A tutorial in the mathematical theory of models for an epidemic with heterogeneity of susceptibility

Paul McKeigue

August 21, 2022

1 Introduction

This tutorial sets out in an accessible form the mathematical theory of epidemics based on a Susceptible-Infected-Removed model with heterogeneous susceptibility. The first general treatment of this was given by Novozhilov [2008a,b, 2012]. Other papers extended his derivation of the size of an epidemic Miller [2012], Katriel [2012] but without application to real data. Novozhilov did not consider the case in which infectivity was correlated with susceptibility; this was developed by Gomes et al. [2020] and Tkachenko et al. [2021].

Applications of this theory to the COVID-19 epidemic were posted as preprints from May 2020 onwards [Gomes et al., 2020, Colombo et al., 2020, Neipel et al., 2020]. A more general commentary on the limitations of mathematical models for guiding policy in the pandemic has been uploaded as a preprint [McKeigue and Wood, 2022].

2 Notation

The notation used here is a simplified version of that used in Novozhilov's articles, with some alterations to maintain consistency between quantities that are dimensionless and quantities that have dimension T^{-1} where T is the time dimension.

- The population is divided into three compartments: Susceptible S, Infected I and Removed V. These fractions sum to 1. Inclusion of an Exposed E compartment as an intermediate stage between S and I does not make any difference to the results derived here. I'll use V for the proportion removed, as we use $\mathcal{R}$ for the reproduction number, defined as the average number of new cases infected by each case.

- S, I and V are functions of time t. I use subscripts to denote dependence on t, so where relevant these are written as S_t, I_t, V_t. At the

start of the epidemic all individuals are susceptible, thus $S_0 = 1$. At the end of the epidemic, the proportion remaining susceptible is S_∞. The final size of the epidemic (final proportion infected) is $1 - S_\infty$.

- The **prevalence** of infection at time t is I_t
- The arrival rate of infection is $-\frac{dS_t}{dt}$
- The **incidence** of infection is the arrival rate divided by the population at risk. $-\frac{dS_t/dt}{S_t}$
- The **transmission function** gives the arrival rate $-\frac{dS_t}{dt}$ as a function of the susceptible fraction S_t and the prevalence of infection I_t. It has dimension T^{-1}
- The reproduction number $\mathcal{R}_t$ is the average number of individuals infected by each new case arising at time t. I use subscripts to denote the value at time t, and parentheses when the argument is the fraction of the population who are susceptible: thus $\mathcal{R}_0$ is the **basic reproduction number** at $t = 0$, and $\mathcal{R}(S)$ is the reproduction number when the fraction susceptible is S.
- **Susceptibility** ω is defined as a multiplicative factor, varying between individuals but a fixed attribute of each individual, that scales the incidence /prevalence ratio. ω is scaled to have mean 1 at $t = 0$.
- The incidence at time t in individuals having susceptibility 1 is the **force of infection**, a rather confusing term as it is a scalar, not a vector.
- $\langle\omega\rangle_t$ is the average of ω taken over its probability distribution among susceptible individuals. This use of angled brackets to denote an expectation is standard notation in physics.
- The ratio β of the incidence of infection in individuals whose susceptiblity is 1 to prevalence is the **transmission coefficient**. β has dimension T^{-1}.

3 Homogeneous model

For this analysis, the classic Susceptible(-Exposed)-Infectious-Removed model formulated by Kermack and McKendrick [1927] can be defined by two differential equations (the intermediate ones for the rate of change of the Exposed and Infectious compartments are not needed for this analysis.

$$\frac{dS}{dt} = -\beta S_t I_t \tag{1}$$

$$\frac{dV}{dt} = \gamma I_t \tag{2}$$

β is the **transmission coefficient**, defined above, and γ is the rate of removal from the infectious compartment I. β and γ have dimension T^{-1}.

At $t = 0$, $S = 1$ and the basic reproduction number $\mathcal{R}_0$ is $\frac{\beta}{\gamma}$

The reproduction number $\mathcal{R}(S)$ when the fraction susceptible is S is $\mathcal{R}_0 S$ and thus at the herd immunity threshold, when $\mathcal{R}(S) = 1$, $S = \frac{1}{\mathcal{R}_0}$

Dividing equation 1 by equation 2 we have

$$\frac{dS}{dV} = -\mathcal{R}_0 S \tag{3}$$

Integrating from $t = 0$ to $t = \infty$ we obtain an equation for the proportion S_∞ of the population who are still susceptible at the end of the epidemic

$$\int_1^{S_\infty} \frac{dS_t}{S_t} = -\mathcal{R}_0 \int_0^{V_\infty} dV_t \tag{4}$$

$$\log S_\infty = -\mathcal{R}_0(1 - S_\infty) \tag{5}$$

Taking exponents we have

$$S_\infty = e^{-\mathcal{R}_0(1-S_\infty)} \tag{6}$$

4 Heterogeneous susceptibility

We allow susceptibility, defined as a quantity that is fixed within individuals and scales the incidence / prevalence ratio, to vary between individuals. The distribution of susceptibility at time t is a mixture of a spike at zero in the fraction $(1 - S_t)$ of the population who are no longer susceptible, and a probability density over values of ω from 0 to infinity in the fraction S_t who are still susceptible. $s_t(\omega)d\omega$ is the fraction of the population who at time t have susceptibility between ω and $\omega + d\omega$.

$$\int_0^\infty s_t(\omega)d\omega = S_t \tag{7}$$

We write $p_t(\omega)$ for the density of susceptibility in the susceptible compartment at time t

$$p_t(\omega) = \frac{s_t(\omega)}{S_t} \tag{8}$$

$p_t(\omega)$ is a probability density that integrates to 1, but $s_t(\omega)$ is not a probability density except at $t = 0$: it integrates to S_t.

The average susceptibility in the susceptible compartment at time t is

$$\langle \omega \rangle_t = \int \omega_t p_t(\omega) d\omega \tag{9}$$

4.1 Incidence at time t in individuals with susceptibility ω

At time t, the incidence in individuals with susceptibility ω is proportional to the incidence in individuals with susceptibility 1.

$$\frac{ds_t(\omega)}{dt} \frac{1}{s_t(\omega)} = -\omega \frac{dq}{dt} \tag{10}$$

where $\frac{dq_t}{dt}$ is the force of infection defined as above.

The dimensionless function q_t is $\int_0^t \frac{dq_t}{dt} dt$ with the initial condition $q_0 = 0$. q_t can be defined as the expected number of infectious contacts received by an individual with susceptibility 1 up to time t (In Novozhilov's notation, $\frac{dq_t}{dt}$ is defined to be negative).

Integrating equation 10 with respect to t and taking exponents

$$s_t(\omega) = s_0(\omega) e^{-\omega q_t} \tag{11}$$

$$\tag{12}$$

As $s_0(\omega) = p_0(\omega)$ and $p_t(\omega) S_t = s_t(\omega)$, we can write the probability density at time t in terms of $p_0(\omega) e^{\omega q_t}$

$$p_t(\omega) = \frac{p_0(\omega) e^{-\omega q_t}}{S_t} \tag{13}$$

4.2 Susceptible fraction S at time t

Integrating equation 13 with respect to ω with initial conditions $S_0 = 1$ and $q_0 = 0$, we have

$$1 = \frac{1}{S_t} \int e^{-\omega q_t} p_o(\omega) d\omega \tag{14}$$

$$S_t = \mathcal{M}(-q_t) \tag{15}$$

where $\mathcal{M}(\theta)$ is the moment generating function of the probability density $p_0(\omega)$.

The moment generating function $\mathcal{M}_t(\theta)$ of the probability density $p_t(\omega)$ is

$$\mathcal{M}_t(\theta) = \int e^{\omega\theta} p_t(\omega) d\omega \tag{16}$$

$$\tag{17}$$

Using equation 13 to substitute for $p_t(\omega)$

$$\mathcal{M}_t(\theta) = \frac{1}{S_t} \int e^{\omega\theta} e^{-\omega q_t} p_0(\omega) d\omega \tag{18}$$

$$= \frac{\mathcal{M}(\theta - q_t)}{\mathcal{M}(-q_t)} \tag{19}$$

This equation was first obtained by Karev [2005] studying the growth of populations. Physicists will recognize this identity as expressing the moment generating function of the distribution of energy states in a canonical ensemble as the ratio of the partition function at coolness $(q_t - \theta)$ to the partition function at coolness q_t. $\mathcal{M}(-q_t)$ is the normalizing constant of the unnormalized density $s_t(\omega)$. We can thus think of q_t as the "coolness" of the epidemic. Initially the "coolness" of the epidemic is 0 (temperature is infinite); as the most susceptible individuals are removed, the epidemic cools (q_t increases).

It follows that the first and second moments of susceptibility ω within the susceptible compartment at time t are given by

$$\langle \omega_t \rangle = \frac{\mathcal{M}'(-q_t)}{S_t} \tag{20}$$

$$\langle \omega_t^2 \rangle = \frac{\mathcal{M}''(-q_t)}{S_t} \tag{21}$$

where $\mathcal{M}'()$ and $\mathcal{M}''()$ are respectively the first and second derivatives of the moment generating function.

From equation 19 we can derive the entire trajectory of the epidemic, including the herd immunity threshold and the final size of the epidemic given only the basic reproduction number $\mathcal{R}_0$ and the initial distribution $p(\omega)$ or its moment generating function. To scale the epidemic trajectory by time, we need to specify one of the rate parameters β or γ.

4.3 Effective susceptible fraction

We can write the transmission function in the form

$$\frac{dS_t}{dt} = -h(S_t)\frac{dq_t}{dt} \tag{22}$$

where $h(S_t)$ is the **effective susceptible fraction**, defined as $\langle\omega\rangle_t S_t$. We can express this equation for the transmission function in words as

Arrival rate = effective susceptible fraction × force of infection

We can write the effective susceptible fraction as

$$h(S) = \mathcal{M}'(-q_t) \tag{23}$$

$$= \mathcal{M}'(-\mathcal{M}^{-1}(S)) \tag{24}$$

as $q_t = -\mathcal{M}^{-1}(S_t)$

$h(S)$ depends only on S, so we can leave off the subscript t. Because the most susceptible individuals are selectively removed from the susceptible compartment, $h(S) < S$ after time $t = 0$ if there is heterogeneity.

The homogeneous model is a special case of this more general model, in which the density $p(\omega)$ is a spike at 1, $\mathcal{M}(\theta) = e^{\theta}$ and $h(S) = S$. Unless the model is homogeneous, $h(S)$ will be a non-linear function of S. **Specifying a transmission function that is a linear function of S is thus equivalent to assuming no heterogeneity of susceptibility (beyond whatever stratification is specified explicitly in the model)**. Any heterogeneity of susceptibility in the population (as long as the moment generating function of the initial probability distribution exists) can be represented by a non-linear transmission function.

The empirical use of of a non-linear transmission function dates back to the 1990s (Anderson and May 1992; Antonovics et al. 1995; Mollison 1995) but no theoretical basis for this was available until 1997, when Dwyer et al. [1997] showed that a gamma distribution of susceptibility would give rise to a power law for the transmission function.

4.4 Reproduction number

The reproduction number $\mathcal{R}_t$ at time t is the average reproduction number over all new cases at time t.

Write $r(\omega)$ for the expected number infected by a single case with susceptibility ω in a fully susceptible population with average susceptibility 1.

At time t, the probability density of susceptibility among new cases will be proportional to $\omega p_t(\omega)$. If the infectious period is short, the distribution of susceptibility in the infectious state at time t can be approximated by the distribution among new cases arising at time t. The expected number infected by a single case with susceptibility ω is then $S_t\langle\omega\rangle_t r(\omega)$.

$$\mathcal{R}_t = \int S_t \langle\omega\rangle_t r(\omega) \frac{\omega p_t(\omega)}{\langle\omega\rangle_t} d\omega \tag{25}$$

$$= S_t \int r(\omega)\omega e^{-\omega q_t} p_0(\omega) d\omega \tag{26}$$

In the limiting cases $r_t(\omega) = \mathcal{R}_0$ (uncorrelated susceptibility) and $r_t(\omega) = \mathcal{R}_0\omega$ (heterogeneous connectivity) this expression can be evaluated in terms of derivatives of the moment generating function as described below.

5 Two limiting cases of heterogeneity

Two possible models for heterogeneity are: (a) susceptibility is uncorrelated with infectivity; (b) susceptibility and infectivity perfectly correlated. Where individual susceptibility depends only on pre-existing resistance to infection, model (a) would be realistic. Where individual susceptibility depends only on the number of contacts, infectivity will be highly correlated with susceptibility. For these two limiting cases (a) and (b), simple expressions can be derived for the reproduction number and the epidemic trajectory.

5.1 Model (a): uncorrelated susceptibility

Extending the mass action kinetics underlying the Kermack-McKendrick model, the equation for the incidence rate at time t in individuals having susceptibility ω is

$$\frac{\partial s_t(\omega)}{\partial t} \frac{1}{s_t(\omega)} = \beta\omega I_t \tag{27}$$

For the total population, the arrival rate is

$$\frac{dS_t}{dt} = -\beta \langle\omega\rangle_t S_t I_t \tag{28}$$

The force of infection $-\dfrac{dq_t}{dt}$ is βI_t and the transmission function is $-\beta \langle\omega\rangle_t S_t$.

The reproduction number at time t is

$$\frac{\mathcal{R}_t}{\mathcal{R}_0} = S_t \int \omega e^{-\omega q_t} p_o(\omega) d\omega \tag{29}$$

$$= S_t \frac{\mathcal{M}'(-q_t)}{S_t} = h(S_t) \tag{30}$$

Thus with uncorrelated susceptibility, the reproduction number when the proportion susceptible is S depends only on the effective susceptible fraction and the basic reproduction number $\mathcal{R}_0$.

$$\frac{\mathcal{R}(S)}{\mathcal{R}_0} = h(S) \tag{31}$$

To obtain an equation for the proportion susceptible at the end of the epidemic we again use $\frac{dS}{dV}$ and integrate from $t = 0$ to $t = \infty$

$$\frac{dS}{dV} = \mathcal{R}_0 h(S) \tag{32}$$

$$\int_1^{S_\infty} \frac{dS_t}{h(S_t)} = \mathcal{R}_0 \int_0^{V_\infty} dV_t \tag{33}$$

$$= -\mathcal{R}_0(1 - S_\infty) \tag{34}$$

5.2 Model (b): heterogeneous connectivity

The incidence rate at time t in infection in individuals having susceptibility ω is proportional to the average value of ω among those infected. If the infectious period is short, we can disregard the change in the distribution of ω in the population from the time at which those who are in the infectious compartment became infected to time t.

The incidence rate at time t in individuals having susceptibility ω (force of infection) is proportional to the average value of ω in the susceptible compartment at time t

$$\frac{ds_t(\omega)}{dt}\frac{1}{s_t(\omega)} = -\beta\omega \left\langle \omega \right\rangle_t I_t \tag{35}$$

$$\tag{36}$$

For the total population, the arrival rate is

$$\frac{dS_t}{dt} = -\beta h(S) \left\langle \omega \right\rangle_t S_t I_t \tag{37}$$

The force of infection $\dfrac{dq_t}{dt}$ is $\beta \left\langle \omega \right\rangle_t I_t$

With heterogeneous connectivity, the reproduction number depends also on the average susceptibility in the susceptible compartment because that determines the average infectivity of those who are infected.

$$\frac{\mathcal{R}_t}{\mathcal{R}_0} = \frac{1}{\langle\omega^2\rangle_0} S_t \int \omega^2 e^{-\omega q_t} p_o(\omega) d\omega \tag{38}$$

$$= \frac{1}{\langle\omega^2\rangle_0} S_t \frac{\mathcal{M}''(-q_t)}{S_t} \tag{39}$$

$$= \frac{\langle\omega^2\rangle_t}{\langle\omega^2\rangle_0} = \frac{\mathcal{M}''(-q_t)}{\mathcal{M}''(0)} \tag{40}$$

The equation for the proportion susceptible at the end of the epidemic is derived as before

$$\frac{dS}{dV} = \mathcal{R}_0 h(S) \langle\omega\rangle_t \tag{41}$$

$$\int_1^{S_\infty} \frac{dS_t}{h(S_t)\langle\omega\rangle_t} = -\mathcal{R}_0(1 - S_\infty) \tag{42}$$

These results, which apply to any distribution of susceptibility as long as the moment generating function of the distribution exists, are summarized in Table 1.

Table 1: Summary table of results under the two limiting models of heterogeneity

Quantity		
Transmission function $\frac{dS_t}{dt}$	$-h(S)\frac{dq_t}{dt}$	
Susceptible fraction S_t	$\int p_0(\omega)e^{-\omega q_t}d\omega = \mathcal{M}(-q_t)$	
Effective susceptible fraction $h(S_t)$	$\mathcal{M}'(-q_t) = \mathcal{M}'(-\mathcal{M}^{-1}(S))$	
Average susceptibility $\langle\omega\rangle_t$ in susceptible individuals at time t	$\frac{\mathcal{M}'(-q_t)}{\mathcal{M}(-q_t)} = \frac{h(S_t)}{S_t}$	
	Uncorrelated susceptibility	Connectivity
Basic reproduction number $\mathcal{R}_0$	$\frac{\beta}{\gamma}$	$\langle\omega^2\rangle_0 \frac{\beta}{\gamma}$
Force of infection $-\frac{dq_t}{dt}$	βI_t	$\beta \langle\omega\rangle_t I_t$
$\frac{\mathcal{R}(S)}{\mathcal{R}_0}$	$S_t\langle\omega\rangle_t = h(S)$	$\frac{\langle\omega^2\rangle_t}{\langle\omega^2\rangle_0}$
$\frac{dS_t}{dV_t}$	$\mathcal{R}_0 h(S)$	$\mathcal{R}_0 h(S) \langle\omega\rangle_t$
Equation for susceptible fraction S_∞ at end of epidemic	$\int_1^{S_\infty} \frac{dS_t}{h(S_t)} = -\mathcal{R}_0(1-S_\infty)$	$\int_1^{S_\infty} \frac{dS_t}{h(S_t)\langle\omega\rangle_t} = -\mathcal{R}_0(1-S_\infty)$

6 Gamma distribution of susceptibility

Susceptibility ω as defined can take only positive values and the mean susceptibility at $t = 0$ is set to 1: otherwise the transmission coefficient β would not be identified. This constrains the form of the distribution of ω.

We expect susceptibility to vary as the result of many factors including immune status and social connectivity. This suggests that the distribution of susceptibility will tend to a state of maximum entropy, subject to whatever constraints exist. Two distributions that maximize the entropy given the mean and one other sufficient statistic are the log-normal distribution and the gamma distribution: the shape of these distributions is very similar except that for a given mean and variance the log-normal distribution has heavier tails. The gamma distribution is more mathematically convenient.

The gamma distribution has two parameters: shape and inverse scale. As the mean of the distribution is equal to the shape parameter divided by the inverse scale parameter. Specifying the density $p_0(\omega)$ as a gamma density with mean 1 specifies that the inverse scale parameter is equal to the shape parameter α, and that the variance is $1/\alpha$.

The shape parameter can take positive values up to infinity. As α tends to infinity, the density tends to a spike at 1 and the model becomes the homogenenous model. Other authors use the coefficient of variation, which is $1/\sqrt{\alpha}$, instead of the shape parameter α.

The moment generating function of a gamma distribution with shape parameter and inverse scale parameter both set to α is

$$\mathcal{M}(\theta) = \left(1 - \frac{\theta}{\alpha}\right)^{-\alpha} \tag{43}$$

The moment generating function for the distribution of susceptibility among those remaining susceptible at time t is As the epidemic ages (q_t becomes more negative), the average susceptibility of those who remain susceptible falls but the shape parameter α does not change.

Expressions for $h(S)$, $\mathcal{R}(S)$, and S_∞ where susceptibility has a gamma distribution with mean 1 are given in Table 2 . Deriving these from $\mathcal{M}(\theta)$ is left as an exercise for the reader.

The ratio $\dfrac{\mathcal{R}(S)}{\mathcal{R}_0}$ is equal to $S^{1+1/\alpha}$ for uncorrelated susceptibility, and $S^{1+2/\alpha}$ for heterogenous connectivity. As a function of the susceptible fraction S, this ratio obeys a power law.

Tkachenko et al. [2021] use the term **immunity coefficient** and the symbol λ for the exponent. The proportion susceptible at the herd immunity threshold is $\mathcal{R}_o^{-1/\lambda}$

The transmission functions generated by specifying other families of distributions for susceptibility are well approximated by a power law. This is not surprising, because the function $\dfrac{\mathcal{R}(S)}{\mathcal{R}_0}$ is constrained: it has to be monotonic and convex downwards, falling from $\mathcal{R}_0$ when $S = 1$ at the start of the epidemic to zero at $S = 0$.

Figure 1: Transmission function and trajectory of epidemic with different values of the immunity coefficient. Vertical dotted lines are at one minus the herd immunity threshold

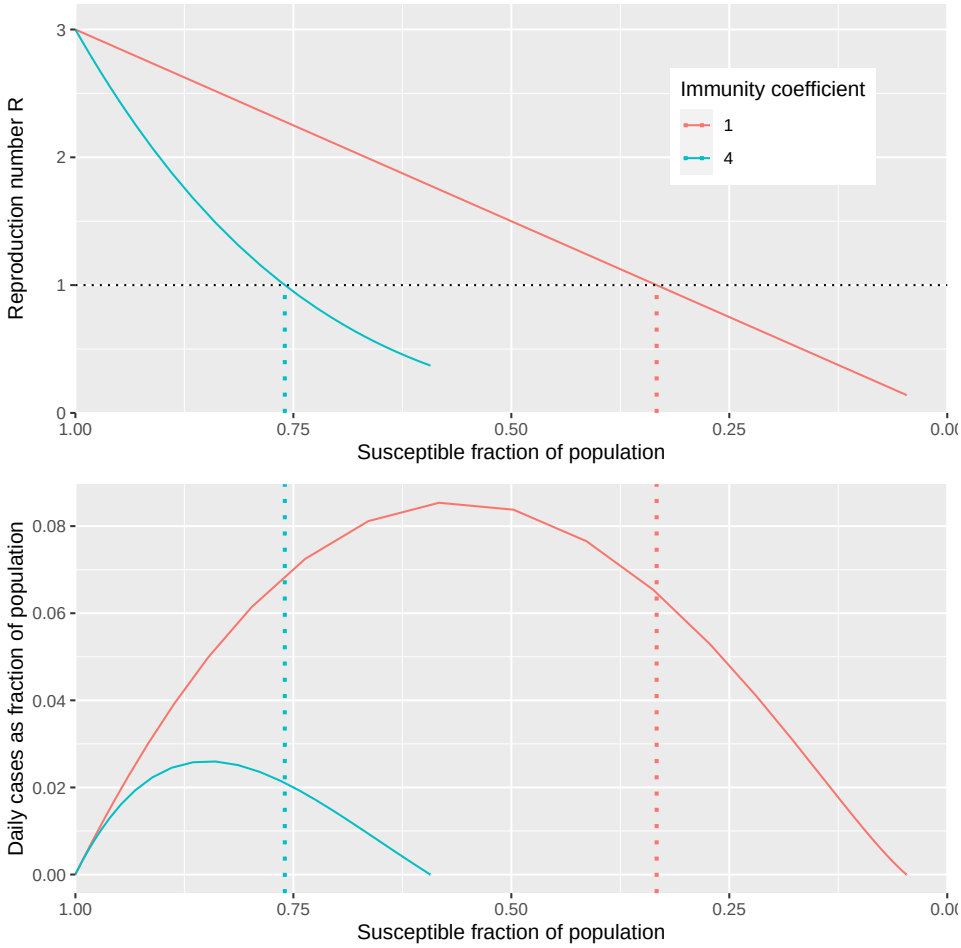

Table 2: Summary table of results for the two limiting models where susceptibility has a gamma distribution with mean 1 and shape parameter α

Quantity	Uncorrelated susceptibility	Connectivity
Moment generating function $\mathcal{M}(\theta)$	$\left(1-\frac{\theta}{\alpha}\right)^{-\alpha}$	
Coolness of epidemic $q_t = \mathcal{M}^{-1}(S_t)$	$\alpha(S_t^{-1/\alpha} - 1)$	
Effective susceptible fraction $h(S)$	$S^{1+1/\alpha}$	
Average susceptibility $\langle\omega\rangle_t$ in susceptible individuals at time t	$S_t^{1/\alpha}$	
Average squared susceptibility $\langle\omega^2\rangle_t$ in susceptible individuals at time t	$\left(1+\frac{1}{\alpha}\right) S_t^{1+2/\alpha}$	
Transmission function	$-\beta S_t^{1+1/\alpha} I_t$	$-\beta\left(1+\frac{1}{\alpha}\right) S_t^{1+2/\alpha} I_t$
$\frac{dS}{dV}$	$-\mathcal{R}_0 S_t^{1+1/\alpha}$	$-\mathcal{R}_0 S_t^{1+2/\alpha}$
$\frac{\mathcal{R}(S)}{\mathcal{R}_0}$	$S^{1+1/\alpha}$	$S^{1+2/\alpha}$
Susceptible fraction at end of epidemic S_∞	$S_\infty = \left(1+\mathcal{R}_0\frac{1-S_\infty}{\alpha}\right)^{-\alpha}$	$S_\infty = \left(1+\mathcal{R}_0\frac{1-S_\infty}{\alpha/2}\right)^{-\alpha/2}$

.

Although models with uncorrelated susceptibility and heterogeneous connectivity cannot be distinguished from the trajectory of a single epidemic wave, these two types of heterogeneity have different practical implications. The uncorrelated susceptibility model predicts that herd immunity will be stable until a new variant appears that has a different susceptibility profile. The connectivity model predicts that herd immunity will be only

transient, and that new waves are likely to arise when the social network is rewired. If variable connectivity contributes to heterogeneity, imposing "circuit-breaker" restrictions that disrupt and rewire social networks is likely to increase the long-term size of the epidemic.

7 References

References

Artem S Novozhilov. On the spread of epidemics in a closed heterogeneous population. *Math Biosci*, 215(2):177–185, October 2008a. ISSN 0025-5564. doi: 10.1016/j.mbs.2008.07.010. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2580825/.

A. Novozhilov. Heterogeneous Susceptibles–Infectives model: Mechanistic derivation of the power law transmission function. In *Proceedings of the 6th International Conference Differential Equations and Dynamical Systems*, Baltimore, MD, 2008b. URL https://arxiv.org/pdf/0809.1578.pdf.

A. S. Novozhilov. Epidemiological Models With Parametric Heterogeneity : Deterministic Theory for Closed Populations. *Math. Model. Nat. Phenom.*, 7(3):147–167, 2012. ISSN 0973-5348, 1760-6101. doi: 10.1051/mmnp/20127310. URL https://www.mmnp-journal.org/articles/mmnp/abs/2012/03/mmnp201273p147/mmnp201273p1

Joel C. Miller. A note on the derivation of epidemic final sizes. *Bull Math Biol*, 74(9):2125–2141, September 2012. ISSN 0092-8240. doi: 10.1007/s11538-012-9749-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3506030/.

Guy Katriel. The size of epidemics in populations with heterogeneous susceptibility. *J. Math. Biol.*, 65(2):237–262, August 2012. ISSN 1432-1416. doi: 10.1007/s00285-011-0460-2. URL https://doi.org/10.1007/s00285-011-0460-2.

M. Gabriela M. Gomes, Rodrigo M. Corder, Jessica G. King, Kate E. Langwig, Caetano Souto-Maior, Jorge Carneiro, Guilherme Goncalves, Carlos Penha-Goncalves, Marcelo U. Ferreira, and Ricardo Aguas. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *medRxiv*, page 2020.04.27.20081893, May 2020. doi: 10.1101/2020.04.27.20081893. URL https://www.medrxiv.org/content/10.1101/2020.04.27.20081893v3.

Alexei V. Tkachenko, Sergei Maslov, Ahmed Elbanna, George N. Wong, Zachary J. Weiner, and Nigel Goldenfeld. Time-dependent heterogeneity

leads to transient suppression of the COVID-19 epidemic, not herd immunity. *Proc Natl Acad Sci U S A*, 118(17):e2015972118, April 2021. ISSN 1091-6490. doi: 10.1073/pnas.2015972118.

Marco Colombo, Joseph Mellor, Helen M. Colhoun, M. Gabriela M. Gomes, and Paul M. McKeigue. Trajectory of COVID-19 epidemic in Europe. *medRxiv*, page 2020.09.26.20202267, September 2020. doi: 10.1101/2020.09.26.20202267. URL https://www.medrxiv.org/content/10.1101/2020.09.26.20202267v1.

Jonas Neipel, Jonathan Bauermann, Stefano Bo, Tyler Harmon, and Frank Jülicher. Power-law population heterogeneity governs epidemic waves. *PLoS One*, 15(10):e0239678, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0239678.

Paul M. McKeigue and Simon N. Wood. Limitations of models for guiding policy in the COVID-19 pandemic, June 2022. URL https://www.medrxiv.org/content/10.1101/2022.06.30.22277091v1.

William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772): 700–721, August 1927. doi: 10.1098/rspa.1927.0118. URL https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118.

Georgy P. Karev. Dynamics of heterogeneous populations and communities and evolution of distributions. *Discrete and Continuous Dynamical Systems*, 2005(Special):487, 2005. doi: 10.3934/proc.2005.2005.487. URL https://www.aimsciences.org/article/doi/10.3934/proc.2005.2005.487.

G. Dwyer, J. S. Elkinton, and J. P. Buonaccorsi. Host heterogeneity in susceptibility and disease dynamics: Tests of a mathematical model. *Am. Nat.*, 150(6):685–707, December 1997. ISSN 0003-0147. doi: 10.1086/286089.

Antonio Montalbán, Rodrigo M. Corder, and M. Gabriela M. Gomes. Herd immunity under individual variation and reinfection. *arXiv:2008.00098 [physics, q-bio]*, January 2022. URL http://arxiv.org/abs/2008.00098.

8 Moment generating functions

The probability density function $p(\omega)$ can be expressed in an alternative form as a moment generating function.

$$\mathcal{M}(\theta) = \int e^{\theta\omega} p(\omega) d\omega \tag{44}$$

Some properties of the moment generating function are:

- There is a one-to-one mapping between the probability density function and the moment generating function: inverting the function $\mathcal{M}(\theta)$ recovers the function $p(\omega)$.
- $\mathcal{M}(\theta)$ can take only positive values
- If ω can take only positive values, $\mathcal{M}(\theta)$ is an increasing function of θ
- $\mathcal{M}(0) = 1$
- the nth moment of the probability distribution of ω can be obtained as the nth derivative of $\mathcal{M}(\theta)$ with respect to θ, evaluated at $\theta = 0$.

$$\langle \omega^n \rangle = \left. \frac{d^n \mathcal{M}(\theta)}{d\theta^n} \right|_{\theta=0} \tag{45}$$

Table 3: Notation used by other writers

This tutorial	Tkachenko et al. [2021]	Montalbán et al. [2022]	Neipel et al. [2020]
Force of infection $-\frac{dq_t}{dt}$	$J(t)$	λ	$\dot{\tau}$
Susceptibility ω	α	x	x
Average susceptibility $\langle\omega\rangle_t$ in susceptible compartment at time t	$\langle\alpha\rangle$	$\bar{\boldsymbol{S}}$ / S	
Effective susceptible fraction $h(S) = S_t\langle\omega\rangle_t$	S_e	$\bar{\boldsymbol{S}}$	$\bar{x}$
Coolness of epidemic q_t	$Z(t)$		τ
Moment generating function $\mathcal{M}(-q_t)$	$\mathcal{M}_\alpha(-Z(t))$		$S(\tau)$