

---

# FUNDAMENTAL BOUND ON EPIDEMIC OVERSHOOT IN THE SIR MODEL

---

**Maximilian M. Nguyen**

Lewis-Sigler Institute  
Princeton University  
Princeton, NJ 08544  
mmnguyen@princeton.edu

**Ari S. Freedman**

Department of Ecology and Evolutionary Biology  
Princeton University  
Princeton, NJ 08544  
arisf@princeton.edu

**Sinan A. Ozbay**

Bendheim Center for Finance  
Princeton University  
Princeton, NJ 08544  
sozbay@princeton.edu

**Simon A. Levin**

Department of Ecology and Evolutionary Biology  
Princeton University  
Princeton, NJ 08544  
slevin@princeton.edu

June 2, 2023

## ABSTRACT

We derive an exact upper bound on the epidemic overshoot for the Kermack-McKendrick SIR model. This maximal overshoot value of 0.2984... occurs at  $R_0^* = 2.151...$ . Using the general analysis framework presented within, we then consider more complex SIR models, such as those that incorporate vaccination or contact heterogeneity. We analyze models that consider vaccinations and show that the presence of vaccinated individuals decreases the maximum possible overshoot. For epidemics where the contact structure is given by a network, we numerically find that increased contact heterogeneity lowers the maximal overshoot value and weakens the dependency of overshoot on transmission.

## Introduction

The overshoot of an epidemic is the proportion of the population that becomes infected after the peak of the epidemic has already passed. Formally, it is given as the difference between the fraction of the population that is susceptible at the peak of infection prevalence and at the end of the epidemic. Intuitively, it is the difference between the herd immunity threshold and the total fraction of the population that gets infected [7, 4]. As it describes the damage to the population in the declining phase of the epidemic (i.e. when the effective reproduction number is less than 1), one might be tempted to dismiss its relative importance. However, a substantial proportion of the epidemic, and thus a large number of people, may be impacted during this phase of the epidemic dynamics.

A natural question to ask then is how large can the overshoot be and how does the overshoot depend on epidemic parameters, such as transmissibility and recovery rate? Surprisingly, within the framework of the SIR ODE model, this question can be answered exactly. In this paper, we first derive the bound on the overshoot in the Kermack-McKendrick limit of the SIR model [8]. Beyond the basic SIR model, we then see if the bound on overshoot holds if we add additional complexity, such as vaccinations, into the model. We then numerically study what happens when we move to a more realistic structural model of epidemics that incorporates contact heterogeneity.

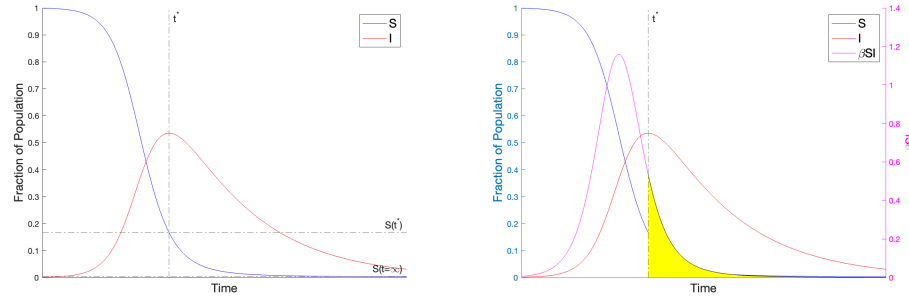


Figure 1: The overshoot can be calculated in two ways: a) Overshoot is calculated as the difference between the fraction of the population that is susceptible at  $t^*$  and infinite time. b) Overshoot is calculated as the integral of the infection incidence curve from  $t^*$  until infinite time. Therefore overshoot corresponds to the area of the region shaded in yellow.

## Results

The classical Kermack-McKendrick SIR model is given by the following set of ODEs:

$$\frac{dS}{dt} = -\beta SI \quad (1)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

where  $S$ ,  $I$ , and  $R$  are the fractions of population in the susceptible, infected, or recovered state respectively. As these are the only possible states within this model, the conservation equation for the whole population is given as  $S + I + R = 1$ . Conceptually, the overshoot can be equivalently calculated in two ways. In the first it is given by the difference in the fraction of susceptible individuals at the peak of infection prevalence ( $S_{t^*}$ ) and at the end of the epidemic ( $S_\infty$ ) (Figure 1a). Alternatively, it can be viewed as the integration of the number of newly infected individuals, which is given by the infection incidence curve ( $\frac{dIncidence}{dt} = \beta SI$ ) from the peak of infection prevalence to the end of the epidemic (Figure 1b). We will make use of the former relationship in the results that follow.

The only two parameters of this model are  $\beta$  and  $\gamma$ . A key parameter in epidemic modeling combines these two into a single parameter by taking their ratio, which is known as the basic reproduction number ( $R_0$ ). While overshoot is a function of both  $\beta$  and  $\gamma$  independently, the behavior of the overshoot can be parameterized in terms of the single parameter  $R_0$  (Figure S1). Plotting the dependency of overshoot on  $R_0$  (Figure 2), we observe a peak in the curve at  $(R_0^*, Overshoot^*)$  that sets an upper bound on the overshoot. From a public health perspective, diseases that have estimated  $R_0$ 's near this peak region in Figure 2 include COVID-19 (ancestral strain) [2], SARS [16], diphtheria [14], monkeypox [6], and ebola [15]. This peak phenomena in the overshoot was first numerically observed by [17], though not explained. We will now derive the solution for this maximum point analytically.

### Deriving the Exact Bound on Overshoot in the Kermack-McKendrick SIR Model

*Theorem: The maximum possible overshoot in the Kermack-McKendrick SIR model is a fraction 0.2984... of the entire population, with a corresponding  $R_0^* = 2.151...$*

*Proof.*

Let  $t^*$  be the time at the peak of the infection prevalence curve. The herd immunity threshold is the difference in the fractions of the population that are susceptible at zero time and at  $t^*$ . Define overshoot as the difference in the fractions of the population that are susceptible at  $t^*$  and at infinite time. This is equivalent to defining overshoot as the cumulative fraction of the population that gets infected after  $t^*$ .

$$\begin{aligned} \text{Overshoot} &\equiv \int_{t^*}^{\infty} -\left(\frac{dS}{dt}\right) dt = \int_{t^*}^{\infty} \beta SI dt \\ &= S_{t^*} - S_\infty \end{aligned} \quad (4)$$

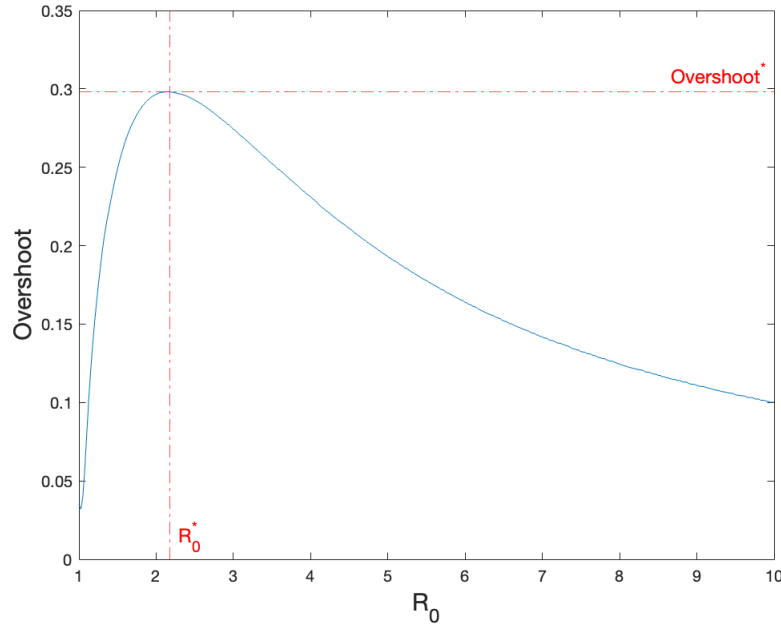


Figure 2: The overshoot as a function of  $R_0$  for the Kermack-McKendrick SIR model.

where  $S_{t^*}$  and  $S_\infty$  are the susceptible fractions at  $t^*$  and infinite time respectively. We will use  $S_{t^*} = \frac{1}{R_0}$  [10], which can be obtained by setting (2) to zero and solving for that critical  $S$ . We will use the notation  $X_t$  to indicate the value of compartment  $X$  at time  $t$ .

$$Overshoot = \frac{1}{R_0} - S_\infty \quad (5)$$

Since we would like to compute maximal overshoot, we can differentiate the overshoot equation (5) with respect to  $S_\infty$  to find the extremum. We will eliminate  $R_0$  from the overshoot equation so that we have an equation only in terms of  $S_\infty$ .

To find an expression for  $R_0$ , we start by deriving the standard final size relation for the SIR model [1, 3]. We solve for the rate of change of  $I$  as a function of  $S$  using (1)-(2) to obtain

$$\frac{dI}{dS} = -1 + \frac{\gamma}{\beta S}$$

from which it follows on integration that  $S + I - \frac{\gamma}{\beta} \ln S$  is constant along any trajectory.

Considering the beginning of the epidemic and the peak of the epidemic yields:

$$S_0 + I_0 - \frac{\gamma}{\beta} \ln S_0 = S_\infty + I_\infty - \frac{\gamma}{\beta} \ln S_\infty$$

hence

$$\frac{\beta}{\gamma} (S_\infty - S_0 + I_\infty - I_0) = \ln\left(\frac{S_\infty}{S_0}\right) \quad (6)$$

We now define the initial conditions:  $S_0 = 1 - \epsilon$  and  $I_0 = \epsilon$ , where  $\epsilon$  is the (infinitesimally small) fraction of initially infected individuals. We use the standard asymptotic of the SIR model that there are no infected individuals at the end of an SIR epidemic:  $I_\infty = 0$ . Hence, recalling that  $R_0 = \frac{\beta}{\gamma}$  we obtain that

$$S_\infty = e^{R_0(S_\infty - 1)} \quad (7)$$

The resulting equation (7) is the final size relation for the Kermack-McKendrick SIR model.

Rearranging for  $R_0$  yields the following expression:

$$\frac{\ln(S_\infty)}{S_\infty - 1} = R_0 \quad (8)$$

We then substitute this  $R_0$  expression (8) into the overshoot equation (5).

$$Overshoot = \frac{S_\infty - 1}{\ln(S_\infty)} - S_\infty \quad (9)$$

Differentiating with respect to  $S_\infty$  and setting the equation to zero to find the maximum overshoot yields:

$$(\ln S_{\infty*})^2 = \ln S_{\infty*} - 1 + \frac{1}{S_{\infty*}} \quad (10)$$

whose solution is

$$S_{\infty*} = 0.1664...$$

and which corresponds to

$$Overshoot^* = 0.2984... \quad (11)$$

using (9). The corresponding  $R_0$  calculated using (8) is

$$R_0^* = 2.151... \quad (12)$$

This concludes the proof ■

Additionally, to find the total recovered fraction is straightforward. In the asymptotic limit of the SIR model, there are no remaining infected individuals, so  $R_{\infty*} = 1 - S_{\infty*}$ .

$$R_{\infty*} = 1 - 0.1664... = 0.8336... \quad (13)$$

In other words, approximately 5 out of every 6 individuals in the population will have experience infection when overshoot is maximized.

## Upper Bounds on Overshoot in Models that Include Vaccinations

Beyond the Kermack-McKendrick SIR model, one can ask if the bound on overshoot still holds if other complexities are added to the model. First, we will consider the addition of vaccinations.

We will consider three qualitatively different types of curves for the vaccination rate (Figure 3). These correspond to different scenarios that might be modeled. The first model assumes a vaccination rate of zero after the outbreak begins, which implies all vaccinations occur before the outbreak. The second model of vaccination assumes a constant per-capita vaccination rate. This is a situation where all susceptible individuals get vaccinated at the same rate. This assumption yields a vaccination curve for the population that is concave down. The third type of model assumes a risk-driven vaccination rate that depends on the number of infected individuals. This yields a non-monotonic vaccination curve for the population that switches from being initially concave up to being concave down. Depending on the scenario being analyzed, one model might be more appropriate to use than others. Below we discuss each model in further detail by providing the corresponding system of equations, relevant scenarios the model might correspond to in reality, and the corresponding maximal overshoot for each model.

### Maximal Overshoot when the Number of Vaccinated Individuals is Constant

The first model of vaccination assumes there are no vaccinations during the outbreak, which implies a fixed number of vaccinated individuals over the course of the epidemic. Such a scenario might be the reintroduction of an infectious disease into a population that has a pre-existing level of immunity.

Since the number of vaccinated individuals is constant, this implies all vaccinations occurred prior to the initial time step. The calculation is then trivial assuming vaccinations provide complete and permanent immunity. In that case, vaccinated individuals can simply be ignored entirely in the dynamics, resulting in the maximal overshoot simply scaling with the unvaccinated fraction.

$$Overshoot_{SIRV}^* = (1 - V)0.2984... \quad (14)$$

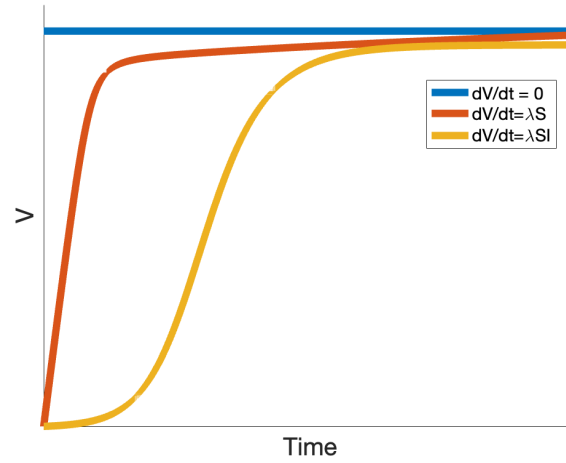


Figure 3: The fraction of population that is vaccinated ( $V$ ) based on different vaccination rates: a vaccination rate of zero over the course of the epidemic (blue), a constant per-capita vaccination rate (red), and a risk-driven vaccination rate (yellow).

#### Maximal Overshoot Under Addition of Constant Per-Capita Vaccination

We next consider a more typical scenario where the vaccination rate per unvaccinated individual is constant per unit time. Barring any additional information about the population or the epidemic, it is reasonable to assume that all susceptible individuals are vaccinated at the same rate. Consider the following SIRV model:

$$\frac{dS}{dt} = -\beta SI - \lambda S \quad (15)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (16)$$

$$\frac{dR}{dt} = \gamma I \quad (17)$$

$$\frac{dV}{dt} = \lambda S \quad (18)$$

In this case, it is easily shown that there is a conserved quantity,  $S + I - \frac{\gamma}{\beta} \ln S + \frac{\lambda}{\beta} \ln I$ , which reduces to (6) when the vaccination rate is zero (i.e.  $\lambda = 0$ ). Unfortunately, having the conserved quantity is not sufficient to compute the overshoot, since there does not appear to be a way to separate infected and vaccinated individuals when trying to extend the previous calculation. Therefore, we turn to numerical computation (Figure 4a). We find that the maximal overshoot is bounded above by the value already obtained in the model without vaccinations. As shown in Figure 4, the overshoot has a complicated dependence on the vaccination parameter  $\lambda$  and  $R_0$ .

#### Maximal Overshoot Under Addition of a Risk-Driven Vaccination Rate

Lastly consider a vaccination rate that is proportional to the number of infected individuals. Such risk-driven behavior may arise for a variety of reasons, including initial vaccine hesitancy, a delay in vaccine availability, or a correlation between willingness to get vaccinated and the number of infected individuals. Consider the following SIRV model:

$$\frac{dS}{dt} = -\beta SI - \lambda SI \quad (19)$$

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (20)$$

$$\frac{dR}{dt} = \gamma I \quad (21)$$

$$\frac{dV}{dt} = \lambda SI \quad (22)$$

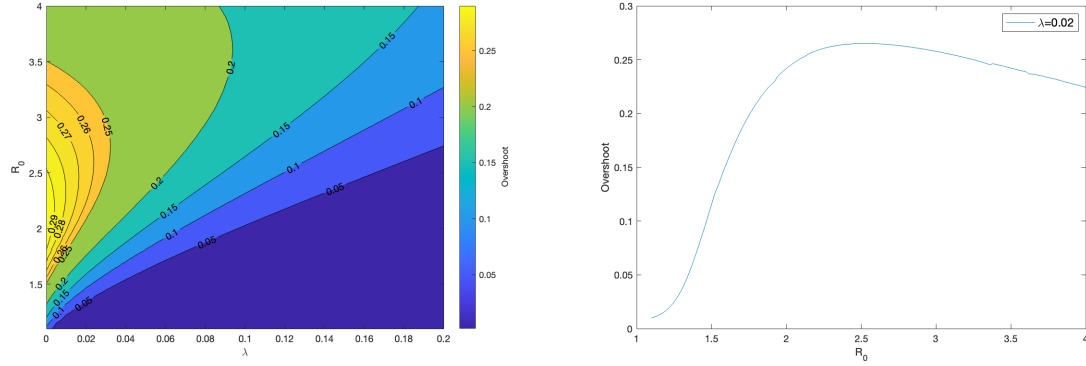


Figure 4: a) Contour plot for the overshoot for the SIRV model with  $\frac{dV}{dt} = \lambda S$  as a function of  $\lambda$  and  $R_0$ . b) Vertical cross-section of the contour plot from (a) for  $\lambda = 0.02$ .

Since the model now has an additional compartment, V, compared with the original SIR model, we must update our definition for overshoot accordingly. Fundamentally, overshoot compares the fraction of people who have not been infected at the epidemic peak and the people who have not been infected at the end of the epidemic. The fraction of people who have not been infected at any particular time,  $t$ , is  $S_t + V_t$ . Thus, overshoot can be redefined as follows.

$$Overshoot = (S_{t^*} + V_{t^*}) - (S_{\infty} + V_{\infty})$$

Since the equation for  $\frac{dI}{dt}$  remains unchanged,  $S_{t^*} = \frac{1}{R_0}$  still applies. Thus, the overshoot equation for models with vaccinated compartments is given by:

$$Overshoot = \left(\frac{1}{R_0} + V_{t^*}\right) - (S_{\infty} + V_{\infty}) \quad (23)$$

To maximize overshoot, we thus need to find expressions for  $R_0$ ,  $V_{t^*}$ , and  $V_{\infty}$  in terms of  $S_{\infty}$ .

To find  $R_0$  we start by taking the ratio  $\frac{dI}{dS}$  and integrating as before. It follows that  $I + \frac{\beta}{\beta+\lambda}S - \frac{\gamma}{\beta+\lambda}\ln S$  is constant along any trajectory. Considering the beginning and the end of the epidemic yields:

$$I_0 + \frac{\beta}{\beta+\lambda}S_0 - \frac{\gamma}{\beta+\lambda}\ln S_0 = I_{\infty} + \frac{\beta}{\beta+\lambda}S_{\infty} - \frac{\gamma}{\beta+\lambda}\ln S_{\infty}$$

Using the same initial conditions, asymptotic behavior, and parameter substitution as before ( $S_0 = 1 - \epsilon$ ,  $I_0 = \epsilon$ ,  $I_{\infty} = 0$ ,  $R_0 = \frac{\beta}{\gamma}$ ) yields the following final size relation.

$$R_0 = \frac{\ln(S_{\infty})}{S_{\infty} - 1} \quad (24)$$

Thus, we see that  $R_0$  for this SIRV model takes on the same expression as the SIR model (8).

To find  $V_{t^*}$ , let us take the ratio of time derivatives of the S and V compartments (19), (22),

$$\begin{aligned} \frac{\frac{dS}{dt}}{\frac{dV}{dt}} &= \frac{-\beta SI - \lambda SI}{\lambda SI} \\ \frac{dS}{dV} &= -\left(\frac{\beta + \lambda}{\lambda}\right) \end{aligned}$$

from which it follows on integration that  $S + \left(\frac{\beta + \lambda}{\lambda}\right)V$  is constant along any trajectory. Considering the beginning and the peak of the epidemic yields:

$$S_0 + \left(\frac{\beta + \lambda}{\lambda}\right)V_0 = S_{t^*} + \left(\frac{\beta + \lambda}{\lambda}\right)V_{t^*}$$

Using the initial conditions ( $S_0 = 1 - \epsilon$ ,  $I_0 = \epsilon$ ,  $V_0 = 0$ ) and recalling that  $S_{t^*} = \frac{1}{R_0}$ , we obtain the following formula for  $V_{t^*}$ .

$$V_{t^*} = \left(1 - \frac{1}{R_0}\right)\left(\frac{\lambda}{\beta + \lambda}\right) \quad (25)$$

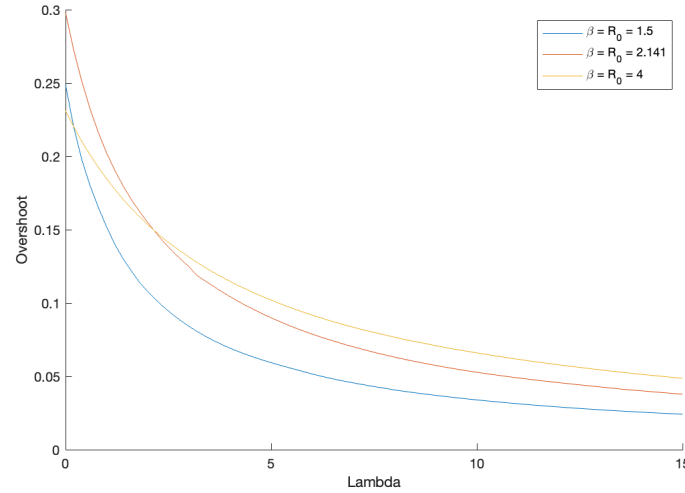


Figure 5: The overshoot for the SIRV model with  $\frac{dV}{dt} = \lambda SI$  as a function of  $\lambda$  for different levels of  $\beta$  (or equivalently  $R_0$ ).

To find  $V_\infty$ , recall that  $S + (\frac{\beta+\lambda}{\lambda})V$  is constant along any trajectory. Considering the peak of the epidemic and the end of the epidemic yields

$$S_{t^*} + (\frac{\beta+\lambda}{\lambda})V_{t^*} = S_\infty + (\frac{\beta+\lambda}{\lambda})V_\infty$$

Using the equation for  $V_{t^*}$  (25) and recalling that  $S_{t^*} = \frac{1}{R_0}$ , we obtain the following equation for  $V_\infty$ .

$$V_\infty = (1 - S_\infty)(\frac{\lambda}{\beta + \lambda}) \quad (26)$$

Substituting the expressions for  $R_0$  (24),  $V_{t^*}$  (25),  $V_\infty$  (26) into the overshoot equation (23) yields:

$$Overshoot = (\frac{S_\infty - 1}{\ln(S_\infty)} - S_\infty)(1 - \frac{\lambda}{\beta + \lambda}) \quad (27)$$

We see that this expression for the overshoot is simply the overshoot expression for the original SIR model (9) scaled by a factor  $1 - \frac{\lambda}{\beta + \lambda}$ .

$$Overshoot_{SIRV(\lambda SI)} = Overshoot_{SIR}(1 - \frac{\lambda}{\beta + \lambda}) \quad (28)$$

Since both  $\beta, \lambda \geq 0$ , then the factor  $1 - \frac{\lambda}{\beta + \lambda}$  can never be greater than 1. This implies that the bound on maximal overshoot given by the theorem holds, becoming exact in the limit of no vaccinations (ie.  $\lambda = 0$ ). For this model, the maximal overshoot decreases as a function of  $\lambda$  in a nonlinear way and has a nonlinear dependence on  $R_0$  (Figure 5).

### Overshoot Behavior Under Addition of Contact Heterogeneity

One of the key assumptions of the Kermack-McKendrick model is a well-mixed population where every infected individual has the same effect on the population. While a mathematically convenient assumption, real-world epidemics happen in more structured populations [13]. This is evidenced by the phenomena of superspreaders, where some fraction of infected individuals infect a disproportionately large number of susceptibles. This is modeled in the literature through the introduction of contact heterogeneity [9, 5].

Here we explore what happens to the overshoot when the contact structure of the population is given by a network graph that is roughly one giant component. While it is possible to construct pathological graphs that produce very complex dynamics, we consider more classical graphs here. Using a parameterization of heterogeneity given by



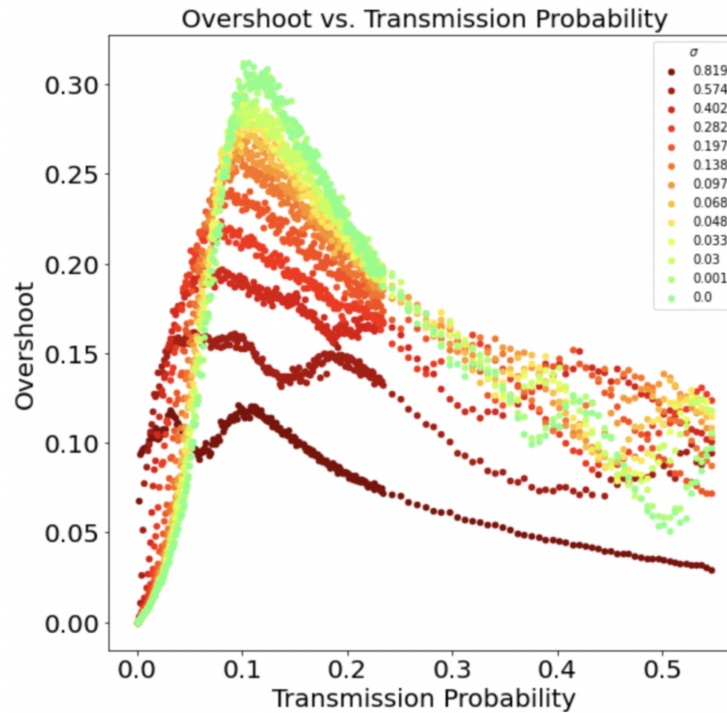


Figure 6: The overshoot for SIR epidemic simulations on networks with varying levels of heterogeneity ( $\sigma$ ) as a function of transmission probability.  $N = 1000$ ,  $\lambda = 5$ . Points represent the average of  $n = 150$  simulation runs.

Ozbay et. al. [12] (see Methods for details), we simulated epidemics on networks with structure ranging from the homogeneous limit (well-mixed, complete graph) to a heterogeneous limit (heavy-tailed degree distributions). We observed what happens to the overshoot on these different graphs as we changed the transmission probability.

On a network model, where contact structure is made explicit, the homogeneous graph is a complete graph, which recapitulates the well-mixed assumption of the Kermack-McKendrick model. It is not surprising then that the overshoot in the homogeneous graph ( $\sigma = 0$ ) peaks also around 0.3 (Figure 6, *green*) and has qualitatively the same shape profile as the ODE model (Figure 2).

We also see that increasing contact heterogeneity qualitatively suppresses the overshoot peak both in terms of the overshoot value and the corresponding transmission probability. Furthermore, increased heterogeneity also flattens out the overshoot curve as a function of transmission probability.

## Discussion

We have proved that the maximum fraction of the population that can be infected during the overshoot phase of an epidemic in the Kermack-McKendrick SIR model is about 0.3, with a corresponding  $R_0 \approx 2.15$ . This upper bound on the overshoot seems to hold in other extensions of SIR models as well. In an SIR model with vaccinations, the upper bound stays the same. It also matches the numerical upper bound seen for SIR dynamics on networks of varying heterogeneity. In the 2-strain with vaccination SIR model of Zarnitsyna et. al. [17], the overshoot depends on both the level of strain dominance and vaccination rate, but from their results it is numerically seen that any amount of vaccination will produce an overshoot lower than the bound found here. Different control measures and strategies may reduce the overshoot as compared to the unmitigated case [7], keeping this upper bound intact. It will be interesting to see how general this bound is for SIR models with other types of complexities or for models beyond the SIR-type.

The mathematical intuition on why there is a peak in the overshoot as a function of  $R_0$  can be seen by an inspection of Equation 5. The first term,  $\frac{1}{R_0}$  monotonically decreases with increasing  $R_0$ . The last term,  $-S_\infty$ , monotonically increases with  $R_0$ . Thus a trade-off in the two terms results in some intermediate peak. The epidemiological intuition behind a peak in the overshoot is that the total number of individuals infected during the epidemic grows monotonically with increasing  $R_0$ . However, too high of an  $R_0$  leads to a sharp growth in the number



of infected individuals, which burns through most of the population before the infection prevalence peak is reached, leaving few susceptible individuals left for the overshoot phase. This is seen by a monotonic decrease in the fraction of infected individuals that occur in the overshoot phase with increased  $R_0$  (Figure S2). Thus the maximal overshoot occurs as a trade-off between those two directions. From a public-health perspective, these results suggest that a significant number of infections may still occur after the peak of infections for diseases that spread relatively slowly (small  $R_0$ ).

## References

- [1] Julien Arino et al. “A final size relation for epidemic models”. In: *Mathematical Biosciences & Engineering* 4.2 (2007). Publisher: American Institute of Mathematical Sciences, p. 159.
- [2] Md Arif Billah, Md Mamun Miah, and Md Nuruzzaman Khan. “Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence”. In: *PLOS ONE* 15.11 (Nov. 11, 2020). Publisher: Public Library of Science, e0242128. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0242128. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0242128> (visited on 12/08/2022).
- [3] Fred Brauer. “Age-of-infection and the final size relation”. In: *Mathematical Biosciences & Engineering* 5.4 (2008). Publisher: American Institute of Mathematical Sciences, p. 681.
- [4] Sarah Cobey. “Modeling infectious disease dynamics”. In: *Science* 368.6492 (May 15, 2020). Publisher: American Association for the Advancement of Science, pp. 713–714. DOI: 10.1126/science.abb5659. URL: <https://www.science.org/doi/10.1126/science.abb5659> (visited on 11/10/2022).
- [5] Ryo Fujie and Takashi Odagaki. “Effects of superspreaders in spread of epidemic”. In: *Physica A: Statistical Mechanics and its Applications* 374.2 (Feb. 1, 2007), pp. 843–852. ISSN: 0378-4371. DOI: 10.1016/j.physa.2006.08.050. URL: <https://www.sciencedirect.com/science/article/pii/S0378437106008703> (visited on 11/10/2022).
- [6] Rebecca Grant, Liem-Binh Luong Nguyen, and Romulus Breban. “Modelling human-to-human transmission of monkeypox”. In: *Bulletin of the World Health Organization* 98.9 (Sept. 1, 2020), pp. 638–640. ISSN: 0042-9686. DOI: 10.2471/BLT.19.242347. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7463189/> (visited on 12/08/2022).
- [7] Andreas Handel, Ira M Longini, and Rustom Antia. “What is the best control strategy for multiple infectious disease outbreaks?” In: *Proceedings of the Royal Society B: Biological Sciences* 274.1611 (Mar. 22, 2007). Publisher: Royal Society, pp. 833–837. DOI: 10.1098/rspb.2006.0015. URL: <https://royalsocietypublishing.org/doi/10.1098/rspb.2006.0015> (visited on 11/10/2022).
- [8] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. “A contribution to the mathematical theory of epidemics”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (Aug. 1927). Publisher: Royal Society, pp. 700–721. DOI: 10.1098/rspa.1927.0118. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118> (visited on 09/16/2022).
- [9] J. O. Lloyd-Smith et al. “Superspreading and the effect of individual variation on disease emergence”. In: *Nature* 438.7066 (Nov. 2005). Number: 7066 Publisher: Nature Publishing Group, pp. 355–359. ISSN: 1476-4687. DOI: 10.1038/nature04153. URL: <https://www.nature.com/articles/nature04153> (visited on 11/10/2022).
- [10] Robert M. May. “Vaccination programmes and herd immunity”. In: *Nature* 300.5892 (Dec. 1982). Number: 5892 Publisher: Nature Publishing Group, pp. 481–483. ISSN: 1476-4687. DOI: 10.1038/300481a0. URL: <https://www.nature.com/articles/300481a0> (visited on 11/10/2022).
- [11] Mark Newman. *Networks*. Google-Books-ID: YdZjDwAAQBAJ. Oxford University Press, July 4, 2018. 793 pp. ISBN: 978-0-19-252749-3.
- [12] Sinan A. Ozbay and Maximilian M. Nguyen. *Parameterizing Network Graph Heterogeneity using a Modified Weibull Distribution*. Dec. 13, 2022. DOI: 10.48550/arXiv.2212.06994. arXiv: 2212.06994[physics, stat]. URL: <http://arxiv.org/abs/2212.06994> (visited on 12/20/2022).
- [13] Romualdo Pastor-Satorras et al. “Epidemic processes in complex networks”. In: *Reviews of Modern Physics* 87.3 (Aug. 31, 2015). Publisher: American Physical Society, pp. 925–979. DOI: 10.1103/RevModPhys.87.925. URL: <https://link.aps.org/doi/10.1103/RevModPhys.87.925> (visited on 09/16/2022).
- [14] Shaun A Truelove et al. “Clinical and Epidemiological Aspects of Diphtheria: A Systematic Review and Pooled Analysis”. In: *Clinical Infectious Diseases* 71.1 (June 24, 2020), pp. 89–97. ISSN: 1058-4838. DOI: 10.1093/cid/ciz808. URL: <https://doi.org/10.1093/cid/ciz808> (visited on 12/08/2022).

- [15] Z. S. Y. Wong et al. “A systematic review of early modelling studies of Ebola virus disease in West Africa”. In: *Epidemiology & Infection* 145.6 (Apr. 2017). Publisher: Cambridge University Press, pp. 1069–1094. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268817000164. URL: <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/systematic-review-of-early-modelling-studies-of-ebola-virus-disease-in-west-africa/154353B9A815326FE3656046AD6390B6> (visited on 12/08/2022).
- [16] World Health Organization. *Consensus document on the epidemiology of severe acute respiratory syndrome (SARS)*. WHO/CDS/CSR/GAR/2003.11. number-of-pages: 46. World Health Organization, 2003. URL: <https://apps.who.int/iris/handle/10665/70863> (visited on 12/08/2022).
- [17] Veronika I. Zarnitsyna et al. “Intermediate levels of vaccination coverage may minimize seasonal influenza outbreaks”. In: *PLOS ONE* 13.6 (June 26, 2018). Publisher: Public Library of Science, e0199674. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0199674. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0199674> (visited on 11/10/2022).

## Methods

### Generating Graphs of Differing Heterogeneity

In Figure 6, we presented the results of SIR simulations of epidemics run on graphs of size  $N = 1000$  and  $\lambda = 5$ , where the parameter of interest is  $\sigma$ . Each curve, from red to green, represents a different value of the graph heterogeneity. We implemented the following procedure from [12] for generating graphs as a function of a continuous parameter ( $\sigma$ ):

The following simple procedure generates a graph that has the desired heterogeneity:

1. Choose values of  $\sigma$  (heterogeneity),  $\lambda$  (mean/median node degree), and  $N$  (number of nodes).
2. Draw  $N$  random samples from the following distribution using  $\sigma$  and  $\lambda$ , rounding these samples to the nearest integer, since the degree of a node can only take on integer values.

$$f(x; \lambda, \sigma) = \frac{-\ln(\sigma)}{\lambda} \left(\frac{x}{\lambda}\right)^{-\ln(\sigma)-1} e^{-(x/\lambda)^{-\ln(\sigma)}}; x \geq 0; \sigma \in (0, 1]; \lambda > \mathbb{R}^+$$

3. With the sampled degree distribution from the previous step, now use the configuration model method [11] (which samples over the space of all possible graphs corresponding to a particular degree distribution) to generate a corresponding graph.

This yields a valid graph with the desired amount of heterogeneity as specified by  $\sigma$ .

### Simulating Epidemics on Graphs of Differing Heterogeneity

We implemented the following simulation procedure from [12] for implementing an SIR epidemic on a graph:

Given a graph  $G(\sigma)$  of heterogeneity  $\sigma$ , fix a transmission probability  $\tau$  and recovery probability  $\gamma$ :

1. At time  $t_0$ , fix a small fraction  $f$  of nodes to be chosen uniformly on the graph and assign them to the Infected state. The remaining  $(1 - f)$  fraction of nodes start as Susceptible.
2. For each  $i \in [1, T]$ , for each pair of adjacent S and I nodes, the susceptible node becomes infected with probability  $\tau$ .
3. For each  $i \in [1, T]$ , each infected node recovers with probability  $\gamma$ .
4. At time  $T$ , record two quantities: The final attack rate and the herd immunity threshold at the peak of the epidemic.
5. Repeat steps (1-4)  $n = 150$  times for each value of  $\tau$ .
6. Repeat steps (1-5) for each value of  $\sigma$ .

### Numerical Solutions and Code

Numerical calculations and epidemic simulations on networks were performed in *MATLAB*. Code for all sections can be provided upon request.

## **Acknowledgements**

The authors would like to acknowledge Bryan Grenfell and Chadi Saad-Roy for their useful suggestions. The authors would like to acknowledge generous funding support provided by the National Science Foundation (CCF1917819 and CNS-2041952), the Army Research Office (W911NF-18-1-0325), and a gift from the William H. Miller III 2018 Trust.

## **Author Contributions**

M.M.N., A.S.F., S.A.O., S.A.L. designed research, performed research, and wrote and reviewed the manuscript.

## **Additional information**

The authors declare no competing interests.

## Supplemental Materials: Fundamental Bound on Epidemic Overshoot in the SIR Model

### Supplemental Figures

**Figure S1. Comparing overshoot in the SIR model as a function of  $R_0$  compared to  $\beta$  and  $\gamma$  individually.**

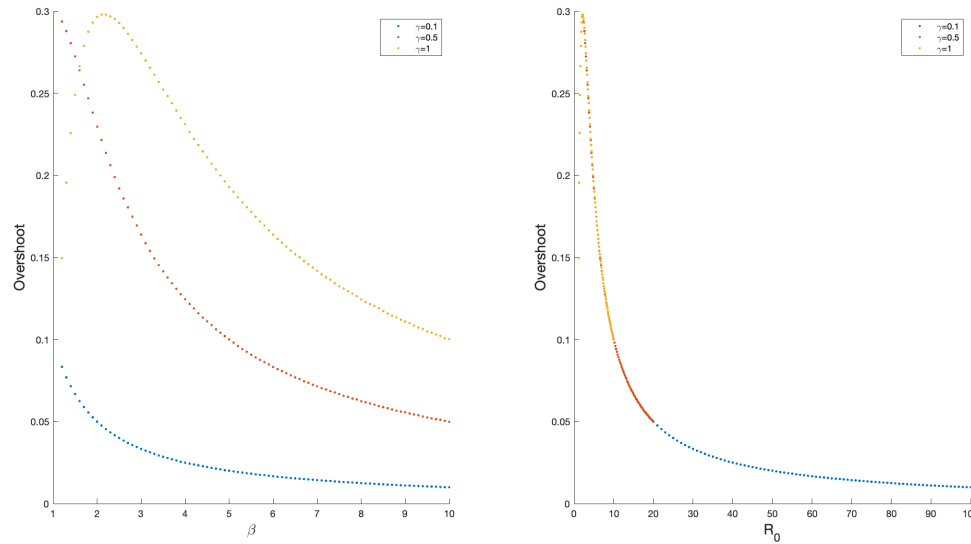


Figure S1: a) The overshoot as a function of parameter  $\beta$  and different  $\gamma$ . b) The same data as in (a) but plotted in terms of  $R_0 = \frac{\beta}{\gamma}$ .

**Figure S2. Fraction of individuals infected in the overshoot out of all individuals infected for SIR model.**

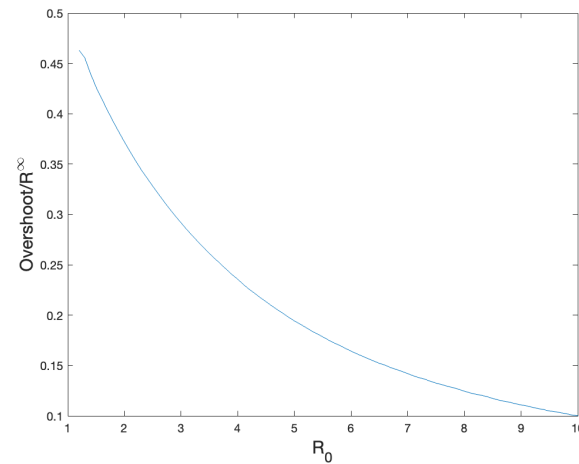


Figure S2: The ratio of overshoot to final attack rate as a function of  $R_0$ .