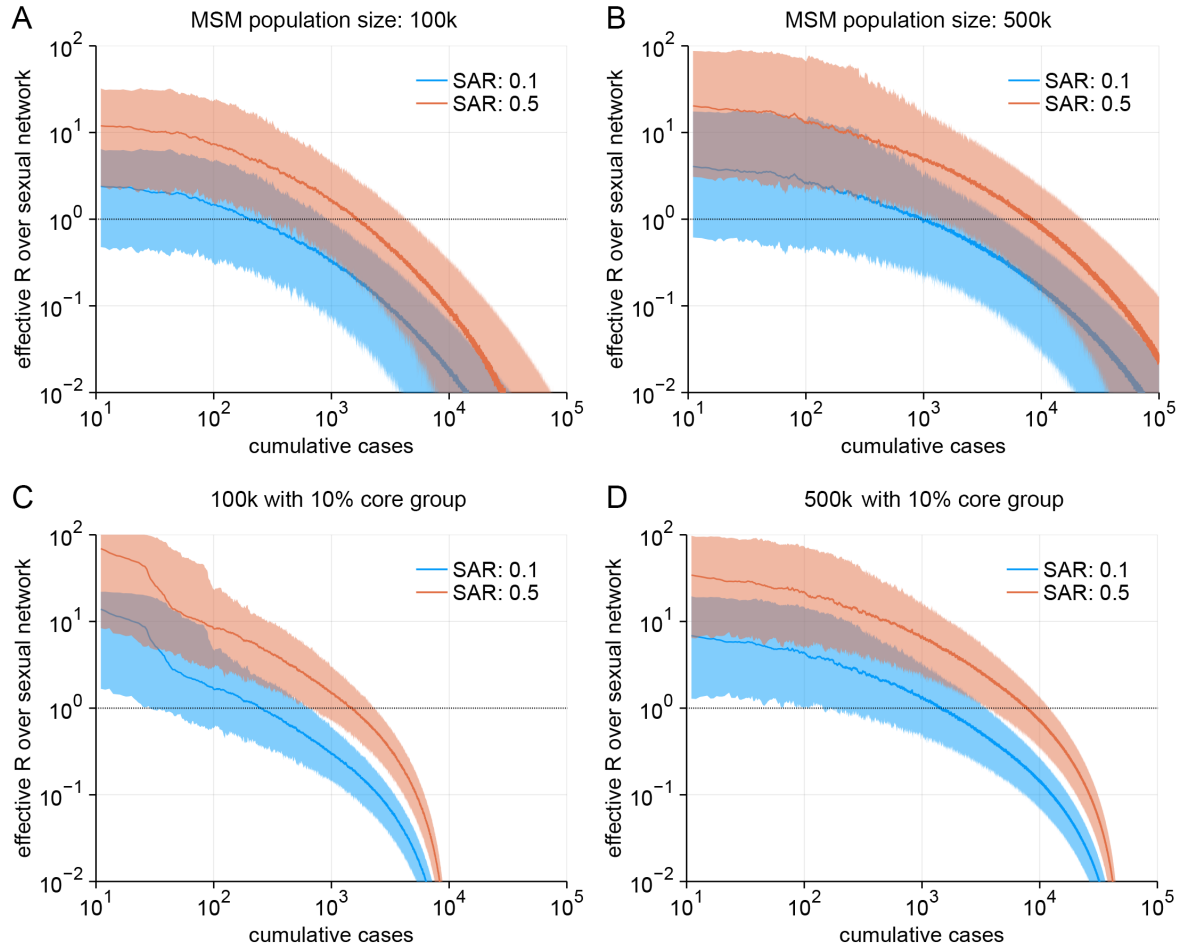

Fig. S5. Effective reproduction number of monkeypox over the MSM sexual contact network in the presence of depletion of susceptibles. (A) (B) The effective reproduction number after observing given cumulative cases in the MSM population of (A) size 100,000 and (B) 500,000. Lines and ribbons represent

median and 95% range. (C) (D) The effective reproduction number for the top 10% exclusive core group after observing given cumulative cases, where the entire MSM population size is (C) 100,000 and (D) 500,000.

Bigger picture of MSM partnership distributions with different levels of tail.

The distributions of MSM partnership with different level of tail (limited to the range $x \geq 5$ in Figure 2E in the main text) are shown for the range $x \geq 1$ (Figure S6). The body part of the distribution is minimally affected between different levels of the length of the distribution tail.

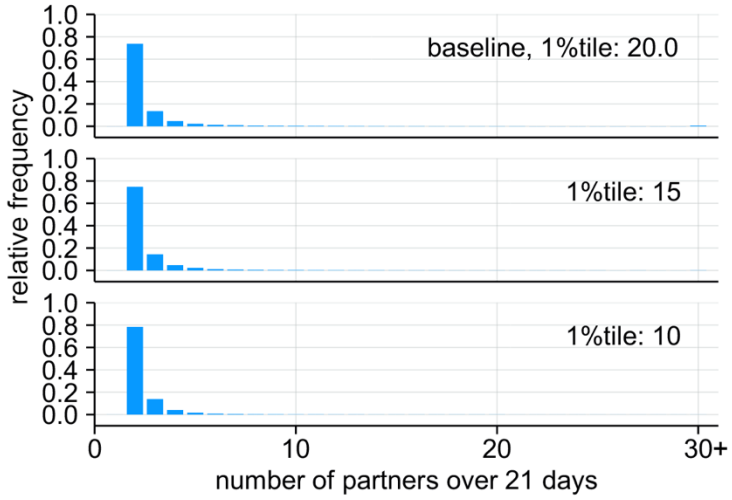

Fig. S6. Modelled 21-day effective sexual partnership distributions among MSM with different levels of distribution tail. Histograms represent modified Weibull distributions under interventions focusing on those with highest numbers of partners (with the upper 1st percentile of 15 and 10). The original distribution fitted to the Natsal datasets (upper 1st percentile: 20.0) is shown as a baseline. See Figure 2E for expanded histograms limited to the range $x \geq 5$.