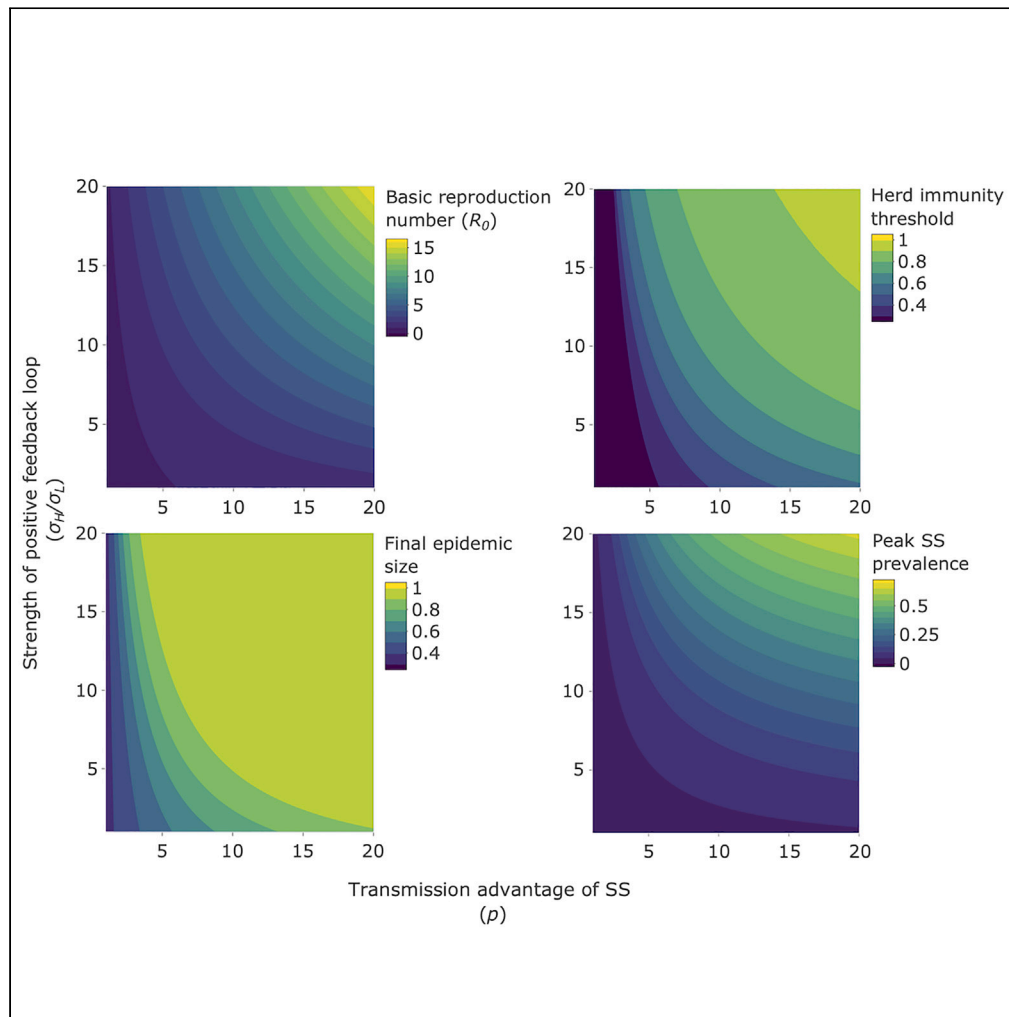# iScience

**Article**

# Positive feedback loops exacerbate the influence of superspreaders in disease transmission



Klara M. Wanelik,
Mike Begon, Andy
Fenton, Rachel A.
Norman, Pablo M.
Beldomenico

klara.wanelik@biology.ox.ac.
uk

### Highlights

We develop a model
where superspreaders are
more likely to generate
superspreaders

Results suggest such
positive feedback can
profoundly affect
epidemic outcomes

Affected outcomes
include final epidemic
size, herd immunity
threshold, and $R_0$

Article

# Positive feedback loops exacerbate the influence of superspreaders in disease transmission

Klara M. Wanelik,[1,2,5,*] Mike Begon,[1] Andy Fenton,[1] Rachel A. Norman,[3] and Pablo M. Beldomenico[4]

## SUMMARY

**Superspreaders are recognized as being important drivers of disease spread. However, models to date have assumed random occurrence of superspreaders, irrespective of whom they were infected by. Evidence suggests though that those individuals infected by superspreaders may be more likely to become superspreaders themselves. Here, we begin to explore, theoretically, the effects of such a positive feedback loop on (1) the final epidemic size, (2) the herd immunity threshold, (3) the basic reproduction number, $R_0$, and (4) the peak prevalence of superspreaders, using a generic model for a hypothetical acute viral infection and illustrative parameter values. We show that positive feedback loops can have a profound effect on our chosen epidemic outcomes, even when the transmission advantage of superspreaders is moderate, and despite peak prevalence of superspreaders remaining low. We argue that positive superspreader feedback loops in different infectious diseases, including SARS-CoV-2, should be investigated further, both theoretically and empirically.**

## INTRODUCTION

Pathogen transmission requires an appropriate contact between a susceptible and infected individual, such that the susceptible individual is exposed to pathogens shed by the infected individual. Most traditional compartmental models of infectious disease dynamics assume that, on average, all individuals behave in the same way.[1] However, it has long been known that some individuals, termed "superspreaders", contribute far more to pathogen transmission, and epidemic spread, than others.[2–5] Superspreaders may be defined as individuals in the upper 1% tail of the distribution of pathogen transmission,[6] but the term has been more widely applied to those with a strongly disproportionate contribution. The importance of accounting for this individual variation in epidemiological models has more recently been recognized, not least because it has played a major role in the ongoing COVID-19 pandemic.[7–13]

Superspreaders exist for most transmissible diseases of humans, livestock, and wildlife,[14] and may originate in two ways. First, "supercontactors" transmit infection to more individuals because they have a larger number of contacts than the average, such as in the transmission of HIV and other sexually transmitted diseases (STDs) in humans, through those who are more sexually active,[15] and of Sin Nombre hantavirus in deer mice.[16] Second, "supershedders" shed more infectious particles than average and so increase the probability of infection once an appropriate contact has been made, such as when humans co-infected with other STDs shed more HIV,[17] cattle shed *Escherichia coli* O157 heterogeneously,[18] or humans have heterogeneous SARS-CoV-2 viral loads.[19] Previous work has shown that model predictions accounting for superspreaders are very different from average-based approaches, with a higher probability of disease extinction and rarer but more explosive disease outbreaks.[6] However, for a closed, fixed-sized population, the formula for the final epidemic size is unchanged by the presence of superspreaders.[20]

Multiple studies have explored such heterogeneities in susceptibility and/or contacts and their consequences for disease transmission.[21–25] Standard models assume that superspreaders occur irrespective of who they were infected by,[20] but there is evidence from some systems of positive feedback, recently proposed as an explanation for the heterogeneous propagation pattern of COVID-19,[26,27] whereby individuals infected by superspreaders are more likely to be superspreaders themselves. For example, supershedders could generate further supershedders if a higher inoculum dose is more likely to overwhelm the mechanisms of resistance, resulting in poor control of viral replication, causing higher viral loads. This hypothesis has been supported by evidence from experiments in other systems[28–31] and, for SARS-CoV-2,

[1]Department of Evolution, Ecology and Behaviour, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK

[2]Department of Biology, University of Oxford, Oxford, UK

[3]Department of Computing Science and Mathematics, Faculty of Natural Sciences, University of Stirling, Stirling, UK

[4]Laboratorio de Ecología de Enfermedades, Instituto de Ciencias Veterinarias del Litoral (Consejo de Investigaciones Científicas y Técnicas - Universidad Nacional del Litoral), Esperanza, Argentina

[5]Lead contact

*Correspondence: klara.wanelik@biology.ox.ac.uk
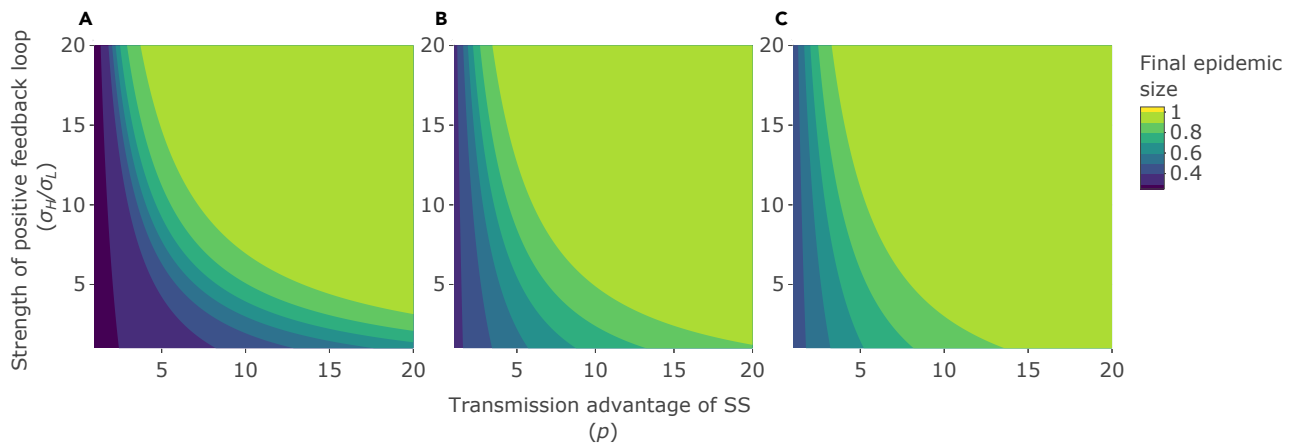
https://doi.org/10.1016/j.isci.2023.106618

**Figure 1. Effects on the final epidemic size**

Final epidemic size (where the maximum is 1, reflecting everyone in the population becoming infected) as the strength of the positive feedback loop ($\sigma_H/\sigma_L$; y axis), the transmission advantage of superspreaders (SS) ($p$; x axis), and the initial number of SS are varied: (A) initial number of SS = 0% of the total population size. (B) initial number of SS = 5% of the total population size. (C) initial number of SS = 10% of the total population size. See Table 1 for other parameter values.

from: infections arising from exposure to high doses being more likely to be symptomatic and to have higher intensity[32,33]; from cases arising from asymptomatic cases being more likely to be asymptomatic[34,35]; from viral loads driving the size and duration of COVID-19 clusters[36–38] and SARS-CoV-2 transmission[39]; and from the ratio of observed-to-expected superspreader-superspreader dyads being greater than expected by chance.[40]

Despite this recognition that superspreaders may generate superspreaders for diseases like COVID-19, no previous study has modeled such positive feedback loops, and hence the role they play in driving the epidemiology of such diseases remains unknown. In this study, therefore, we develop a generic model (for a hypothetical acute viral infection, like SARS-CoV-2 or MERS-CoV) to begin to explore, theoretically, how positive feedback loops affect (1) the final epidemic size, (2) the herd immunity threshold, (3) the basic reproduction number, $R_0$, and (4) the peak prevalence of superspreaders, of an emerging epidemic. By doing so, we hope to stimulate further work, both theoretical and empirical, on this potentially important phenomenon.

## RESULTS

### Modeling positive superspreader feedback loops

To understand the role of the positive feedback loop between superspreaders, we developed a generic model for a hypothetical acute viral infection. The rationale underlying the formulation of this model is based on the hypothesis (detailed above) that supershedders could generate further supershedders. The model includes four classes of hosts: susceptible hosts, infected hosts with low-titre infections (non-superspreaders), infected hosts with high-titre infections (superspreaders; with a transmission advantage over non-superspreaders of magnitude $p$), and recovered hosts.

The model includes two routes for the generation of superspreaders. It assumes that some proportion of infections ($\sigma_L$) from a non-superspreader result in a superspreader—representing the background, spontaneous generation of superspreaders. There are multiple mechanisms by which this could happen. For example, through prolonged contact between the non-superspreader and the recipient, or through the recipient being more susceptible to infection (due to e.g. genetics or an underlying medical conditions).[41] The model also assumes that some proportion of infections ($\sigma_H$) from a superspreader result in another superspreader. Both proportions ($\sigma_H$ and $\sigma_L$) vary between 0 and 1, but if $\sigma_H > \sigma_L$, then superspreaders are more likely to generate new superspreaders, and hence, we use $\sigma_H/\sigma_L$ throughout to represent the strength of the positive feedback loop of superspreader infections.

The full mathematical description of the model and details of its implementation are available in the STAR Methods. The rationale behind the choice of parameter values and initial conditions are also discussed in the STAR Methods.

**Table 1. Model parameters and their values**

| Parameter | Symbol | Per capita rate | Comment |
|---|---|---|---|
| Death rate due to disease | $\alpha$ | 0 | For simplicity, assumed no excess death due to infection. |
| Recovery rate | $\gamma$ | 0.1 | Assumed an average infectious period of one week, as acute viral infections like SARS-CoV-2[38] and influenza[42] have infectious periods of approximately one week. |
| Baseline transmission rate arising from non-superspreaders ($L$-infected hosts) | $\beta_L$ | $8 \times 10^{-6}$ | Used an arbitrary $\beta_L$ that put the baseline $R_0$ in the absence of superspreading (i.e. when $p = 1$, $\sigma_H = \sigma_L$) below 1 (see STAR Methods; for the values used here, $R_0^L = 0.8$). From this basis, we can explore whether superspreading and/or a positive feedback loop has the potential to drive an epidemic that would not otherwise occur. |
| Proportion of infections from a non-superspreader ($L$-infected host) that result in a superspreader ($H$-infected host) | $\sigma_L$ | 0.05 | Assumed that 5% of infections from a non-superspreader result in a superspreader. |
| The number of times more superspreaders ($H$-infected hosts) generated by a superspreader ($H$-infected host) than a non-superspreader ($L$-infected host), i.e. the strength of the positive feedback loop | $\sigma_H/\sigma_L$ | 1–20 | Varied from 1 (equal generation of superspreaders by superspreaders and non-superspreaders) to the maximum feasible value of 20 (20 times more superspreaders generated by a superspreader than a non-superspreader). This maximum feasible value arises since $\sigma_H$ cannot exceed 1; given the default value of $\sigma_L = 0.05$, $\sigma_H/\sigma_L$ cannot exceed 20. |
| The number of times by which the transmission rate from a superspreader ($H$-infected host) is greater than the transmission rate from a non-superspreader ($L$-infected host; the transmission advantage of superspreaders) | $p$ | 1–20 | Varied from 1 (equal transmission rate from a superspreader and a non-superspreader) to 20 (20 times higher transmission rate from a superspreader than from a non-superspreader). This upper bound is consistent with literature on the superspreading of aerosols for acute viral infections, like SARS-CoV-2[43] |

### Epidemic outcomes reach their highest values when both superspreader advantage and positive superspreader feedback are high

The final epidemic size (where the maximum is 1, reflecting everyone in the population becoming infected; Figures 1A–1C), the herd immunity threshold (the proportion of the population that needs to be immune to achieve herd immunity; Figures 2A–2C), and the basic reproduction number, $R_0$ (Figure 3), all increase with increases in the transmission advantage of superspreaders, $p$, and with the strength of the positive feedback loop whereby superspreaders generate further superspreaders, $\sigma_H/\sigma_L$. Hence, all these epidemic outcomes reach their highest values when both $p$ and $\sigma_H/\sigma_L$ are high.

### Positive superspreader feedback loops can have little or no effect on epidemic outcomes in some scenarios

Of course, in the absence of superspreading ($p = 1$), all epidemiological measures are insensitive to $\sigma_H/\sigma_L$ since there is no superspreader advantage on which a feedback loop might act (along the vertical axes of the figures). Indeed, as shown in STAR Methods, if the pathogen is unable to sustain itself in the absence of
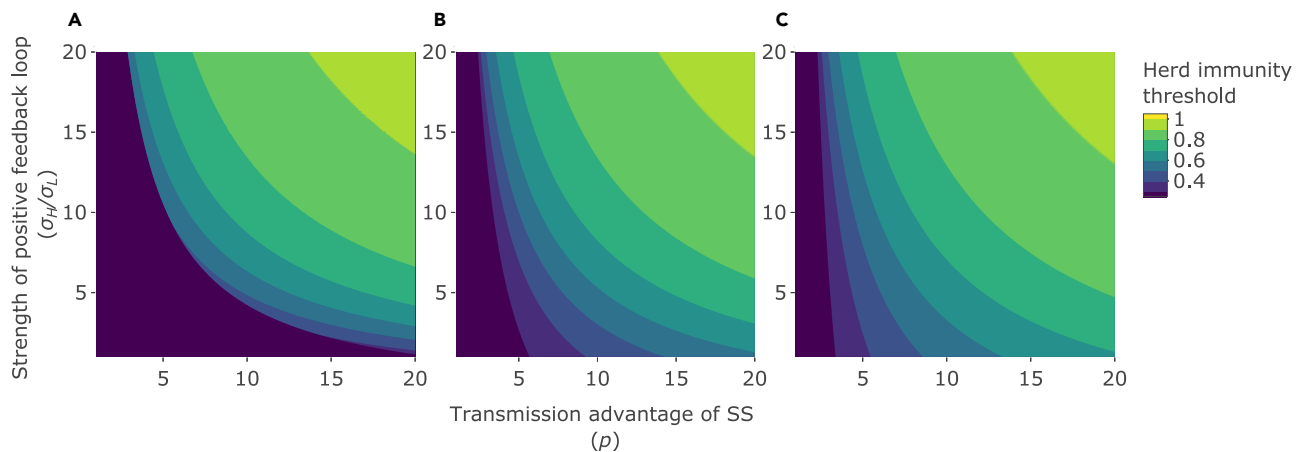
**Figure 2. Effects on the herd immunity threshold**

Herd immunity threshold (the proportion of the population that need to be immune to achieve herd immunity) as the strength of the positive feedback loop ($\sigma_H/\sigma_L$; y axis), the transmission advantage of superspreaders (SS) ($p$; x axis), and the initial number of superspreaders are varied: (A) initial number of SS = 0% of the total population size. (B) initial number of SS = 5% of the total population size. (C) initial number of SS = 10% of the total population size. See Table 1 for other parameter values.

superspreading (i.e. if $R_0 < 1$ for a pure low-titre pathogen), there is a lower threshold value of $p$, given by $\frac{(\alpha+\gamma)}{N\beta_L}$, below which no viable amount of positive feedback can drive $R_0 > 1$ (for our baseline parameter values, this lower threshold value of $p$ is 1.25, suggesting superspreaders need at least a 25% transmission advantage for an epidemic to occur, regardless of the extent of positive feedback).

### Positive superspreader feedback loops can have a profound effect on epidemic outcomes in other scenarios

Conversely, in the absence of a feedback loop, as assumed in previous studies ($\sigma_H/\sigma_L = 1$, along the horizontal axis in the figures), these predicted outcomes, especially $R_0$, may change little with superspreader advantage. For example, in the absence of feedback, and using the baseline parameter values to generate Figure 3, $R_0$ is less than 1 (the pathogen fails to cause an epidemic) unless $p$ exceeds 6 (see STAR Methods), and only reaches 1.6 as $p$ approaches 20 (the maximum illustrative value used here). But when there is feedback, and especially when the feedback loop is strong (e.g. $\sigma_H/\sigma_L = 20$ in Figure 3), $R_0$ rises rapidly, reaching 16 in Figure 3 as $p$ approaches 20, and causing an epidemic ($R_0 > 1$) with only very moderate levels of superspreader advantage ($p \geq 1.25$; see STAR Methods).

### Positive superspreader feedback loops can have a profound effect on epidemic outcomes even when the proportion of superspreaders in the population remains low

It is also noteworthy that this newly identified effect of the feedback loop on our chosen epidemic outcomes can be profound even when the proportion of superspreaders in the population remains low. For example, at intermediate levels of superspreader advantage ($p = 5$), increasing feedback loop strength from 1 to 5, an intermediate level of feedback, has a powerful effect on both final epidemic size (Figures 1A–1C) and $R_0$ (Figure 3), with $R_0$, for example, increasing from less than 1 (pathogen failing to cause an epidemic), to more than 1 (pathogen causing an epidemic). This is despite the peak prevalence of superspreaders (as a proportion of total population size) not exceeding ~0.5%–10%, depending on the initial number of superspreaders (Figures 4A–4C).

### DISCUSSION

Superspreaders are recognized as being important drivers of disease spread. However, models to date have assumed random generation of superspreaders, irrespective of whom they were infected by. Here, by contrast, and supported by a range of studies, we explore the importance of positive superspreader feedback loops, whereby those individuals infected by superspreaders are more likely to be superspreaders themselves. We show that superspreaders in their own right (i.e. without positive superspreader feedback), contrary to widespread perception, may have little effect on our chosen epidemic outcomes (namely, the final epidemic size, the herd immunity threshold, and the basic reproduction number, $R_0$), but when positive superspreader feedback is included, the effect may be profound, even when the
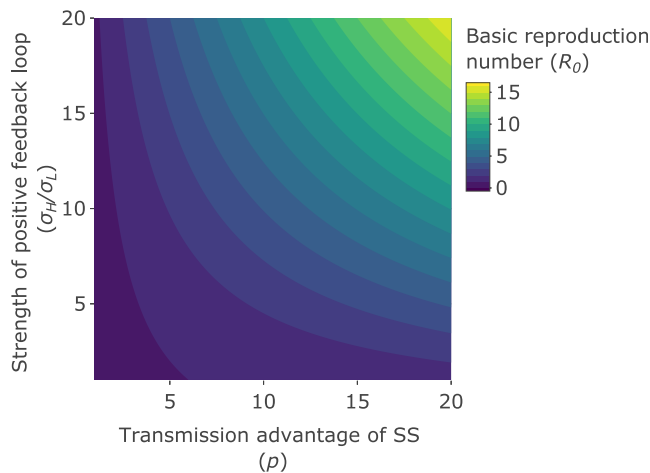
**Figure 3. Effects on the basic reproduction number**

The basic reproduction number ($R_0$) as the strength of the positive feedback loop ($\sigma_H/\sigma_L$; y axis) and the transmission advantage of superspreaders (SS) ($p$; x axis) are varied. See Table 1 for other parameter values.

transmission advantage of superspreaders is moderate, and despite peak prevalence of superspreaders remaining low. Thus, our theoretical framework formally supports and enhances the ideas proposed by Beldomenico,[26] whereby positive feedback loops have the potential to drive the heterogeneous propagation pattern of a range of infectious diseases, including SARS-CoV-2.

In the context of COVID-19, previous work exploring the effects of population heterogeneity on herd immunity has suggested that assortative mixing, whereby supercontactors tend to preferentially contact other supercontactors, has no effect on the herd immunity threshold.[24] Furthermore, heterogeneity itself may decrease the threshold, when the proportion of infected (and subsequently immune) individuals in subgroups with the highest contact rates is higher than in subgroups with low contact rates.[25] In contrast, here, we show that positive feedback loops, whereby supershedders tend to generate more supershedders, *increase* the herd immunity threshold, potentially to a significant degree. This difference reflects supershedding not being a fixed attribute of an individual in our model, such that superspreaders continue to be generated during the course of an epidemic, rather than being among the first to become immune (or die).

Our model is necessarily simple. Our aim was, as a first step, to identify the potential effect of positive superspreader feedback loops on a number of epidemic outcomes. Given the results from this initial study,
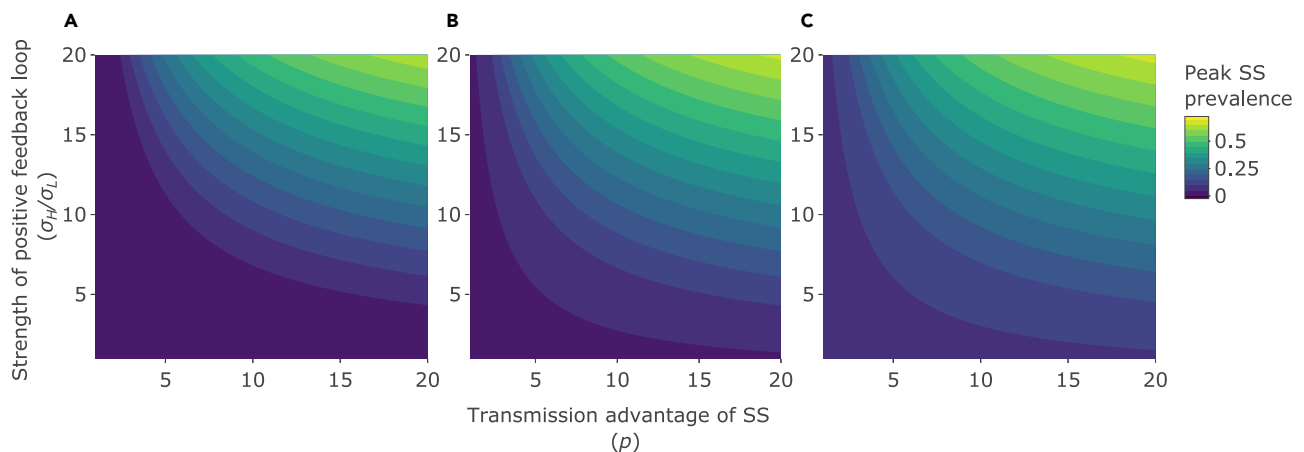


**Figure 4. Effects on the peak prevalence of superspreaders**

Peak prevalence of superspreaders (as a proportion of total population size, $N$) as the strength of the positive feedback loop ($\sigma_H/\sigma_L$; y axis), the transmission advantage of superspreaders (SS) ($p$; x axis), and the initial number of SS are varied: (A) initial number of SS = 0% of the total population size. (B) initial number of SS = 5% of the total population size. (C) initial number of SS = 10% of the total population size. See Table 1 for other parameter values.

we contend that the existence of positive feedback loops in different host-pathogen systems warrants further investigation. It will be important to test our model using empirical data, and to further develop our model theoretically, to quantify more precisely the effects of positive feedback loops, and to incorporate further biological details.

### Limitations of the study

We present results (generated entirely via numerical simulation) from a generic model, parameterized using illustrative values to represent a hypothetical acute viral infection, as an initial exploration of this potentially important phenomenon. Unfortunately, we did not have access to sufficient data to parameterize our model, or to test our model, for a specific pathogen. However, we hope that others who do, will adapt and parameterize our model for a specific pathogen and, in so doing, further interrogate our model and its predictions.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- METHOD DETAILS
  - ○ Mathematical description of the model
  - ○ Calculating epidemic outcomes
  - ○ Choice of parameter values and initial conditions
  - ○ Additional calculations
- QUANTIFICATION AND STATISTICAL ANALYSIS

### AUTHOR CONTRIBUTIONS

P.M.B. conceived the analysis. M.B., K.M.W., A.F., and R.A.N. designed the model. K.M.W., A.F., and R.A.N. analyzed the data. K.M.W. led the writing of the manuscript. All authors contributed critically to analysis discussion and manuscript drafts and approved a final version for publication.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

1. Anderson, R., and May, R. (1981). The population dynamics of microparasites and their invertebrate hosts. Philos. Trans. R. Soc. Lond. B Biol. Sci. *291*, 451–524. https://doi.org/10.1098/rstb.1981.0005.

2. Anderson, R., and May, R. (1991). Infectious Diseases of Humans: Dynamics and Control (Oxford University Press).

3. Woolhouse, M.E., Dye, C., Etard, J.F., Smith, T., Charlwood, J.D., Garnett, G.P., Hagan, P., Hii, J.L., Ndhlovu, P.D., Quinnell, R.J., et al. (1997). Heterogeneities in the transmission of infectious agents. Proc. Natl. Acad. Sci. USA *94*, 338–342. https://doi.org/10.1073/pnas.94.1.338.

4. May, R.M., and Anderson, R.M. (1987). Transmission dynamics of HIV infection. Nature *326*, 137–142. https://doi.org/10.1038/326137a0.

5. Kemper, J.T. (1980). On the identification of superspreaders for infectious disease. Math. Biosci. *48*, 111–127. https://doi.org/10.1016/0025-5564(80)90018-8.

6. Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., and Getz, W.M. (2005). Superspreading and the effect of individual variation on disease emergence. Nature *438*, 355–359. https://doi.org/10.1038/nature04153.

7. Miller, D., Martin, M.A., Harel, N., Tirosh, O., Kustin, T., Meir, M., Sorek, N., Gefen-Halevi, S., Amit, S., Vorontsov, O., et al. (2020). Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. Nat. Commun. *11*, 5518. https://doi.org/10.1038/s41467-020-19248-0.

8. Zhang, Y., Li, Y., Wang, L., Li, M., and Zhou, X. (2020). Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China. Int. J. Environ. Res. Publ. Health 17, 3705. https://doi.org/10.3390/ijerph17103705.

9. Adam, D.C., Wu, P., Wong, J.Y., Lau, E.H.Y., Tsang, T.K., Cauchemez, S., Leung, G.M., and Cowling, B.J. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nat. Med. 26, 1714–1719. https://doi.org/10.1038/s41591-020-1092-0.

10. Liu, Y., Eggo, R.M., and Kucharski, A.J. (2020). Secondary attack rate and superspreading events for SARS-CoV-2. Lancet 395, e47. https://doi.org/10.1016/S0140-6736(20)30462-1.

11. Gómez-Carballa, A., Bello, X., Pardo-Seco, J., Martinón-Torres, F., and Salas, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. Genome Res. 30, 1434–1448. https://doi.org/10.1101/GR.266221.120.

12. Correa-Martínez, C.L., Kampmeier, S., Kümpers, P., Schwierzeck, V., Hennies, M., Hafezi, W., Kühn, J., Pavenstädt, H., Ludwig, S., and Mellmann, A. (2020). A pandemic in times of global tourism: superspreading and exportation of COVID-19 cases from a ski area in Austria. J. Clin. Microbiol. 58, 19–21. https://doi.org/10.1128/JCM.00588-20.

13. Mushanyu, J., Chukwu, W., Nyabadza, F., and Muchatibaya, G. (2022). Modelling the potential role of super spreaders on COVID-19 transmission dynamics. Int. J. Math. Model. Numer. Optim. 12, 191–209. https://doi.org/10.1101/2021.08.30.21262341.

14. Stein, R.A. (2011). Super-spreaders in infectious diseases. Int. J. Infect. Dis. 15, e510–e513. https://doi.org/10.1016/j.ijid.2010.06.020.

15. Aral, S.O. (2004). Sexual risk behaviour and infection: epidemiological considerations. Sex. Transm. Infect. 80, 8–12. https://doi.org/10.1136/sti.2004.011866.

16. Clay, C.A., Lehmer, E.M., Previtali, A., St Jeor, S., and Dearing, M.D. (2009). Contact heterogeneity in deer mice: implications for Sin Nombre virus transmission. Proc. Biol. Sci. 276, 1305–1312. https://doi.org/10.1098/rspb.2008.1693.

17. Cohen, M.S., Hoffman, I.F., Royce, R.A., Kazembe, P., Dyer, J.R., Daly, C.C., Zimba, D., Vernazza, P.L., Maida, M., Fiscus, S.A., and Eron, J.J., Jr. (1997). Reduction of concentration of HIV-1 in semen after treatment of urethritis: implications for prevention of sexual transmission of HIV-1. Lancet 349, 1868–1873. https://doi.org/10.1016/S0140-6736(97)02190-9.

18. Omisakin, F., MacRae, M., Ogden, I.D., and Strachan, N.J.C. (2003). Concentration and prevalence of Escherichia coli O157 in cattle feces at slaughter. Appl. Environ. Microbiol. 69, 2444–2447. https://doi.org/10.1128/AEM.69.5.2444-2447.2003.

19. Yang, Q., Saldi, T.K., Gonzales, P.K., Lasda, E., Decker, C.J., Tat, K.L., Fink, M.R., Hager, C.R., Davis, J.C., Ozeroff, C.D., et al. (2021). Just 2% of SARS-CoV-2−positive individuals carry 90% of the virus circulating in communities. Proc. Natl. Acad. Sci. USA 118, e2104547118. https://doi.org/10.1073/pnas.2104547118.

20. Ma, J., and Earn, D.J.D. (2006). Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull. Math. Biol. 68, 679–702. https://doi.org/10.1007/s11538-005-9047-7.

21. Nold, A. (1980). Heterogeneity in disease-transmission modeling. Math. Biosci. 52, 227–240. https://doi.org/10.1016/0025-5564(80)90069-3.

22. Del Valle, S.Y., Hyman, J.M., and Chitnis, N. (2013). Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. Math. Biosci. Eng. 10, 1475–1497. https://doi.org/10.3934/mbe.2013.10.1475.

23. Gomes, M.G.M., Ferreira, M.U., Corder, R.M., King, J.G., Souto-Maior, C., Penha-Gonçalves, C., Gonçalves, G., Chikina, M., Pegden, W., and Aguas, R. (2022). Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. J. Theor. Biol. 540, 111063. https://doi.org/10.1016/j.jtbi.2022.111063.

24. Aguas, R., Gonçalves, G., Ferreira, M.U., Gonçalves, G., Ferreira, M.U., and Gabriela Gomes, M.M. (2021). Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics. medRxiv. https://doi.org/10.1101/2020.07.23.20160762.

25. Britton, T., Ball, F., and Trapman, P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science 369, 846–849. https://doi.org/10.1126/science.abc6810.

26. Beldomenico, P.M. (2020). Do superspreaders generate new superspreaders? A hypothesis to explain the propagation pattern of COVID-19. Int. J. Infect. Dis. 96, 461–463. https://doi.org/10.1016/j.ijid.2020.05.025.

27. van Damme, W., Dahake, R., van de Pas, R., Vanham, G., and Assefa, Y. (2021). COVID-19: does the infectious inoculum dose-response relationship contribute to understanding heterogeneity in disease severity and transmission dynamics? Med. Hypotheses 146, 110431. https://doi.org/10.1016/j.mehy.2020.110431.

28. Strong, R., la Rocca, S.A., Paton, D., Bensaude, E., Sandvik, T., Davis, L., Turner, J., Drew, T., Raue, R., Vangeel, I., and Steinbach, F. (2015). Viral dose and immunosuppression modulate the progression of acute BVDV-1 infection in calves: evidence of long term persistence after intra-nasal infection. PLoS One 10, e0124689. https://doi.org/10.1371/journal.pone.0124689.

29. Gaskell, R.M., and Povey, R.C. (1979). The dose response of cats to experimental infection with Feline Viral Rhinotracheitis virus. J. Comp. Pathol. 89, 179–191. https://doi.org/10.1016/0021-9975(79)90057-4.

30. Zarkov, I., and Bochev, I. (2008). Influence of inoculation dose of avian H6N2 influenza A virus on virus shedding and humoral immune response of chickens after artificial experimental intravenous infection. Rev. Med. Vet. 159, 489–493.

31. Mumford, J.A., Hannant, D., and Jessett, D.M. (1990). Experimental infection of ponies with equine influenza (H3N8) viruses by intranasal inoculation or exposure to aerosols. Equine Vet. J. 22, 93–98. https://doi.org/10.1111/j.2042-3306.1990.tb04217.x.

32. Dabisch, P.A., Biryukov, J., Beck, K., Boydston, J.A., Sanjak, J.S., Herzog, A., Green, B., Williams, G., Yeager, J., Bohannon, J.K., et al. (2021). Seroconversion and fever are dose-dependent in a nonhuman primate model of inhalational COVID-19. PLoS Pathog. 17, e1009865. https://doi.org/10.1371/journal.ppat.1009865.

33. Bjorkman, K.K., Saldi, T.K., Lasda, E., Bauer, L.C., Kovarik, J., Gonzales, P.K., Fink, M.R., Tat, K.L., Hager, C.R., Davis, J.C., et al. (2021). Higher viral load drives infrequent severe acute respiratory syndrome coronavirus 2 transmission between asymptomatic residence Hall roommates. J. Infect. Dis. 224, 1316–1324. https://doi.org/10.1093/infdis/jiab386.

34. Wu, P., Liu, F., Chang, Z., Lin, Y., Ren, M., Zheng, C., Li, Y., Peng, Z., Qin, Y., Yu, J., et al. (2021). Assessing asymptomatic, presymptomatic, and symptomatic transmission risk of severe acute respiratory syndrome coronavirus 2. Clin. Infect. Dis. 73, e1314–e1320. https://doi.org/10.1093/cid/ciab271.

35. Ge, Y., Martinez, L., Sun, S., Chen, Z., Zhang, F., Li, F., Sun, W., Chen, E., Pan, J., Li, C., et al. (2021). COVID-19 transmission dynamics among close contacts of index patients with COVID-19: a population-based cohort study in zhejiang province, China. JAMA Intern. Med. 181, 1343–1350. https://doi.org/10.1001/jamainternmed.2021.4686.

36. Ladoy, A., Opota, O., Carron, P.N., Guessous, I., Vuilleumier, S., Joost, S., and Greub, G. (2021). Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: a spatio-temporal investigation in Vaud state, Switzerland. Sci. Total Environ. 787, 147483. https://doi.org/10.1016/j.scitotenv.2021.147483.

37. Hay, J.A., Kennedy-Shaffer, L., Kanjilal, S., Lennon, N.J., Gabriel, S.B., Lipsitch, M., and Mina, M.J. (2021). Estimating epidemiologic dynamics from cross-sectional viral load distributions. Science 373, eabh0635. https://doi.org/10.1126/science.abh0635.

38. Jones, T.C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J., Bleicker, T., Tesch, J., Schmidt, M.L., Sander, L.E., et al. (2021). Estimating infectiousness throughout SARS-CoV-2 infection course. Science 373, eabi5273. https://doi.org/10.1126/science.abi5273.

39. Marks, M., Millat-Martinez, P., Ouchi, D., Roberts, C.H., Alemany, A., Corbacho-Monné, M., Ubals, M., Tobias, A., Tebé, C., Ballana, E., et al. (2021). Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. Lancet Infect. Dis. *21*, 629–636. https://doi.org/10.1016/S1473-3099(20)30985-3.

40. Taube, J.C., Miller, P.B., and Drake, J.M. (2022). An open-access database of infectious disease transmission trees to explore superspreader epidemiology. PLoS Biol. *20*, e3001685. https://doi.org/10.1371/journal.pbio.3001685.

41. Wong, G., Liu, W., Liu, Y., Zhou, B., Bi, Y., and Gao, G.F. (2015). MERS, SARS, and ebola: the role of super-spreaders in infectious disease. Cell Host Microbe *18*, 398–401. https://doi.org/10.1016/j.chom.2015.09.013.

42. Cori, A., Valleron, A.J., Carrat, F., Scalia Tomba, G., Thomas, G., and Boëlle, P.Y. (2012). Estimating influenza latency and infectious period durations using viral excretion data. Epidemics *4*, 132–138. https://doi.org/10.1016/j.epidem.2012.06.001.

43. Lelieveld, J., Helleis, F., Borrmann, S., Cheng, Y., Drewnick, F., Haug, G., Klimach, T., Sciare, J., Su, H., and Pöschl, U. (2020). Model calculations of aerosol transmission and infection risk of COVID-19 in indoor environments. Int. J. Environ. Res. Publ. Health *17*, 8114. https://doi.org/10.3390/ijerph17218114.

44. Kermack, W., and McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society A *115*, 700–721. https://doi.org/10.1098/rspa.1927.0118.

45. R Core Team (2020). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

46. Soetaert, K., Petzoldt, T., and Setzer, R.W. (2010). Solving differential equations in R: package deSolve. J. Stat. Software *33*. https://doi.org/10.18637/jss.v033.i09.

47. Diekmann, O., Heesterbeek, J.A.P., and Roberts, M.G. (2010). The construction of next-generation matrices for compartmental epidemic models. J. R. Soc. Interface *7*, 873–885. https://doi.org/10.1098/rsif.2009.0386.

48. Kim, Y., Ryu, H., and Lee, S. (2018). Agent-based modeling for super-spreading events: a case study of MERS-CoV transmission dynamics in the Republic of Korea. Int. J. Environ. Res. Publ. Health *15*, 2369. https://doi.org/10.3390/ijerph15112369.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Software and algorithms | | |
| R version 4.0.3 | R Foundation for Statistical Computing | https://www.r-project.org |
| Deposited Data | | |
| R code to perform simulations | This paper | https://github.com/kwanelik/Positive-superspreader-feedback-loops |
| | | https://doi.org/10.5281/zenodo.7767626 |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Klara M Wanelik (klara.wanelik@biology.ox.ac.uk).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- No empirical data were used in this paper.

- All original code for numerics and figure production has been deposited on GitHub and is publicly available. DOI is listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## METHOD DETAILS

We initially provide a broad overview of the model structure and metrics we used to quantify the impacts of positive superspreader feedback loops on an epidemic; we provide more detailed explanations and calculations relating to those metrics further down in an 'additional calculations' section.

### Mathematical description of the model

To understand the role of the positive feedback loop between superspreaders we developed the following model:

$$\frac{dS}{dt} = -S\beta_L(L + pH), \tag{Equation 1}$$

$$\frac{dL}{dt} = S\beta_L\big((1 - \sigma_L)L + p(1 - \sigma_H)H\big) - (\alpha + \gamma)L, \tag{Equation 2}$$

$$\frac{dH}{dt} = S\beta_L\big(\sigma_L L + \sigma_H pH\big) - (\alpha + \gamma)H, \tag{Equation 3}$$

$$\frac{dR}{dt} = \gamma(H + L). \tag{Equation 4}$$

where $S$ refers to the number of susceptible hosts, $L$ to infected hosts with low-titre infections (non-superspreaders), $H$ to infected hosts with high-titre infections (superspreaders) and $R$ to recovered hosts. $\beta_L$ represents the baseline transmission rate arising from $L$-infected hosts. This baseline rate is then increased through transmission by superspreaders ($H$-infected hosts) by a magnitude $p$ (we assume $p \geq 1$) representing the factor by which the transmission rate from a superspreader is greater than the transmission rate

from a non-superspreader (i.e. the transmission advantage of superspreaders); Mathematically, $p$ has no upper bound.

$\sigma_L$ is the proportion of infections from a non-superspreader that result in a superspreader, and so represents the background, spontaneous generation of superspreaders. $\sigma_H$ is the proportion of infections from a superspreader that result in another superspreader. $\sigma_H$ and $\sigma_L$ both vary between 0 and 1, but if $\sigma_H > \sigma_L$ then superspreaders are more likely to generate new superspreaders, and hence we use $\sigma_H / \sigma_L$ to represent the strength of the positive feedback loop of superspreader infections.

$\alpha$ is the death rate due to disease, and $\gamma$ the recovery rate from the disease (which may be through natural recovery, or recovery through hospitalisation), both assumed to be the same for superspreaders and non-superspreaders. For this model, there are no natural births or deaths, so the population only changes in size due to deaths resulting from the disease. So, if $\alpha = 0$ then the population remains constant, size $N = S + L + H + R$.

## Calculating epidemic outcomes

For an infection spreading through a closed, fixed-sized population, it is possible to calculate the final epidemic size – the total number of individuals infected throughout the epidemic.[44] Ma and Earn[20] showed that the formula for the final epidemic size, $Z$, is unchanged by the presence of a fixed proportion of superspreaders:

$$Z = S[0]\left(1 - Exp\left(-Z\frac{\beta_L}{\alpha+\gamma}(1 - \sigma_H + p\sigma_L) - \left(\frac{\beta_L}{\gamma+\alpha}\right)(L[0] + p\, H[0])\right)\right). \qquad \text{(Equation 5)}$$

Rescaling to express the sizes of each class as proportions of a fixed population size of 1, and in the limit $I(0) \to 0$, $S(0) \to 1$ the final epidemic size, $Z$, for all of the models they considered, is given by $Z = 1 - e^{-R_0 Z}$.[20]

For the current model, when $\sigma_H = \sigma_L$ (i.e. when there is no positive feedback loop) the final epidemic size, $Z$, is given by the same formula (see 'additional calculations' section below). However, when $\sigma_H > \sigma_L$, and so the positive feedback loop does exist, it is not possible to use the methods of Ma and Earn[20] to calculate the final epidemic size, since additional non-linearities in this model disqualify the simplifications they use. However, in the 'additional calculations' section below we show that when $\sigma_H > \sigma_L$, the final epidemic size will be larger than in the absence of such a feedback loop, and this difference increases as the difference between $\sigma_H$ and $\sigma_L$ increases. In the Results, we quantify the final epidemic size numerically by running simulations in R version 4.0.3[45] using the de-Solve package[46] while varying the ratio $\sigma_H/\sigma_L$ (the strength of the positive feedback loop, assuming $\sigma_H/\sigma_L \geq 1$) and $p$ (the transmission advantage of superspreaders, assuming $p \geq 1$). We also calculate numerically the peak prevalence of superspreaders and the herd immunity threshold i.e. the proportion of the population that needs to be immune to achieve herd immunity. The latter is calculated by subtracting from one the proportion of the population that remains susceptible when the number of infecteds peaks.[23] Finally, we calculate the basic reproduction number ($R_0$), i.e. the number of new infections generated by one infectious individual in a completely susceptible population for each of these scenarios. Using the next generation method of Diekmann et al.[47] (see 'additional calculations' section below), $R_0$ for this model is given by:

$$R_0 = \frac{N\beta_L\left(1 + p\sigma_H - \sigma_L + \sqrt{(\sigma_L - 1 - p\sigma_H)^2 - 4p(\sigma_H - \sigma_L)}\right)}{2(\alpha+\gamma)}. \qquad \text{(Equation 6)}$$

## Choice of parameter values and initial conditions

Parameter values are chosen to represent a hypothetical acute viral infection for illustrative purposes, and are consistent with the literature on acute viral infections like SARS-CoV-2, MERS-CoV and influenza (see Table 1). Again, for illustrative purposes, the initial number of non-superspreader infecteds, $L$, was kept constant at 10% of the total population size, $N$, as was the initial number of recovereds, $R$, at 0%. We varied the initial number of superspreaders, $H$, between 0–10% (consistent with previous modelling of acute viral infections, like MERS-CoV[48]). The total population size, $N$, was set to 10,000 – the population size of a small

town in the UK. All models were run for sufficient periods of time to ensure that the epidemic had completed in all cases.

## Additional calculations

From Equation (1) dividing both sides by $S$ and integrating both sides with respect to $t$

$$\log\left[\frac{S(\infty)}{S(0)}\right] = -\beta_L \int_0^\infty L\,dt - p\beta_L \int_0^\infty H\,dt. \qquad \text{(Equation 7)}$$

To solve Equation (2) we rearrange Equation (1) to get $\beta_L p SH$ in terms of $S$ and $L$ and substitute into Equation (2). Integrating both sides with respect to $t$ and assuming that $L(\infty) = 0$ gives:

$$-L(0) = (\sigma_H - 1)[S(\infty) - S(0)] + \int_0^\infty S\beta_L L(\sigma_H - \sigma_L)\,dt - (\gamma + \alpha)\int_0^\infty L\,dt. \qquad \text{(Equation 8)}$$

To solve Equation (3) we rearrange Equation (1) to get $\beta_L SL$ in terms of $S$ and $H$ and substitute into Equation (3). Integrating both sides with respect to $t$ and assuming that $H(\infty) = 0$ gives:

$$-H(0) = -\sigma_L[S(\infty) - S(0)] + \int_0^\infty S\beta_L p H(\sigma_H - \sigma_L)\,dt - (\gamma + \alpha)\int_0^\infty H\,dt. \qquad \text{(Equation 9)}$$

The final epidemic size is the difference between the number of susceptibles at $t = 0$ and $t = \infty$ i.e. $Z = S(0) - S(\infty)$. Rearranging Equations (8) and (9) in order to get the integrals of $H$ and $L$ with respect to time, and substituting into Equation (7) gives:

$$\log\left[\frac{S(\infty)}{S(0)}\right] = -\frac{\beta_L}{\alpha+\gamma}\left(L(0) + pH(0) - (1 - \sigma_H + p\sigma_L)Z + (\sigma_H - \sigma_L)\beta_L\int_0^\infty SL + p^2 SH\,dt\right).$$

$$\text{(Equation 10)}$$

When $\sigma_H = \sigma_L$ this gives the same solution as in Ma and Earn,[20] and for a fixed population size of 1 and in the limit $I(0) \to 0$, $S(0) \to 1$ the final epidemic size, $Z$, for all of the models they considered, is given by $Z = 1 - e^{-R_0 Z}$. However, in a model where there is a positive feedback loop, with superspreaders generating more superspreaders, there is an extra positive term which makes the exponent of the exponential term larger and still negative and so makes the final epidemic size bigger; the larger the difference between $\sigma_H$ and $\sigma_L$, the larger the impact on the final epidemic size.

2. Calculating the basic reproduction number, $R_0$, using the method of Diekmann et al.[47]

We calculated $R_0$ following the next generation approach of Diekmann et al.,[47] where $R_0$ is the dominant eigenvalue of the matrix $K_L$, obtained from the transmission matrix $T$ and the inverse of the transition matrix $\Sigma$ ($K_L = -T\Sigma^{-1}$), derived from the equations of the system presented above, and evaluated at the disease-free equilibrium (population of size $N$, comprising all susceptible individuals):

$$T = \begin{pmatrix} N\beta_L(1 - \sigma_L) & pN\beta_L(1 - \sigma_H) \\ N\beta_L\sigma_L & pN\beta_L\sigma_H \end{pmatrix}. \qquad \text{(Equation 11)}$$

$$\Sigma = \begin{pmatrix} -(\alpha+\gamma) & 0 \\ 0 & -(\alpha+\gamma) \end{pmatrix}. \qquad \text{(Equation 12)}$$

and so

$$K_L = \begin{pmatrix} \dfrac{N\beta_L(1 - \sigma_L)}{(\alpha+\gamma)} & \dfrac{pN\beta_L(1 - \sigma_H)}{(\alpha+\gamma)} \\ \dfrac{N\beta_L\sigma_L}{(\alpha+\gamma)} & \dfrac{pN\beta_L\sigma_H}{(\alpha+\gamma)} \end{pmatrix}. \qquad \text{(Equation 13)}$$

The dominant eigenvalue ($R_0$) of $K_L$ is, as presented above:

$$R_0 = \frac{N\beta_L\left(1 + p\sigma_H - \sigma_L + \sqrt{(\sigma_L - 1 - p\sigma_H)^2 - 4p(\sigma_H - \sigma_L)}\right)}{2(\alpha+\gamma)}. \qquad \text{(Equation 6)}$$

This expression always exists and is non-negative if $1 \geq \sigma_L$ and $1 \geq \sigma_H$ (i.e. if superspreaders are at least as likely to generate new superspreaders as non-superspreaders are, as assumed for our model).

We also note that in the absence of superspreading ($p = \sigma_H = \sigma_L = 0$), the basic reproduction number for a pathogen purely transmitting via low-titre, non-superspreaders, is:

$$R_0^L = \frac{N\beta_L}{\alpha + \gamma}. \qquad \text{(Equation 14)}$$

3. Threshold superspreading rates for an epidemic

From the above expression for $R_0$ we can calculate analytical expressions for key threshold values of $p$ (the relative transmission advantage of superspreading) for an epidemic to take off ($R_0 \geq 1$):

(i) How strong does superspreading need to be to allow an epidemic, when transmission from super-spreaders always generates new superspreaders (i.e. when $\sigma_H = 1$)?

Setting $\sigma_H = 1$ in the above expression for $R_0$, setting that expression equal to 1 (the threshold for an epidemic to occur), and solving for $p$ gives:

$$p_1 = \frac{(\alpha + \gamma)}{N\beta_L} = \frac{1}{R_o^L}. \qquad \text{(Equation 15)}$$

For the baseline parameter values used to generate Figure 3, $p_1 = 1.25$.

Hence, even if superspreaders always generate new superspreaders ($\sigma_H = 1$), if $R_o^L < 1$ (such that non-superspreaders alone are not able to cause an epidemic) there is a lower limit of $p$ that is greater than 1, that must be exceeded in order for an epidemic to occur. In other words, even 100% transmission of super-spreading is not necessarily sufficient to drive an epidemic of a pathogen with only mild levels of superspreading.

(ii) How strong does superspreading need to be to drive an epidemic, when there is no positive feedback loop between superspreaders (i.e. when $\sigma_H = \sigma_L$)?

Setting $\sigma_H = \sigma_L$ in the above expression for $R_0$, setting that expression equal to 1 and solving for $p$ gives:

$$p_2 = \frac{\alpha + \gamma - N\beta_L(1 - \sigma_L)}{N\beta_L\sigma_L} = \frac{1/R_o^L - (1 - \sigma_L)}{\sigma_L} = \frac{p_1 - (1 - \sigma_L)}{\sigma_L}. \qquad \text{(Equation 16)}$$

Hence in the absence of a positive feedback loop in superspreader transmission, the rate of superspreader advantage ($p$) needs to exceed a threshold value determined by the baseline $R_0$ in the absence of super-spreading ($R_0^L$) and the rate at which non-superspreaders spontaneously generate superspreaders ($\sigma_L$), in order for an epidemic to occur. This threshold is shown in Figure 3 by the value of $p$ along the x-axis (when $\sigma_H = \sigma_L$), at the boundary where $R_0 = 1$. For the baseline parameter values used to generate Figure 3, $p_2 = 6$.

## QUANTIFICATION AND STATISTICAL ANALYSIS
Results were generated via numerical simulation in R version 4.0.3.