Supplementary material for Infectious disease dynamics and restrictions on social gathering size

Christopher B Boyer[1,*], Eva Rumpler[1,*], Stephen M Kissler[*], and Marc Lipsitch[*]

[*]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

August 30, 2022

A Appendix

A.1 Theory

A.1.1 Derivation of relationship between expected cases and gathering size

Assuming fixed transmission probability τ for contacts between susceptibles and infectious individuals, the number of secondary cases generated by a single infectious individual with $K_s = k_s$ susceptible contacts is binomially distributed, i.e.

$$X_{it} \mid K_s = k_s \sim \text{Binomial}(k_s, \tau). \tag{A.1}$$

If susceptibles, infectious, and recovered individuals attend gatherings at rates equivalent to their population proportions then attendance at a gathering of size $K = k$ can be represented by a multinomial sampling model of the form

$$(K_s, K_i, K_r)' \mid K = k \sim \text{Multinomial}(k, (p_s, p_i, p_r)') \tag{A.2}$$

where $p_s = \frac{S(t)}{N}$, $p_i = \frac{I(t)}{N}$, and $p_r = \frac{R(t)}{N}$. Under the model, the expected number of susceptibles is kp_s, the expected number of infectious is kp_i, and the expected number of recovereds is kp_r. To calculate the expected number of secondary cases, note that, on average, only the K_s susceptibles are at risk of infection and they are exposed to K_i infectious individuals. Then the probability that the K_s susceptibles "escape", i.e. that they are not infected by any of the K_i infectious individuals, is $(1-\tau)^{K_i}$ and thus, by extension, the probability that they are infected by at least one of the K_i infectious individuals in attendance is $1-(1-\tau)^{K_i}$. Therefore, the total number of secondary cases at a gathering of size $K = k$ is

$$\begin{aligned} (K_s, K_i, K_r)' \mid K = k &\sim \text{Multinomial}(k, (p_s, p_i, p_r)') \\ X \mid K_s = k_s, K_i = k_i &\sim \text{Binomial}(k_s, 1-(1-\tau)^{k_i}) \end{aligned} \tag{A.3}$$

and taking iterated expectations, the expected number of secondary cases given a gathering of size $K = k$ is simply

$$E(X \mid K = k) = kp_s(1-(1-\tau)^{kp_i}). \tag{A.4}$$

For a more intuitive way to think about this equation, notice that kp_s is the expected number of susceptibles and kp_i is the expected number of infectious individuals; the expression $kp_s(1-(1-\tau)^{kp_i})$ is then just the expected number of susceptibles times the probability of being infected by any of the infectious individuals who attend, where the latter is equivalent to the one minus the "escape" probability, i.e. the probability that no susceptible is infected by any of the infectious individuals expected to attend.

A.1.2 Binomial Approximation

More specifically, when $|\tau kp_i| \ll 1$ a Binomial approximation gives

$$(1-\tau)^{kp_i} \approx 1 - kp_i\tau$$

and thus

$$kp_s(1-(1-\tau)^{kp_i}) \approx kp_s(1-(1-kp_i\tau)) \approx k^2 p_s p_i \tau$$

A.2 Sensitivity analyses

In the main analysis, for simplicity of presentation and to fix concepts, we consider fixed values of τ as well as p_s, p_i, and p_r. While we attempted to use values consistent with the literature for COVID-19, in practice these may vary across settings and for different pathogens. Here we conduct a series of sensitivity analyses to explore the role that these parameters play in determining impact of gathering size restrictions. In the main text, we assumed that the distribution of gatherings follows a discrete power law distribution. Here, we instead consider another heavy-tailed distribution, the log-normal distribution and show it fits to the various empirical data.

A.2.1 Varying τ

The probability of transmission given contact, τ, can vary either in constant value —for instance, if a new variant emerges that is more infectious— or more often it may simply be heterogeneous across settings —for example, a crowded indoor gathering versus an outdoor gathering. In the case of the former, in Figure A.1 we range τ over a suitable range, for instance 0.01 to 0.25 and find that our results are not substantially changed. We chose 0.01 as a lower bound, 0.08 as a medium

value from the secondary attack rate during meals (Bi et al., 2020), and 0.25 as a higher bound from the secondary attack rate in households during the Omicron wave (Jørgensen et al., 2022). In the case of the latter, we can consider τ to be drawn from a distribution reflecting the population of gathering settings at any given time. Given that τ must be between 0 and 1, a natural starting point for incorporating heterogeneity in τ into our prior model is to draw it from a beta distribution, i.e.

$$\begin{aligned} \tau &\sim \text{Beta}(\mu, \phi) \\ (K_s, K_i, K_r)' \mid K = k &\sim \text{Multinomial}(k, (p_s, p_i, p_r)') \\ X \mid K_s = k_s, K_i = k_i, \tau &\sim \text{Binomial}(k_s, 1 - (1 - \tau)^{k_i}) \end{aligned} \tag{A.5}$$

where here we've parameterized it such that μ is the mean of the beta distribution and ϕ is a dispersion parameter representing how concentrated values of τ are around the mean[1]. For now, we assume that τ is independent of gathering size. Figure A.2 shows example draws from beta distributions with same value of τ but different dispersion. As the dispersion parameter increases, τ is increasingly concentrated around the mean, as it decreases τ is more variable. In practice, increasing variability in τ would indicate that there are a small number of gatherings with very high transmission and a larger number with little to no transmission.

We show the effects of dispersion in τ on our conclusions regarding gathering size restrictions in Figure A.3. Given that our derivation of equation A.4 relies only on the mean of τ and we assume that tau is independent of gathering size, we should expect that the expected reduction in incident cases for a k_{max} restriction is unchanged by varying ϕ as

$$E(X \mid K = k) = kp_s(1 - (1 - E(\tau))^{kp_i}) = kp_s(1 - (1 - \mu))^{kp_i}). \tag{A.6}$$

Practically, this implies that as long as τ is independent of gathering size (and the mean value of τ is defined) our long run conclusions about the effect of gathering size restrictions is unchanged. However, as highlighted by the shaded regions of the 95% simulation intervals, increasing variability in τ leads to greater variance in the effect of restrictions. In terms of policy-making, if there's strong evidence for heterogeneity in τ decision-makers may want to consider planning with these intervals

[1]The canonical definition of the beta distribution is in terms of shape parameters α and β where $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$. Here we use the following transformation $\mu = \frac{\alpha}{\alpha+\beta}$ and $\phi = \alpha + \beta$, where ϕ is sometimes also called the sample size.

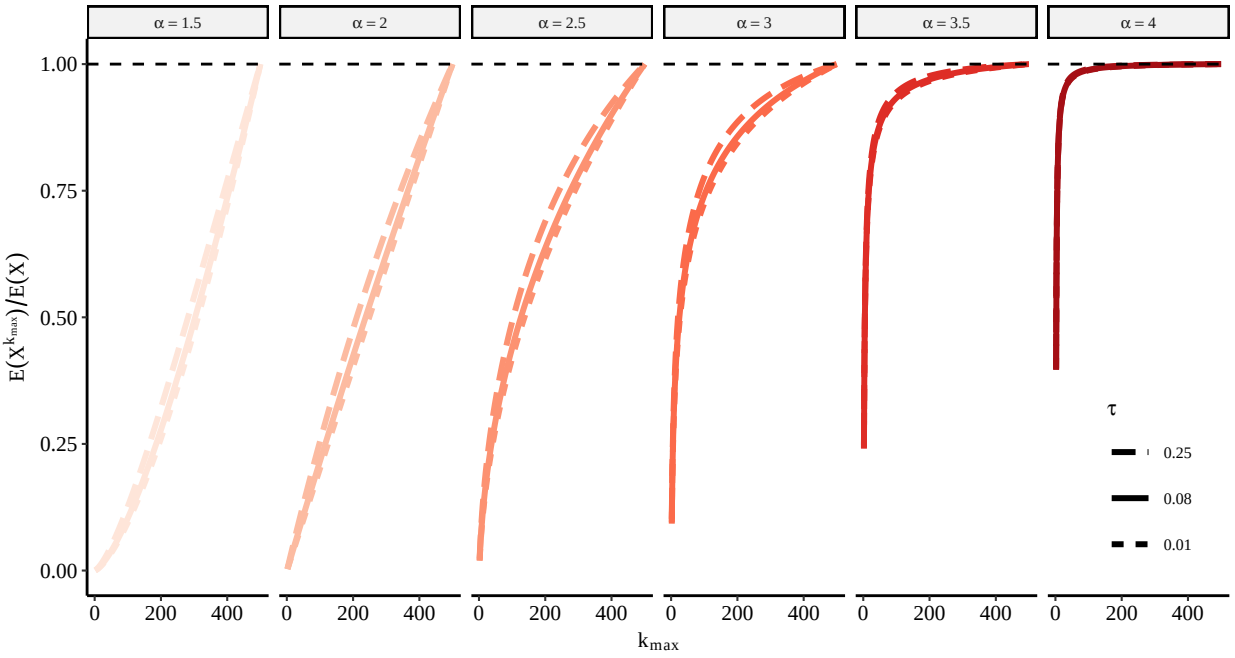

Figure A.1: Effect of different values of constant τ on relative rate of incident cases under k_{max} gathering size restrictions for different power law distributions. Here we re-create Figure 2 but vary the values of the constant τ from 0.01 to 0.25.

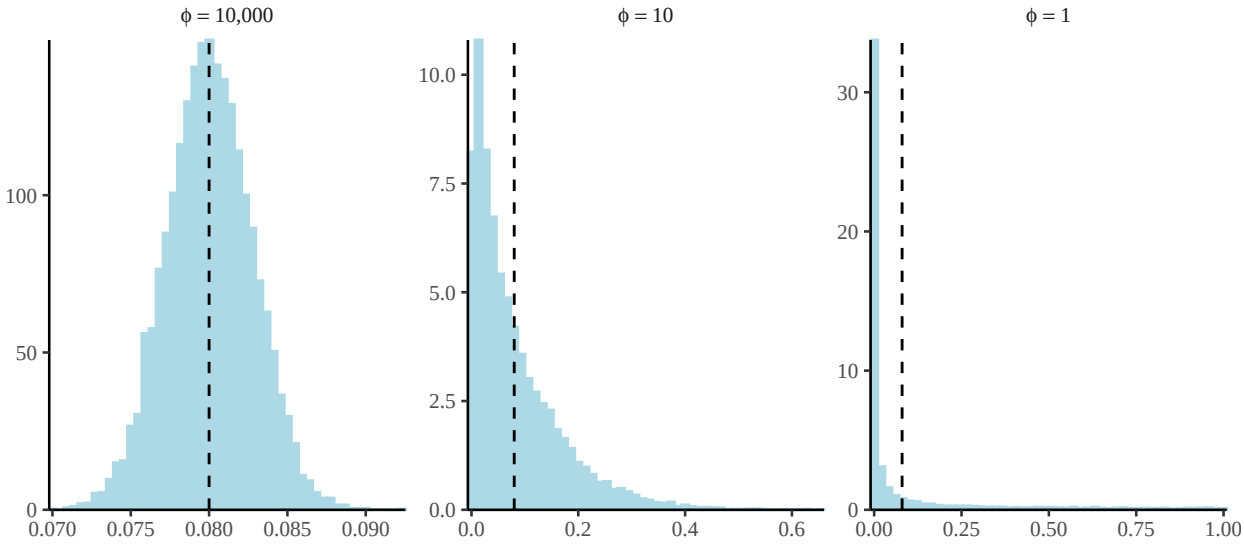

Figure A.2: Examples of the beta distribution with $\mu = 0.08$ under different values of ϕ. Based on 10,000 draws from the beta distribution with μ fixed to 0.08; dashed line shows the position of μ.

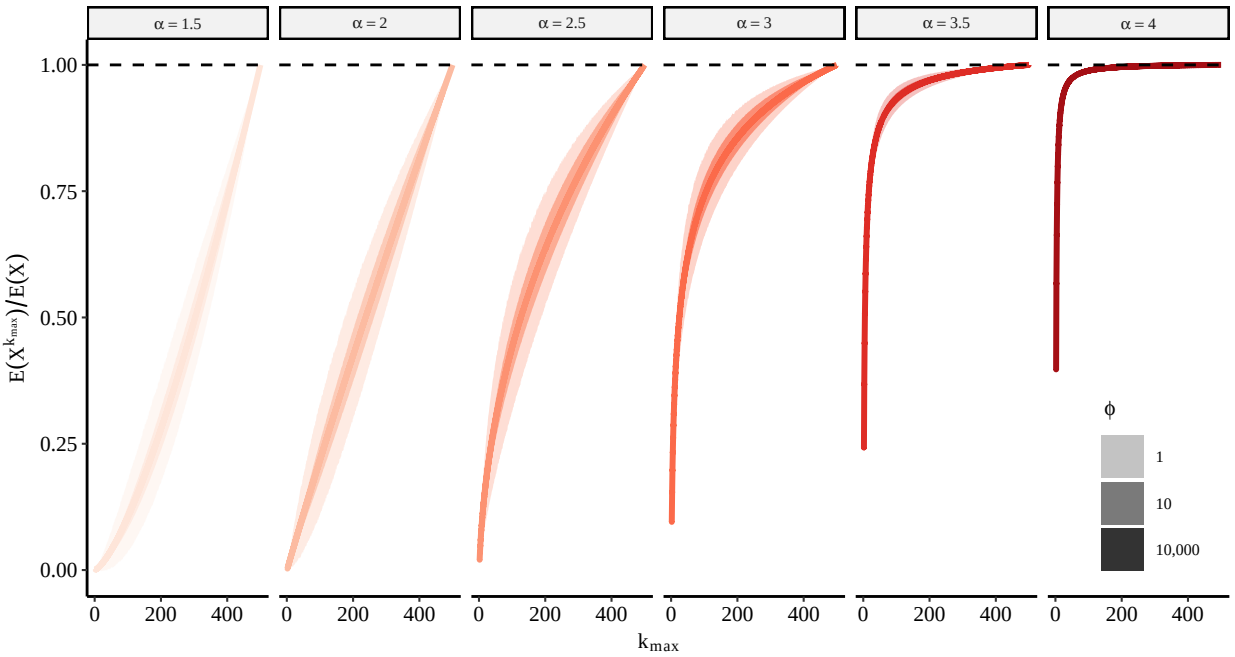

Figure A.3: Effect of dispersion in τ on relative rate of incident cases under k_{max} gathering size restrictions for different power law distributions. Here we fix transmission parameters to following values $p_i = 0.01$, $p_s = 0.99$, but allow τ to vary. We draw 10,000 values of τ from a beta distribution with $\mu = 0.08$ and varying ϕ. As previously, the figure shows the relative rate of incident cases calculated using equation 6 and comparing restrictions with k_{max}-level thresholds to unrestricted rate (e.g. a value of 0.5 implies a 50% fewer per capita incident cases relative to unrestricted rate). The lines represent the mean relative rate across all simulations while the shaded areas show the 95% simulation intervals.

in mind (e.g. using upper bound from an appropriately chosen interval such that there's an p% chance that restriction leads to reduction of desired size).

Finally, we consider what happens when τ is allowed to vary with the size of gathering. Given the paucity of data, it's unclear what one might expect the relationship to be between τ and gathering size *a priori*. On the one hand, τ could increase with gathering size if larger gatherings tend to be longer or in settings more conducive to spread. On the other hand, one could make an equally compelling case that smaller gatherings, which may be in more intimate settings may have higher τ. In their paper describing the Copenhange Network Study data, (Sekara et al., 2016) find no association between the size and duration of gatherings, suggesting no relationship on at least one proxy for τ. Absent reliable sources, here we vary the relationship across three representative scenarios: (1) τ decreases with gathering size K, (2) τ independent of gathering size K, τ increases with gathering size K. To keep it simple, we group gatherings into three

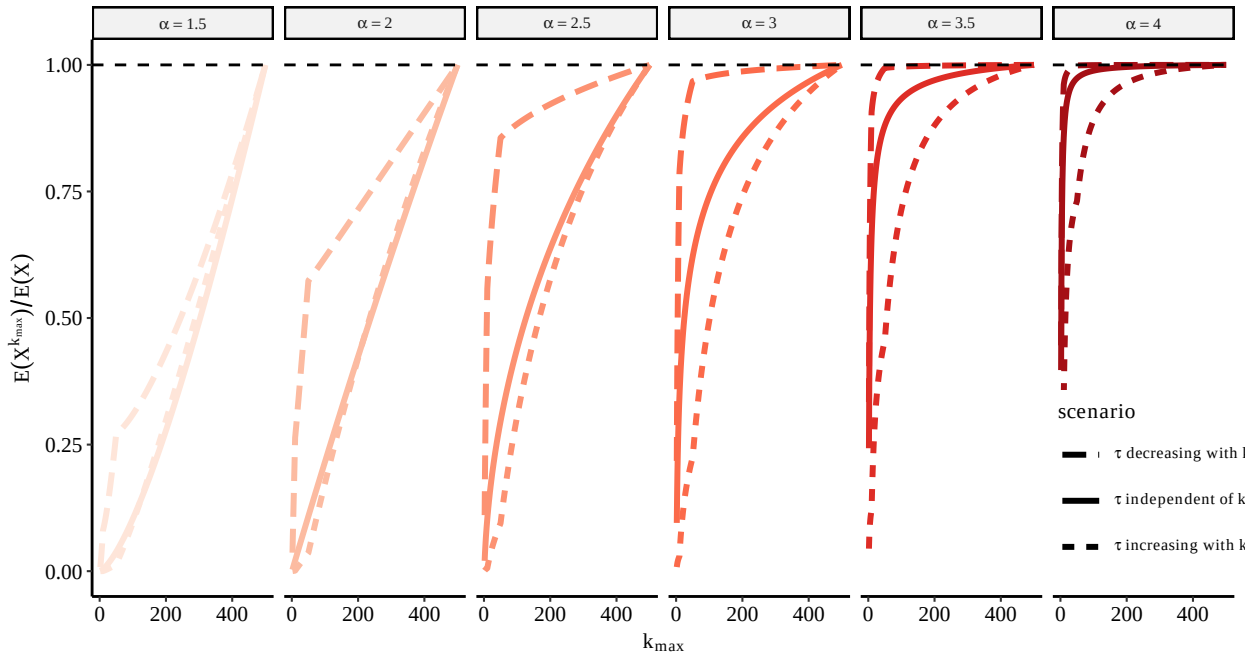

Figure A.4: Effect of relationship between τ and gathering size k on relative rate of incident cases under k_{max} gathering size restrictions for different power law distributions. Here we consider three scenarios for how the transmission probability τ varies with gathering size k. For τ decreasing with k, $\tau = 0.25$ for gatherings with less than 10 people, $\tau = 0.08$ for gatherings with 11 to 50 people, and $\tau = 0.01$ for gatherings with more than 51 people. For τ increasing with k, $\tau = 0.01$ for gatherings with less than 10 people, $\tau = 0.08$ for gatherings with 11 to 50 people, and $\tau = 0.25$ for gatherings with more than 51 people. The solid line is the reference case where $\tau = 0.08$ at all sizes.

mostly arbitrary groupings (less than 10 people, 11 to 50 people, and larger then 51 people) and choose τ for each from $\tau \in \{0.01, 0.08, 0.25\}$. Figure A.4 shows the results. Relative to our results presented in the main text, represented by the solid line where τ is independent of k, if τ decreases with gathering size our estimates are too optimistic and harsher restrictions would be required to achieve the same reductions. If τ increases with gathering size, then our estimates are too pessimistic and comparable reductions could be achieved with looser restrictions on size.

A.2.2 Varying p_i and p_r

The proportion of susceptible, infected and recovered individuals in each gathering are inputs to our gathering size restriction equation. In the main analysis, we fix values of p_s, p_i and p_r to 0.99, 0.01 and 0 respectively, which seemed reasonably illustrative for our purposes. However, in sensitivity analyses presented below, we vary the values of p_i between 0.001, 0.01 and 0.1; and that

of p_r between 0, 0.25 and 0.75. Across all nine combinations of p_i and p_r, the value of p_s is equal to $1 - (p_i + p_r)$ and thus varies from 0.15 to 0.999.

As shown in Figure A.5, for different power law distributions, the value of p_r seems to have only a small impact on the relative rate of incident infections comparing a scenario with restrictions to scenarios without. However, values of p_i seem to have a larger effect, especially for power law distributions with smaller α values. For very high proportions in infected individuals (i.e. $p_i = 0.1$), a given gathering size restriction leads to a lower relative reduction in the number of incident cases, when compared to lower proportion on infected individuals (i.e. $p_i = 0.01$ or 0.001). This makes sense as with such high proportion of infected attendees, and under distributions of gathering sizes where larger gatherings are frequent, most large gatherings will lead to a substantial number of infections. It is only when considering distribution of gatherings where larger gatherings are rare (i.e. $\alpha = 3.5$ or 4) that the proportion of infected individual attending p_i matters less, as in smaller gatherings less individuals can be newly infected.

As shown in Figure A.6, across all five empirical distribution of gathering sizes, the values of p_i and p_r have very limited impact on the the relative rate of incident cases under restriction to gatherings of size k_{max} compared to that in absence of restrictions. Indeed, the dashed lines representing the varying values of p_i overlap in most plots, only showing a slight shift when using the distribution derived from CNS data source. Similarly, ranging values of p_r also appears to have minor impact.

Overall, these results suggest that the proportion of susceptible, infected and recovered individuals attending gatherings has limited to no effect on our relative reduction in incident cases when implementing restrictions. This suggests that our results may be robust to those parameters that our conclusion may hold for various stages of epidemics.

A.2.3 Alternative gathering size distributions

In the main text we illustrate our theoretical points using the discrete power law distribution and then follow them up with results using empirical distributions. In practice we find that, while informative from a theoretical standpoint, a power law may not provide the best fit empirically. Therefore, when informing policy, rather than intuition, we encourage the use of realistic and preferably empirically-derived distributions. However, such data may not always be available and

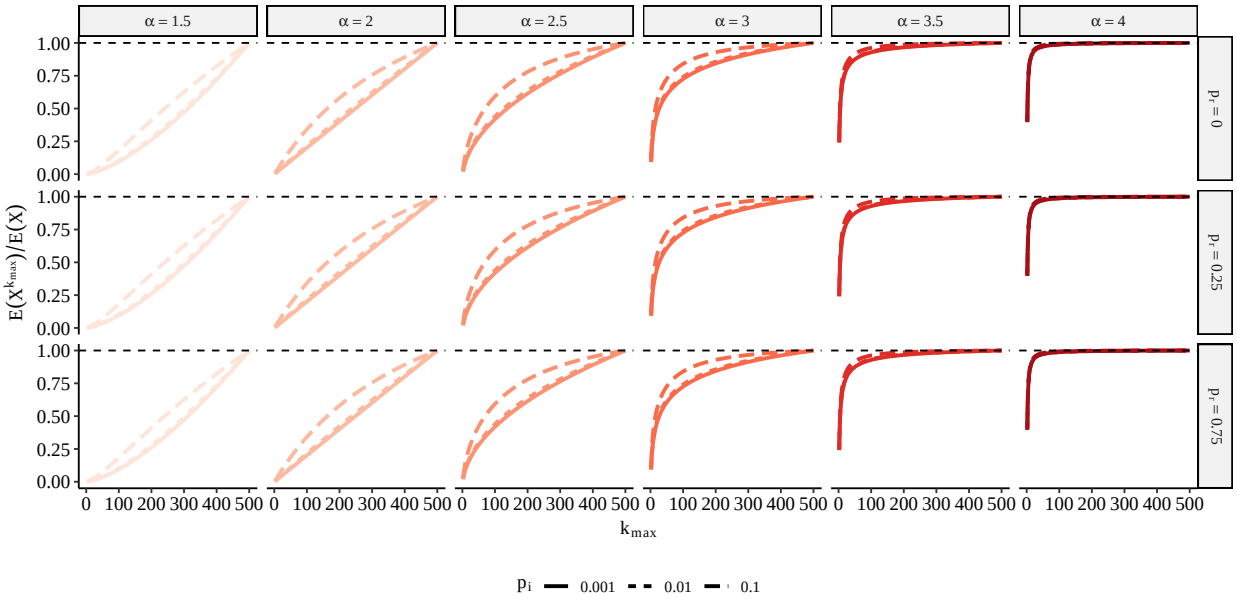

Figure A.5: Relative rate of incident cases under restriction which prohibits gatherings above size k_{max} for different power law distributions and varying the values of p_i and p_r. The value of p_i varies between 0.001, 0.01 and 0.1. That of p_r varies between 0, 0.25 and 0.75 and p_s is equal to $1 - (p_i + p_r)$. τ is equal to 0.08. Similarly to Figure 2, we assume that power law behavior starts at $k_{min} = 1$ and truncate the power law above gatherings of size 500 both to make the sum tractable and given that gathering sizes must at minimum be less than population size. The figure shows the relative rate of incident cases calculated using equation 6 and comparing restrictions with k_{max}-level thresholds to unrestricted rate (e.g. a value of 0.5 implies a 50% fewer per capita incident cases at time t relative to unrestricted rate).

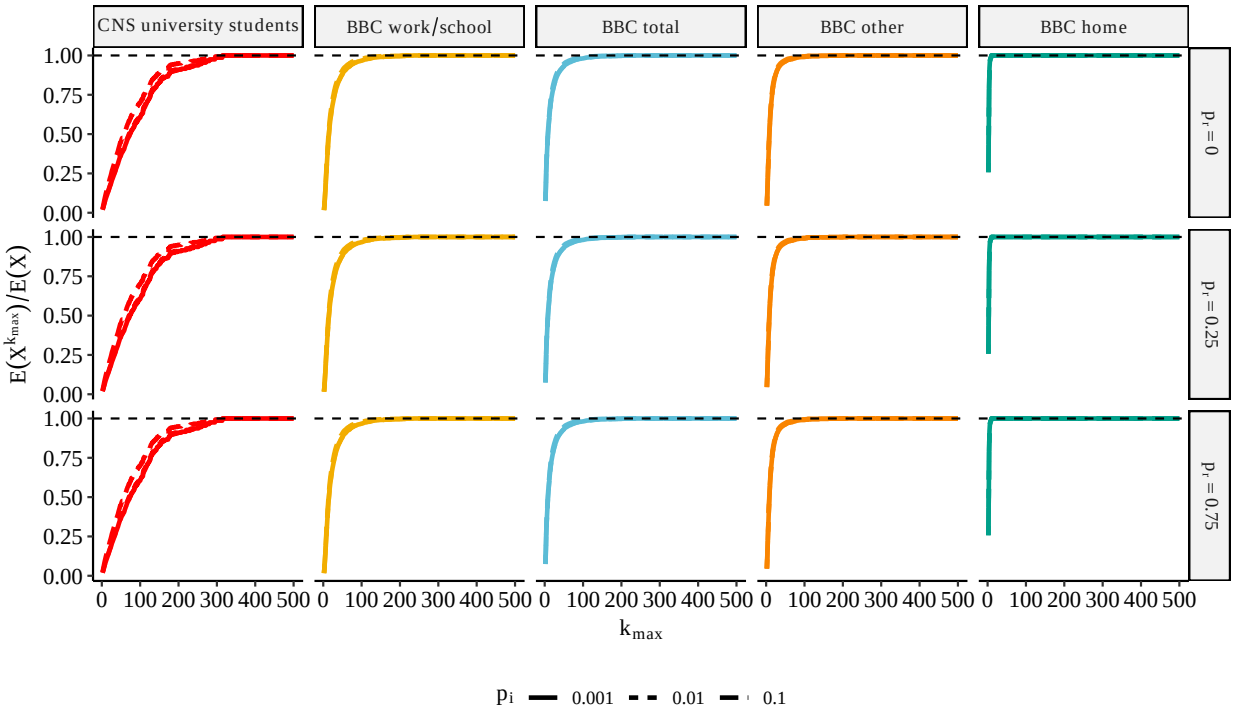

Figure A.6: Relative rate of incident cases under restriction which prohibits gatherings above size k_{max} using different data sources for the distributions of gathering sizes and varying the values of p_i and p_r. The value of p_i varies between 0.001, 0.01 and 0.1. That of p_r varies between 0, 0.25 and 0.75 and p_s is equal to $1 - (p_i + p_r)$. τ is equal to 0.08. Similarly to Figure 5, we use draws from empirical distributions. The figure shows the relative rate of incident cases calculated using equation 6 and comparing restrictions with k_{max}-level thresholds to unrestricted rate (e.g. a value of 0.5 implies a 50% fewer per capita incident cases at time t relative to unrestricted rate).

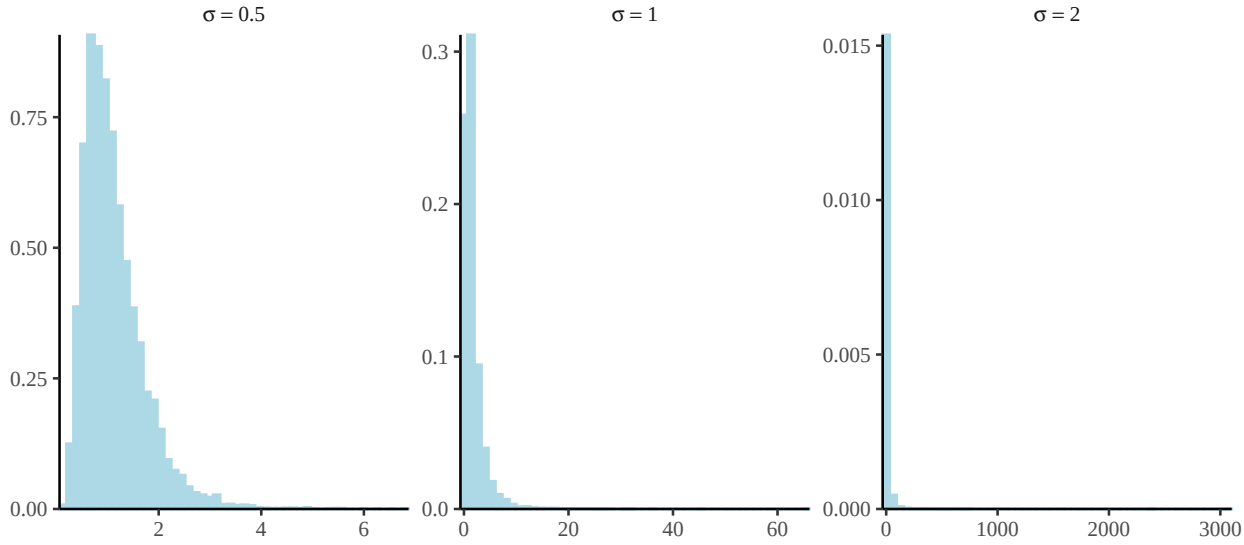

Figure A.7: Examples of the log-normal distribution with $\mu = 0$ under different values of σ. Based on 10,000 draws from the beta distribution with μ fixed to 0.

thus for completeness here we also consider other heavy-tailed distributions. A relatively common heavy-tailed alternative is the log-normal distribution, i.e.

$$f(k) = \frac{1}{k\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log k - \mu)^2}{\sigma^2}\right\},$$

where the parameters μ and σ^2 describe the mean and the variance of a log-transformed normal random variable. Figure A.7 provides some examples of parameter values with increasing tail mass.

As in the main text, we fit the log-normal distribution to the empirical distributions from the BBC Pandemic and Copenhagen Networks Study using maximum likelihood. We again consider the possibility that the empirical data may only follow a log-normal distribution above a certain lower threshold (k_{min}). Figure A.8 shows the best fitting log-normal distributions and their parameter values. Visually the log-normal distribution seems like a better fit than the discrete power law considered in the main text, mostly because there does seem to be some nonlinear drop-off in the extreme tails. However, this could also reflect influence of measurement error[2] and sampling variability in the extreme tail. We can test this using Vuong's likelihood ratio test which compares the Kullback-Leibler criteria for the two fits. In Table A.1, we find strong evidence that the log-

[2]For instance, BBC Pandemic dataset has individuals record their daily contacts. It's likely that they are better at estimating small group sizes relative to big ones and may lump or round estimates for larger groups.

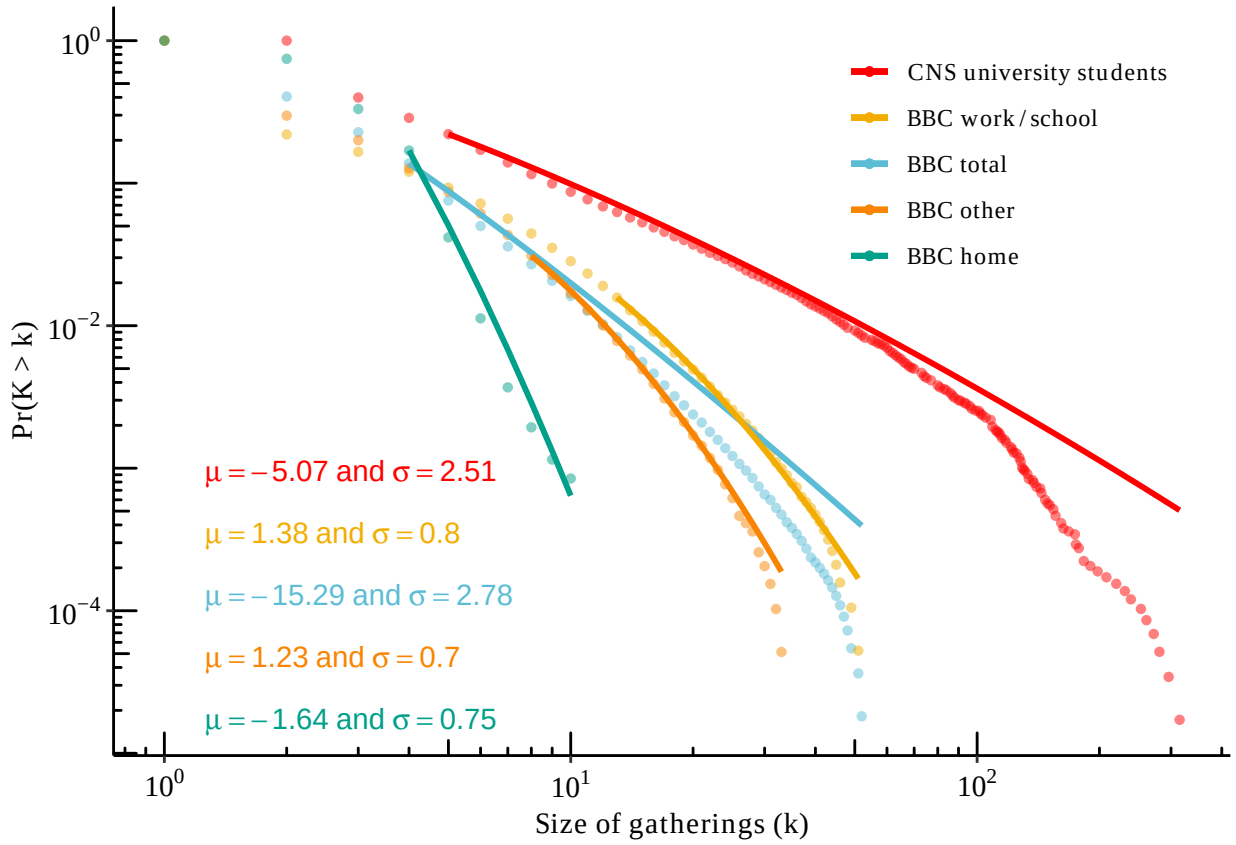

Figure A.8: Estimates of log-normal parameters for the Copenhagen Networks Study (CNS) and the BBC Pandemic study by setting. Plot is complementary cumulative distribution function versus gathering size with lines showing fitted power law distribution. Estimates for α and k_{min} obtained using maximum likelihood for discrete power law using the `poweRlaw` package in R.

normal is a better fit for all but the household contacts.

Table A.1: Vuong's test comparing discrete log-normal and power-law distribution fits to empirical data on gathering size.

Data source	R-statistic	p-value
BBC Pandemic		
Home	1.24	0.216
Work / school	6.66	<0.001
Other	5.37	<0.001
Total	4.20	<0.001
Copenhagen Networks Study	5.89	<0.001

Notes: R-statistic is the ratio of log-likelihoods of the discrete log-normal and power-law fits, positive values favor the log-normal distribution. p-values are for two-sided hypothesis that log-normal is a better fit.

References

Bi, Q., Wu, Y., Mei, S., Ye, C., Zou, X., Zhang, Z., Liu, X., Wei, L., Truelove, S. A., Zhang, T., Gao, W., Cheng, C., Tang, X., Wu, X., Wu, Y., Sun, B., Huang, S., Sun, Y., Zhang, J., ... Feng, T. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *The Lancet Infectious Diseases*, *20*(8), 911–919. https://doi.org/10.1016/S1473-3099(20)30287-5

Jørgensen, S. B., Nygård, K., Kacelnik, O., & Telle, K. (2022). Secondary Attack Rates for Omicron and Delta Variants of SARS-CoV-2 in Norwegian Households. *JAMA*, *327*(16), 1610. https://doi.org/10.1001/jama.2022.3780

Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the National Academy of Sciences*, *113*(36), 9977–9982. https://doi.org/10.1073/pnas.1602803113