Superspreading Poisson mixture branching process
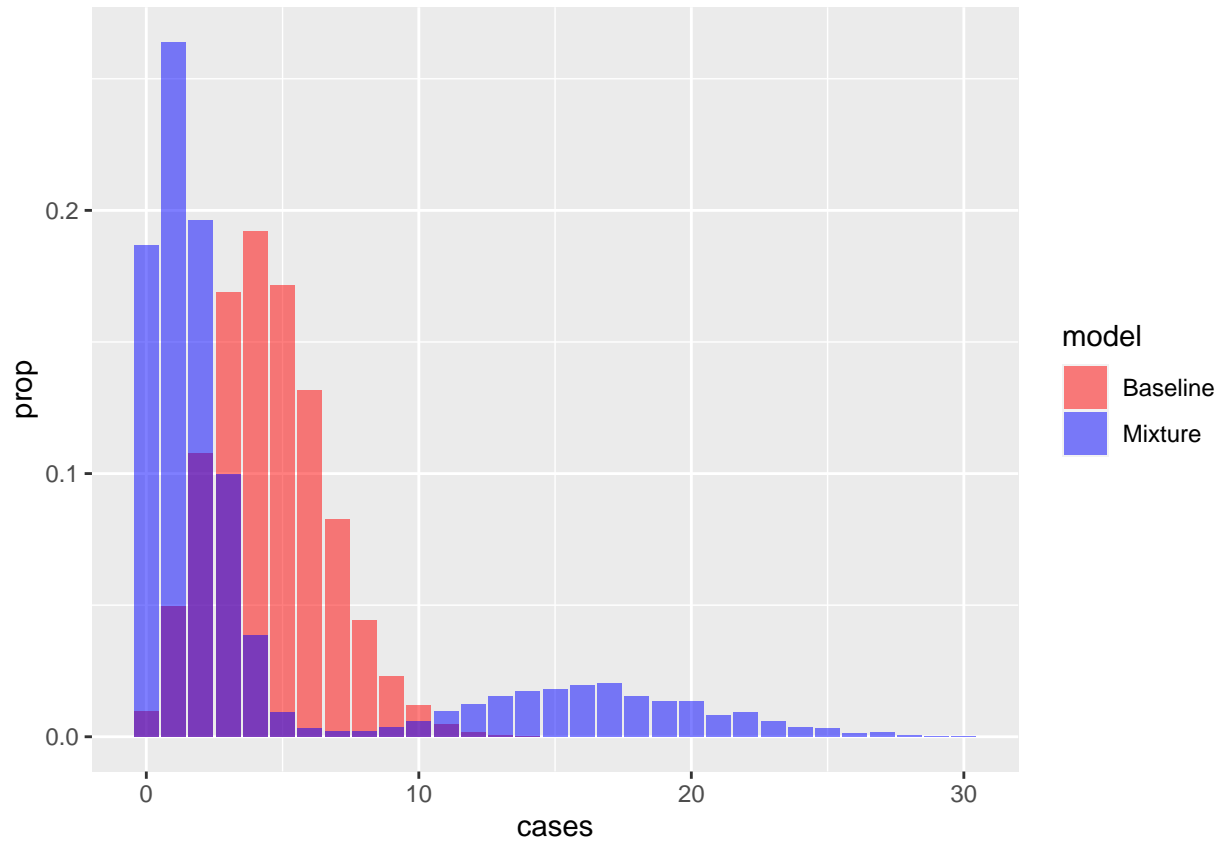
Offspring distributions

Figure. Poisson mixture offspring distribution using actuar package compared with Poisson offspring distribution with same mean $R_0 = 4.5$.
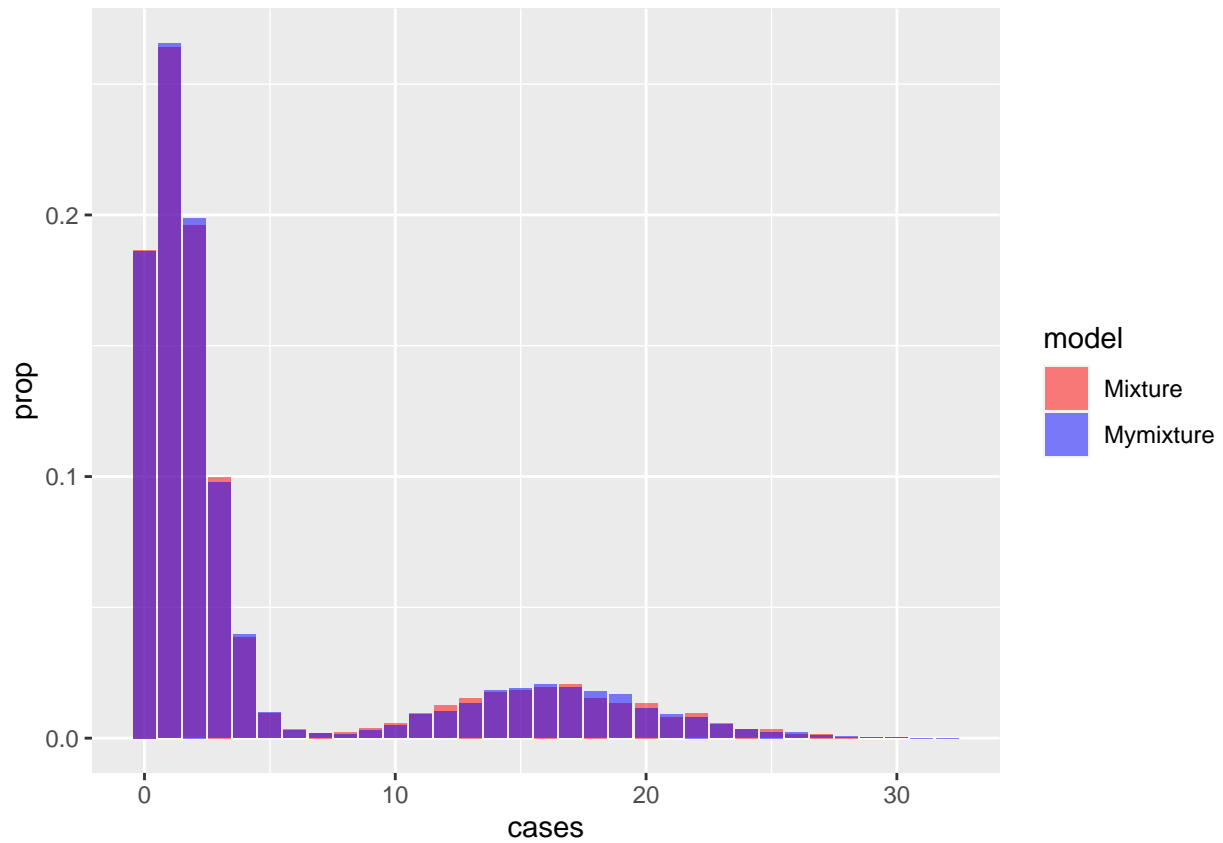
Figure: Poisson mixture offspring distribution using actuar package compared with Poisson offspring distribution generated using my function. My function works as well as the Poisson mixture function from the actuar library.
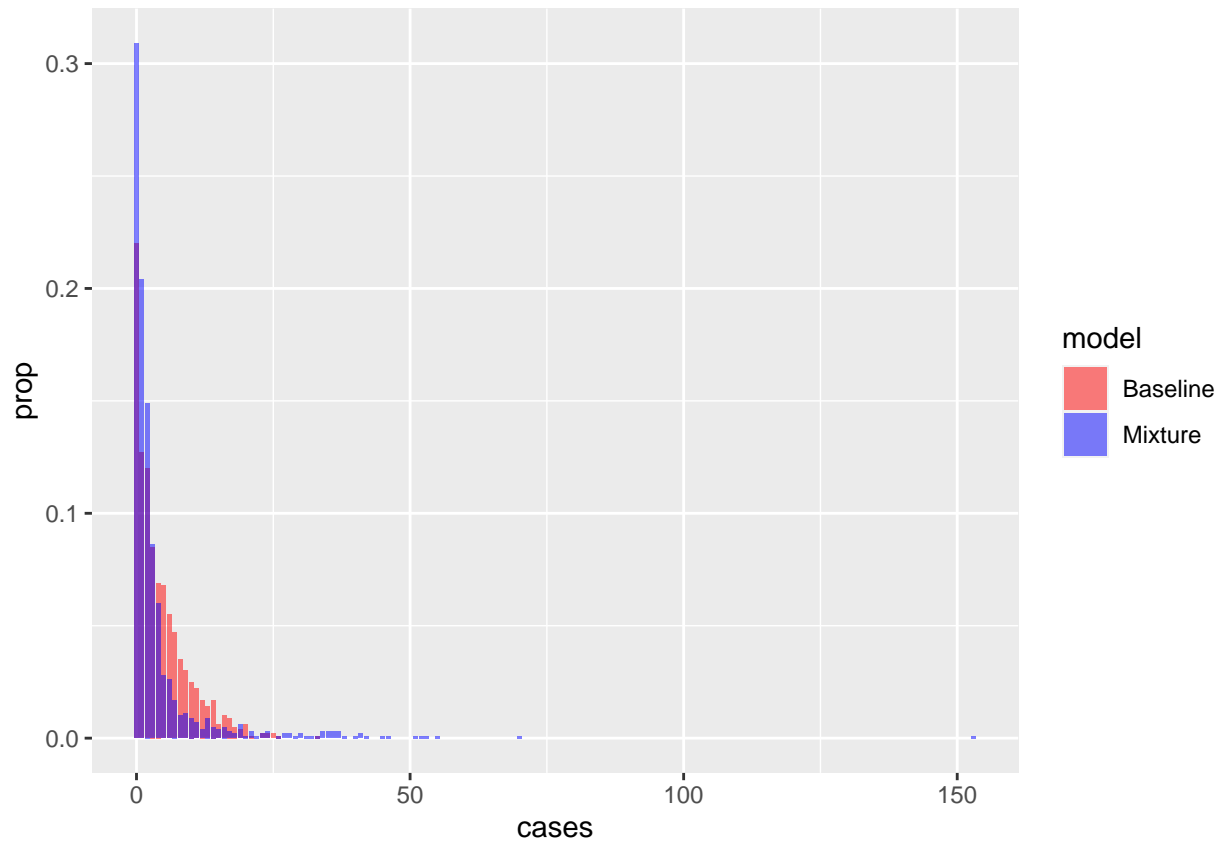
Figure. Mixture of geometric distributions (actuar package) has longer tail than geometric distribution, and the probability that cases is zero is much higher.
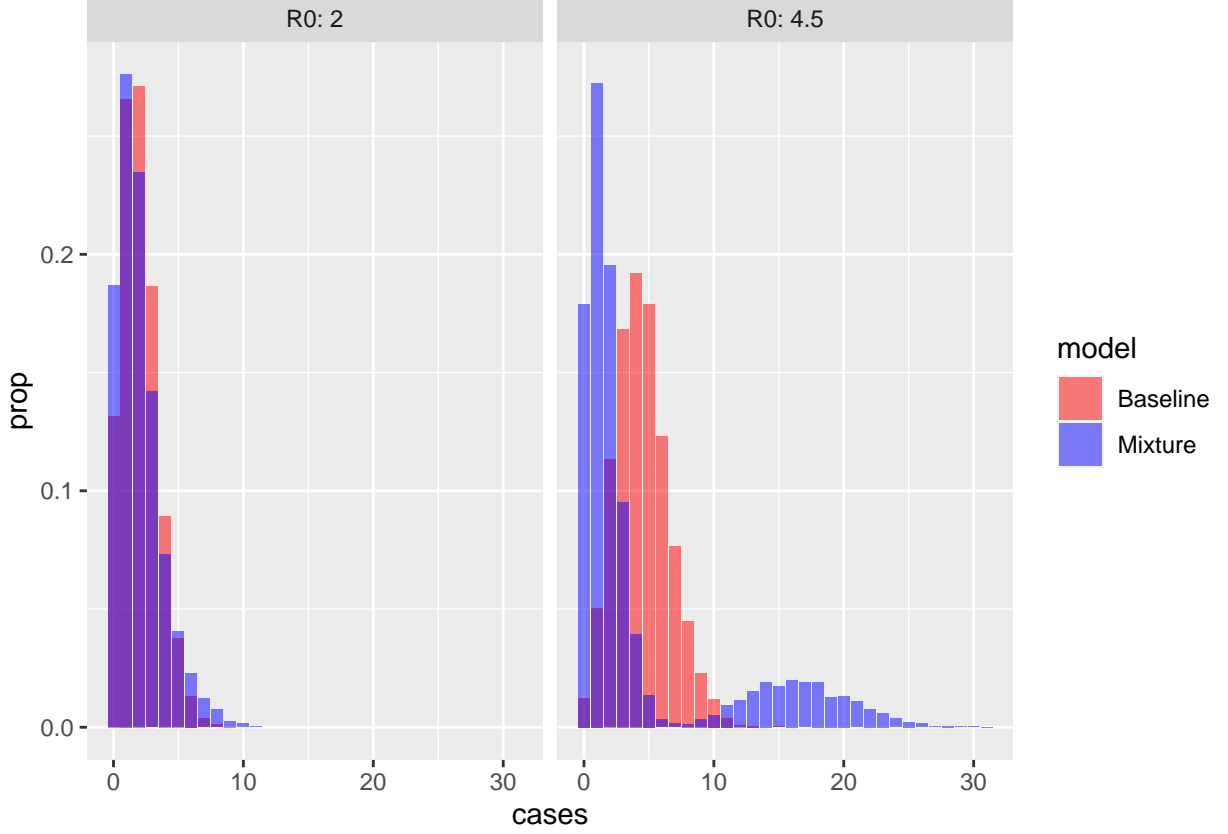
Figure 1. Comparing offspring distributions for R_0. Poisson mixture is bimodal if R_0 is sufficiently large

Probability mass function of the superspreading branching process

- The probability mass function that describes the superspreading branching process is a finite Poisson mixture (Johnson and Hoboken, 2005)

$$P(X = k) = \sum_{k=0}^{\infty} \left(p \frac{(R_0^D)^k}{k!} e^{-R_0^D} + (1-p) \frac{(R_0^A)^k}{k!} e^{-R_0^A} \right)$$

$$= p \sum_{k=0}^{\infty} \frac{(R_0^D)^k}{k!} e^{-R_0^D} + (1-p) \sum_{k=0}^{\infty} \frac{(R_0^A)^k}{k!} e^{-R_0^A}.$$

- The proportion of individuals that do not transmit the disease may be obtained from transmission trees (Lloyd Smith et al. 2005). Figure 1 indicates that the proportion of cases that lead to no secondary transmission is greater in the Poisson mixture process than for the Poisson process with the same R_0, and it decreases with the proportion of aerosol transmission $1 - p$.
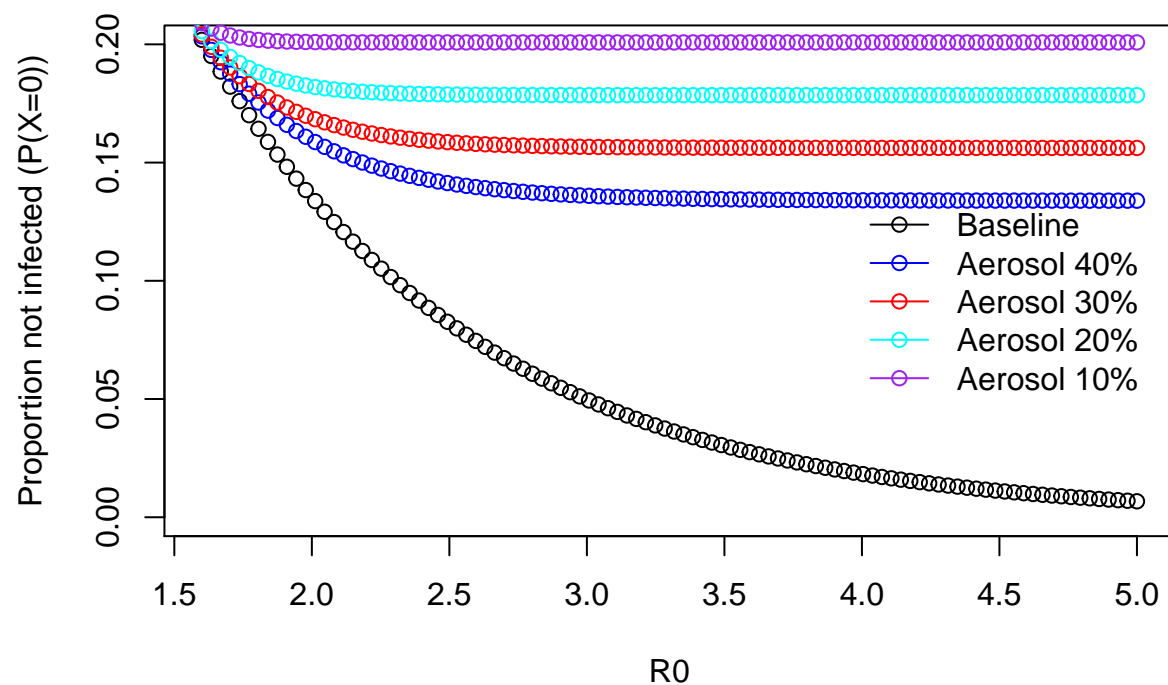
Figure. For each value of R_0, the mixture models predict that a higher proportion of individuals remain uninfected than the baseline Poisson model, and the proportion decreases with increasing aerosol transmission.

Probability of extinction of the superspreading branching process when $R_0 > 1$
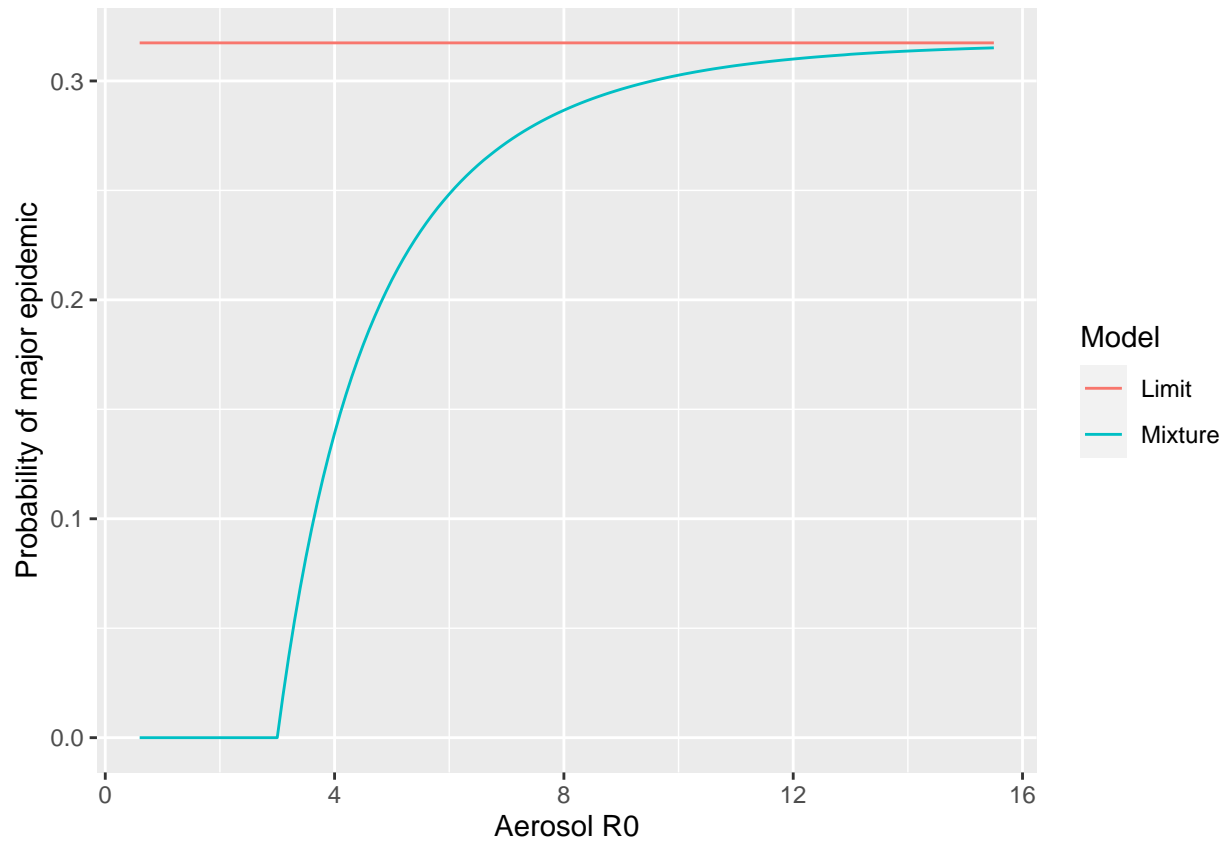

Figure. The probability of a major epidemic is an increasing function of R_0^A provided $R_0 > 1$. As R_0^A increases, the probability of a minor epidemic z_∞ decreases, and saturates to the solution of $z = pe^{R_0^D(z-1)}$. Then the probability of a major epidemic $1 - z_\infty$ saturates to the limit indicated by the horizontal line. Here $p = 0.8$, $R_0^D = 0.5$ and the limiting solution is $1 - z_\infty = 1 - 0.6826 = 0.3174$. At $R_0^A = 3$, $R_0 = 1$, and for values of $R_0^A < 3$, all outbreaks go extinct.
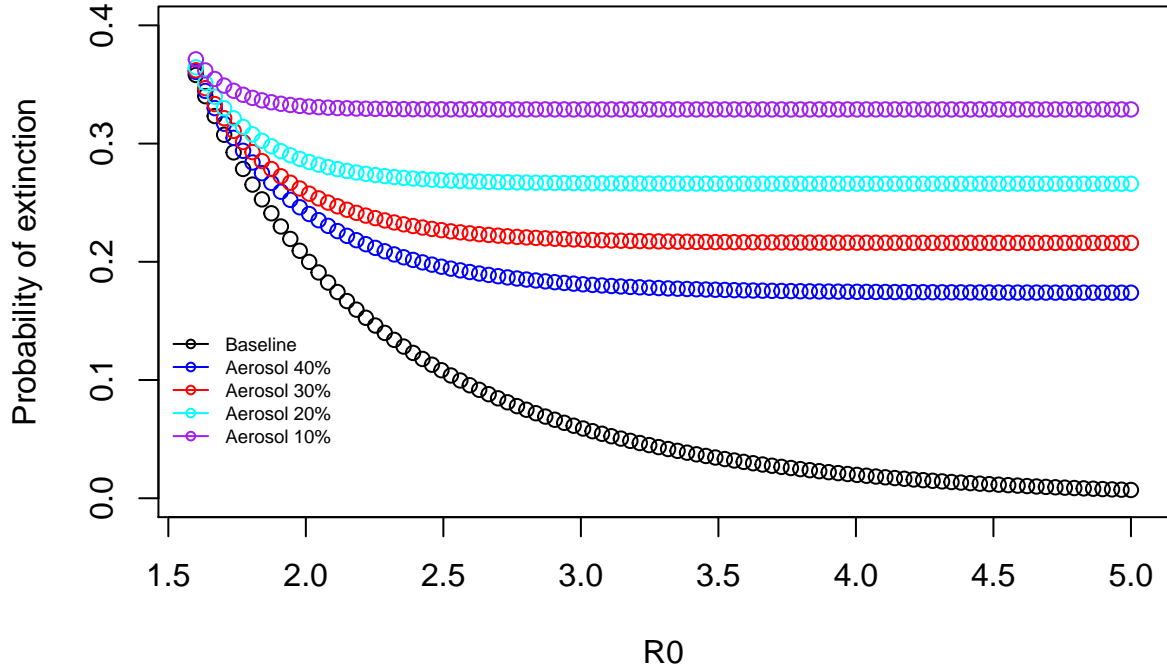
Probability of extinction of a minor outbreak

Figure The probability of extinction for the baseline Poisson model declines as R_0 increases. The mixture models have much higher probability of extinction for the same R_0 (note here $R_0^D = 1.5$.) The probability of extinction increases as the proportion of aerosol transmission decreases.

Variance of the superspreading branching process

- The probability generating function can be used to find the mean and variance of the offspring distribution. The expected number of cases is

$$G'(1) \ = pR_0^D + (1-p)R_0^A \ = R_0.$$

The mean is an increasing function of R_0^D and R_0^A. As R_0^D approaches zero from the right, R_0 approaches $(1-p)R_0^A$, and as R_0^A approaches R_0^D from the right, R_0 approaches R_0^D.

- The variance of the number of cases is

$$
\begin{aligned}
V(X) &= G''(1) + G'(1) - (G'(1))^2 \\
&= p(R_0^D)^2 + (1-p)(R_0^A)^2 + pR_0^D + (1-p)R_0^A - (pR_0^D + (1-p)R_0^A)^2 \\
&= p(R_0^D)(1 + R_0^D) + (1-p)(R_0^A)(1 + R_0^A) - (pR_0^D + (1-p)R_0^A)^2.
\end{aligned}
\tag{1}
$$

Assuming no direct contact (i.e., $p = 0$) for any value of R_0^A the mixture reduces to a Poisson process with mean R_0^A and variance is $V(X) = R_0^A$. Assuming no aerosol transmission ($p = 1$) for any value of R_0^D the mixture reduces to a Poisson process with mean R_0^D and variance R_0^D. Note that the variance can also be obtained using expectations, $V(X) = E(X^2) - (E(X))^2$.

- Letting $R_0^A = R_0^D + \delta$ and $R_0 = R_0^D + (1-p)\delta$ yields the following expression for the variance,

$$V(X) = R_0^D + (1-p)\delta + p(1-p)\delta^2. \tag{2}$$

Clearly, the variance is greater than the mean R_0 provided $0 < p < 1$.

- The second derivative test yields $V_{pp}(X) = -2\delta^2 < 0$ for $\delta > 0$ and so the variance has a local maximum. The variance has a local maximum at

$$p = \frac{1}{2}\left(1 - \frac{1}{\delta}\right),\qquad(3)$$

which is positive if $\delta > 1$ and approaches $1/2$ as $\delta \to \infty$ (i.e., as R_0^A gets increasingly large.) If $0 < \delta < 1$, then the local maximum occurs for negative p and therefore the variance is a decreasing function of p (an increasing function of the proportion of aerosol transmission). If $\delta > 1$ then the variance is non-monotonic on the interval $0 < p < 1$.

- The variance is a concave-up function of δ, with local minimum at $\delta = -1/2p$, and therefore the variance is a increasing function of δ on the interval $0 \le p \le 1$. Therefore the variance is an increasing function of R_0^A.

- The skewness of the offspring distribution is

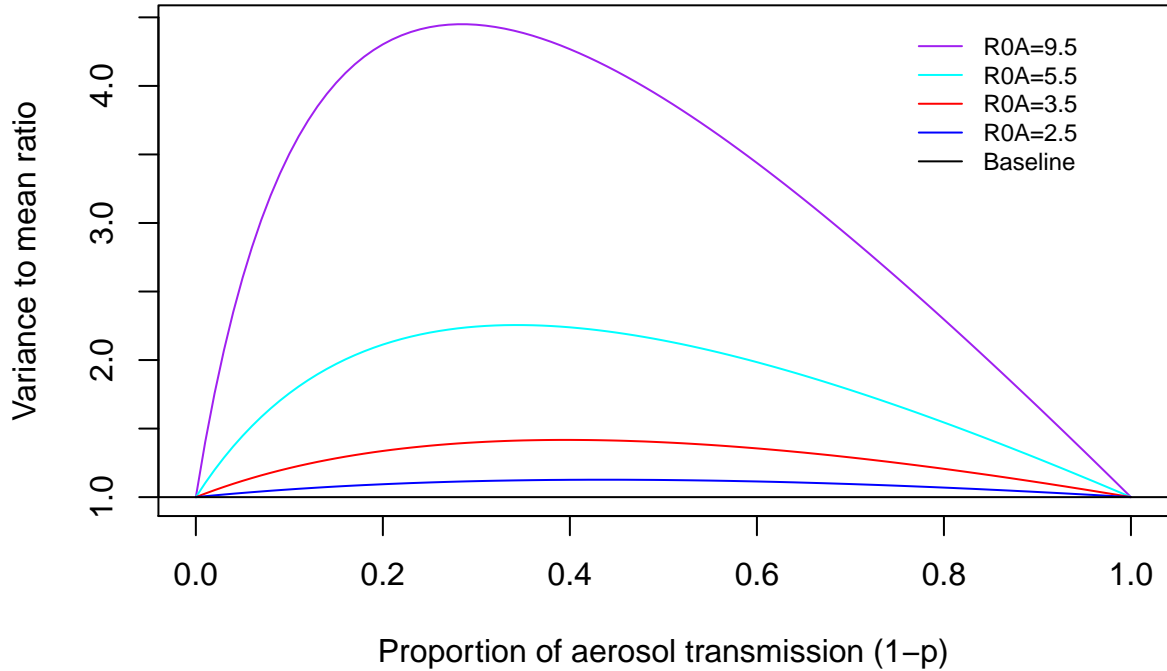$$S(X) = \frac{R_0^D + (1-p)\delta + 3p(1-p)\delta^2 + p(1-p)(2p-1)\delta^3}{V(X)^{3/2}}\qquad(4)$$

Figure. The variance to mean ratio of the offspring distribution is a non-monotonic function of the proportion of aerosol transmission, and is at its highest for low to moderate levels of aerosol transmission. The Poisson mixture models are overdispersed for $0 < p < 1$.
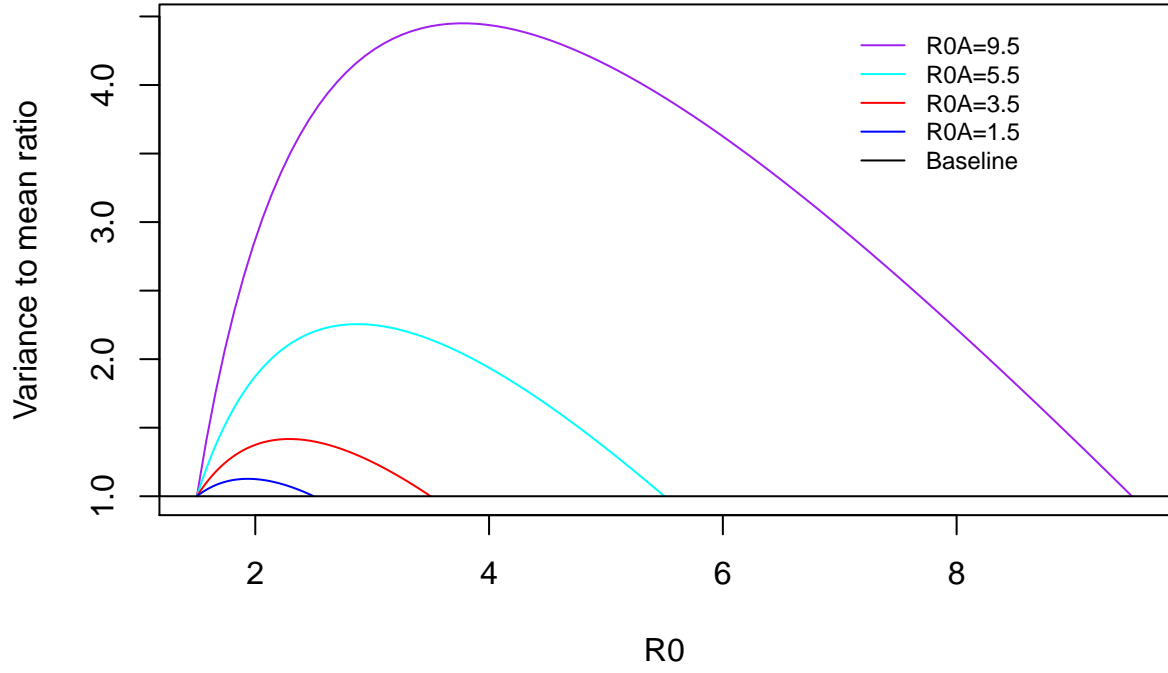
8

Figure. The variance to mean ratio of the offspring distribution is a non-monotonic function of R_0, and is larger for larger values of R_0^A.
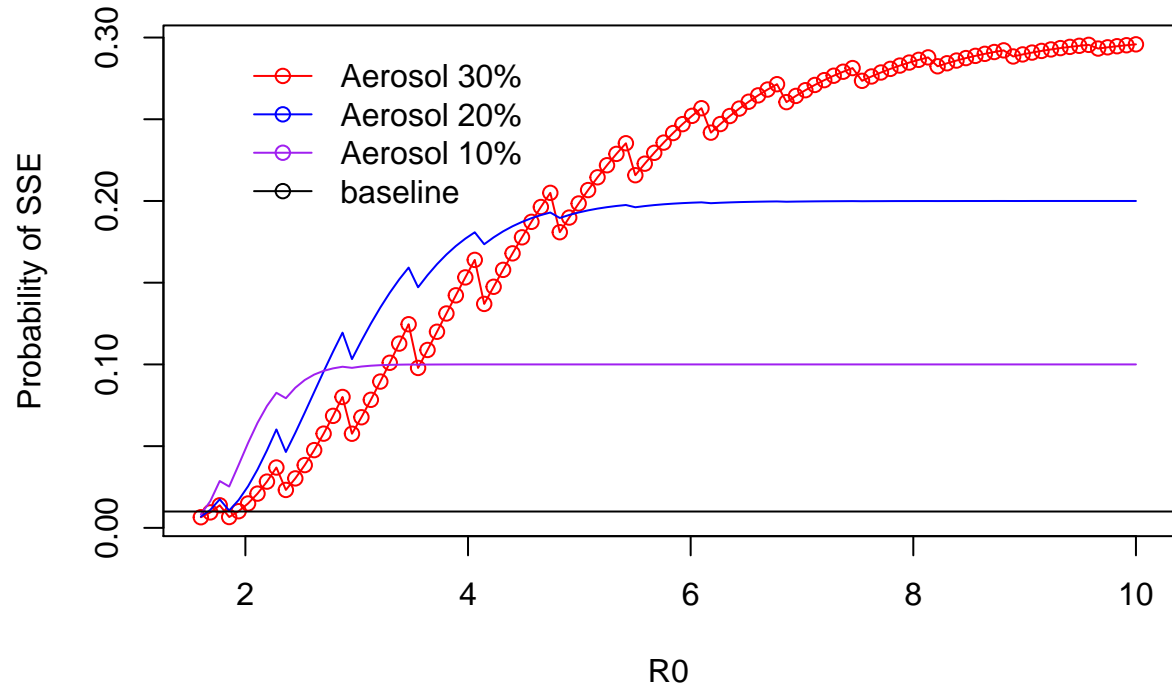
Figure. The probability of a superspreading event (the probability that the number of cases generated by a given Poisson mixture offspring distribution is greater than the 99th percentile of the baseline Poisson model with the same R_0) is an increasing function of R_0 and saturates to $1 - p$ for large R_0.
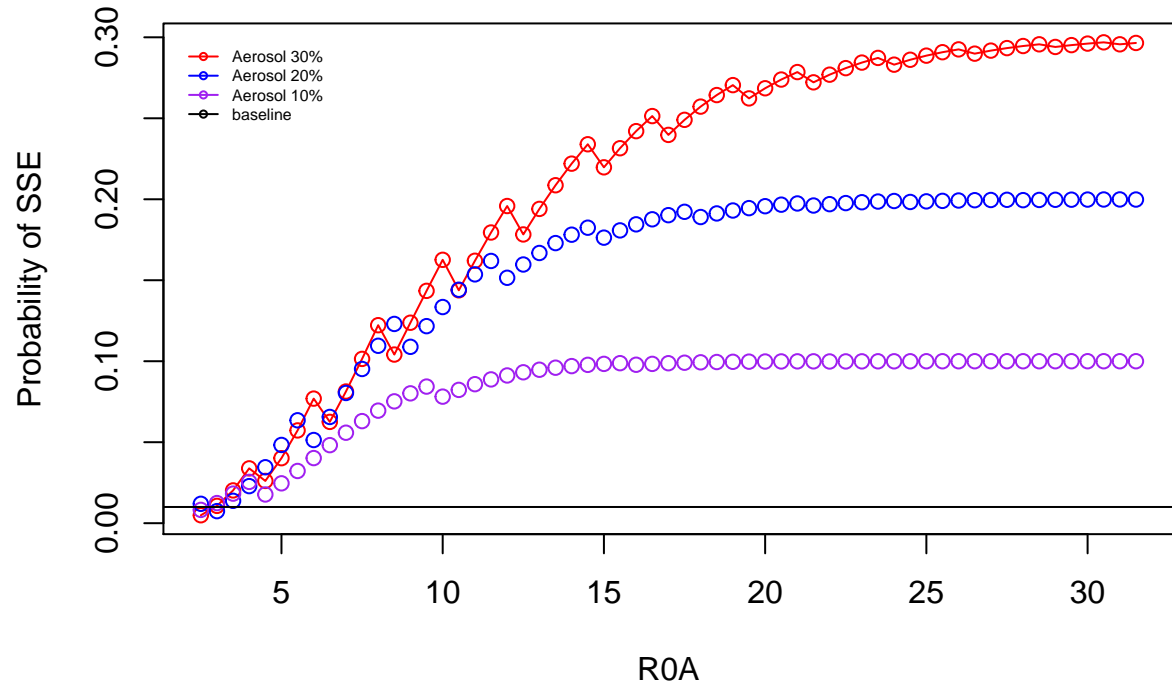
Figure. The probability of a superspreading event is an increasing function of aerosol reproduction number R_0^A and saturates to $1 - p$ for large R_0^A.
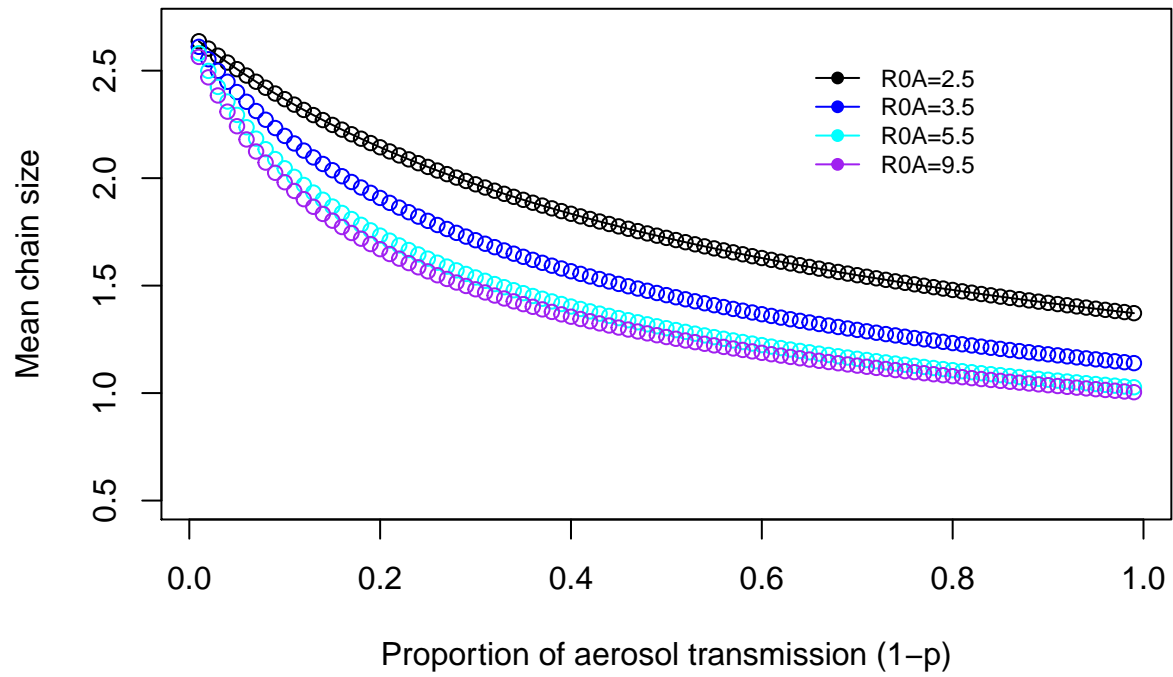
Chain size distribution statistics

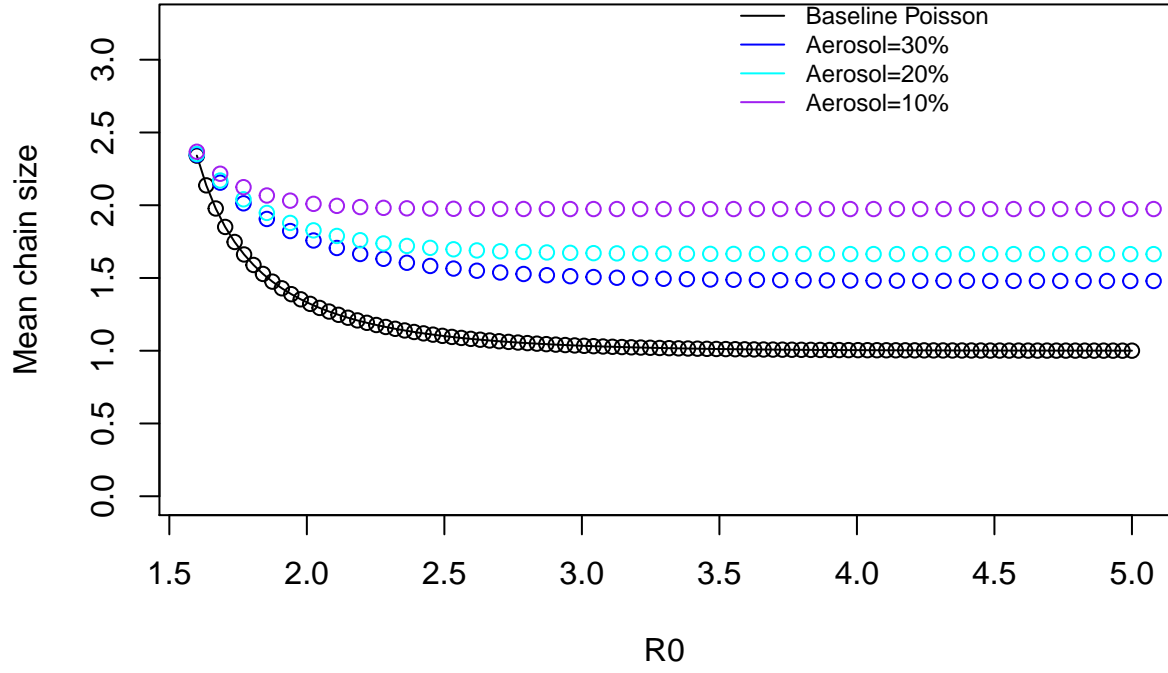**Figure.** The mean chain size decreases with increasing proportion of aerosol transmission.

Figure. The mean chain size decreases with increasing R_0. This is because the probability of a major epidemic increases with R_0 (less likely to have a transmission chain that does not result in a major epidemic).
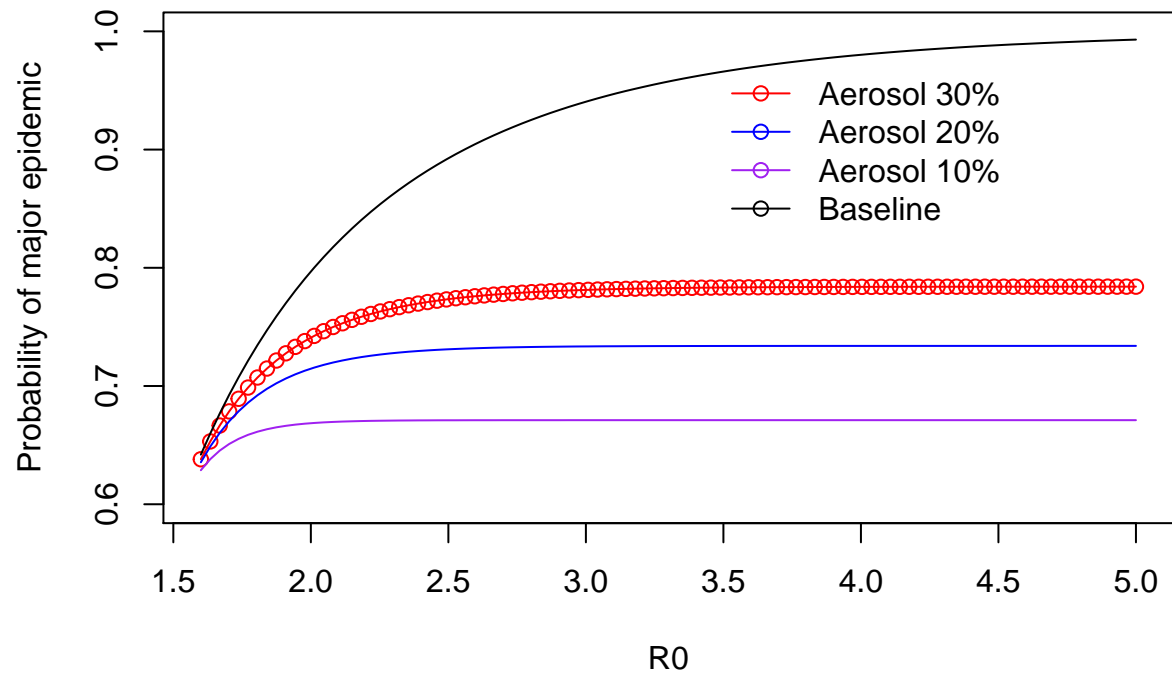
Figure. The probability of a major epidemic increases with R_0 (less likely to have a transmission chain that does not result in a major epidemic). T
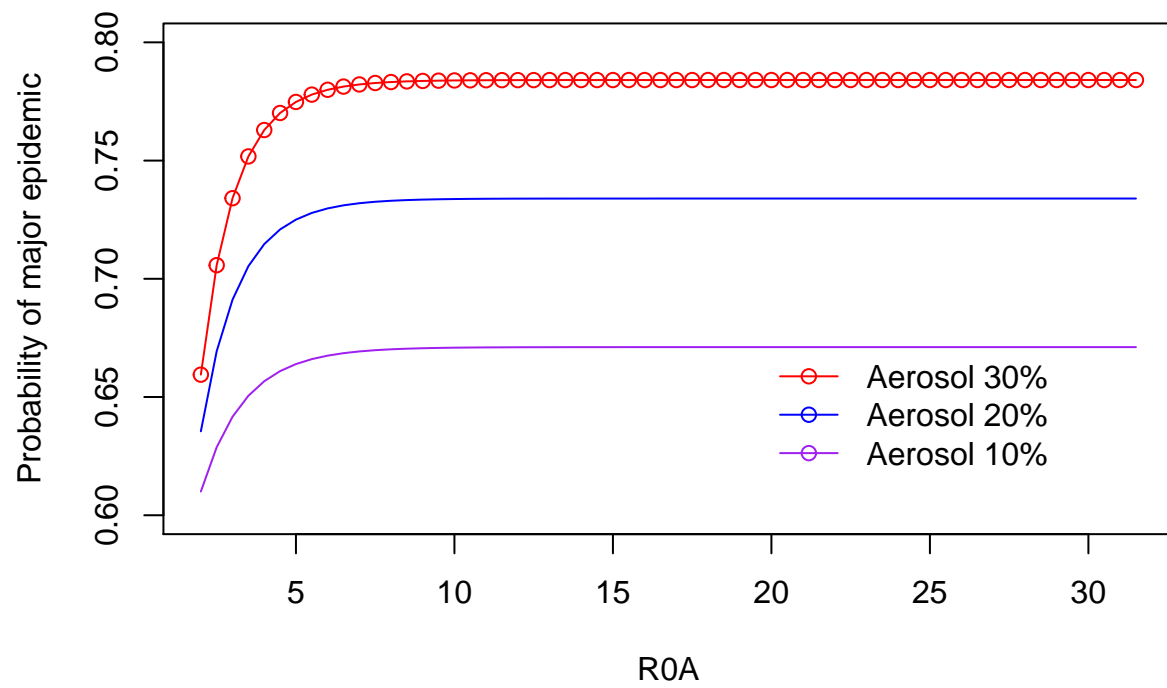
Figure. The probability of a major epidemic increases with R_0^A (less likely to have a transmission chain that does not result in a major epidemic).
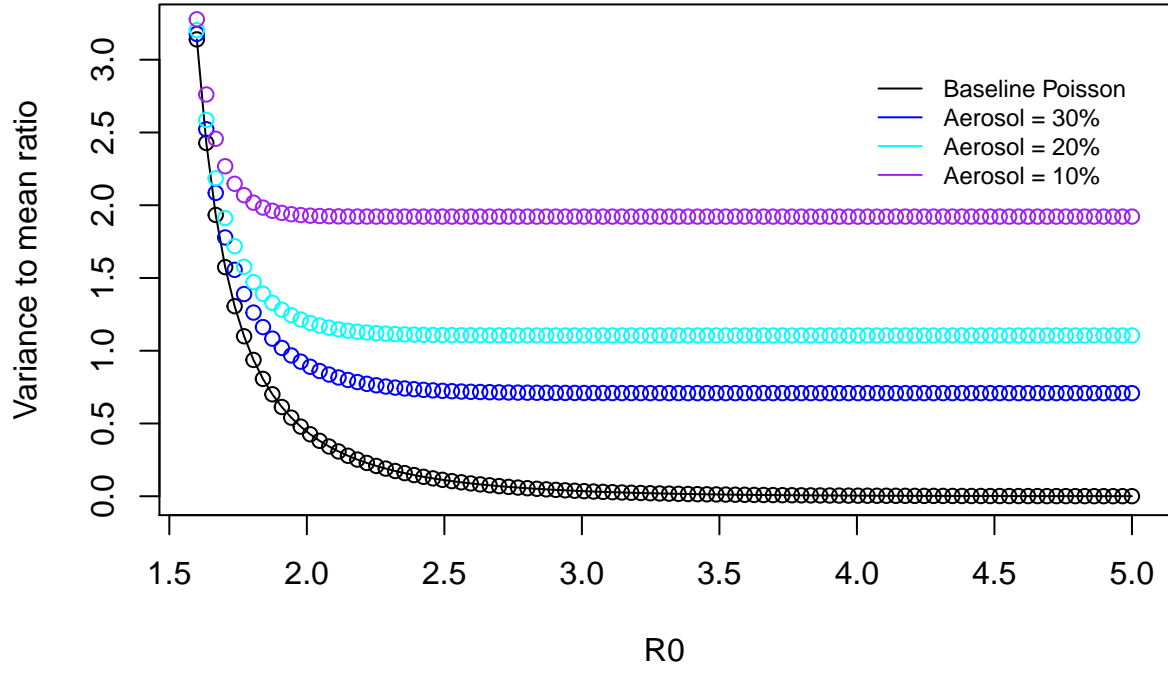
Figure. The variance to mean ratio of the chain size declines as a function of R_0. The level of heterogeneity (overdispersion) decreases as R_0 increases.
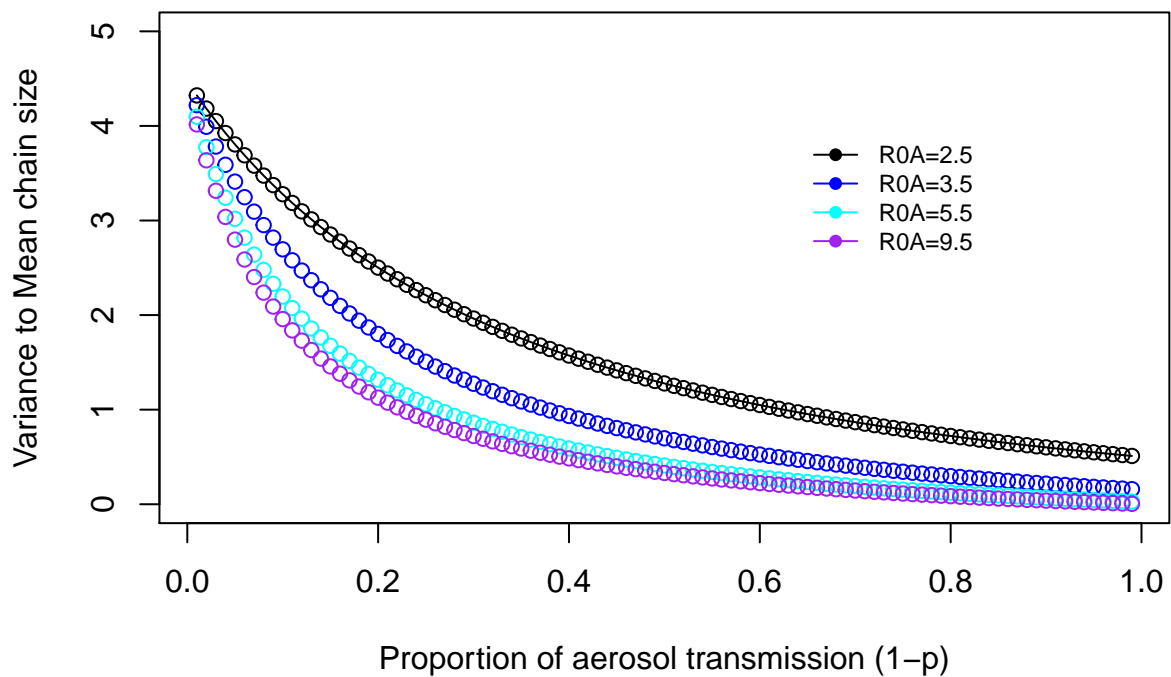
Figure. The variance to mean ratio of the chain size declines as a function of the level of aerosol transmission because the probability of a major epidemic increases as a function of the level of aerosol transmission.
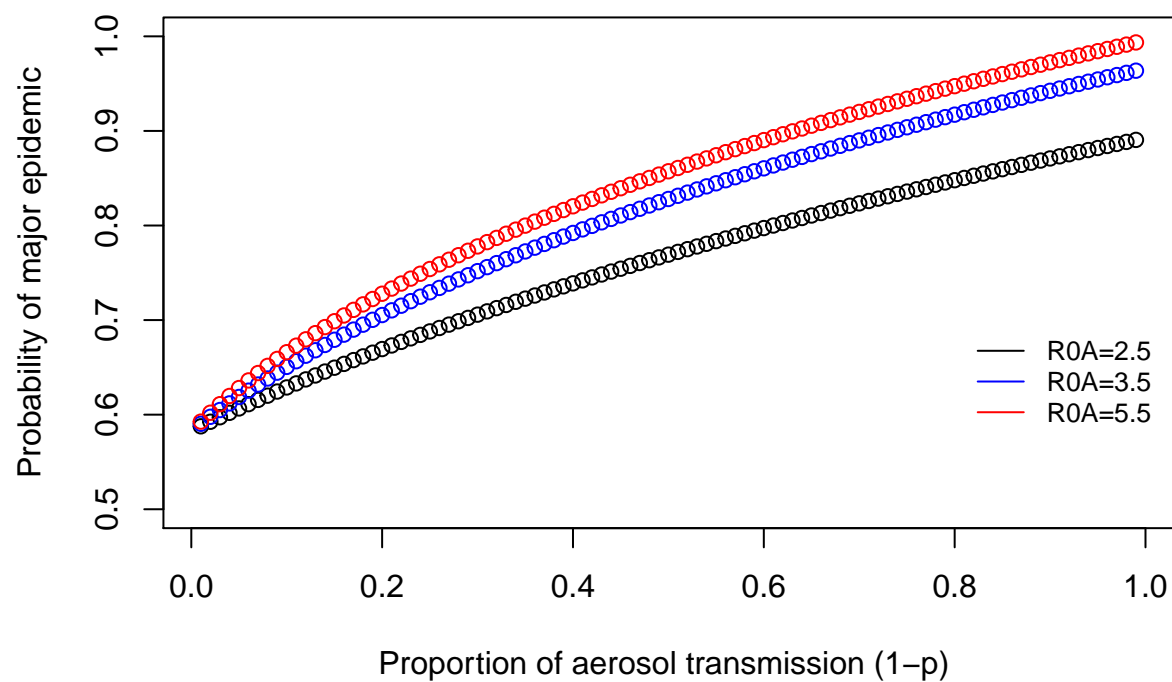
Figure. The probability of a major epidemic increases as a function of the level of aerosol transmission.

Derivation of chain size distribution
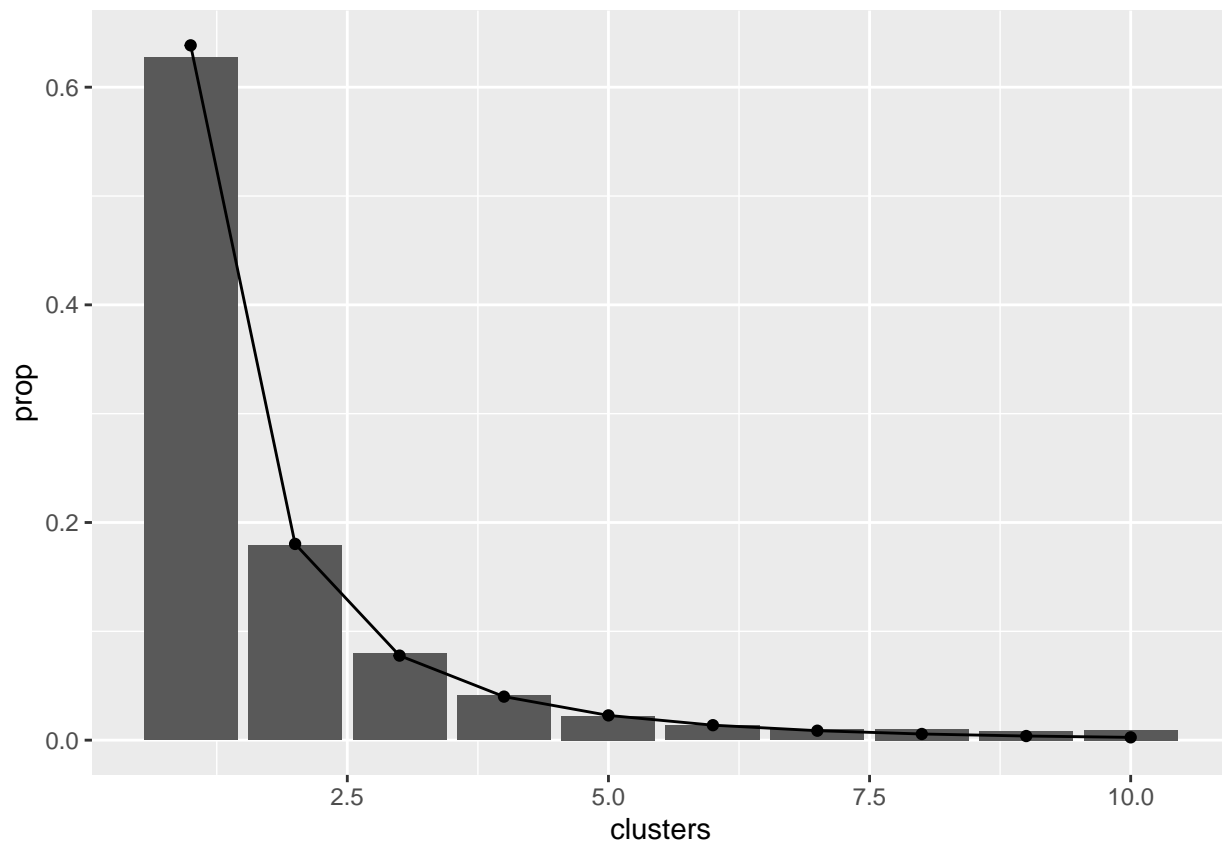
```
## [1] 28604        1
```

Figure. Cluster size distribution obtained from simulation (gray bars) compared to the theoretical prediction (points). We generated 100000 simulations of a mixture branching process with $p = 0.8$, $R_0^D = 1.5$ and $R_0^A = 4$, retaining those that went extinct within 6 generations ($n = 28625$). Points are the theoretical predictions for the probability of observing a cluster of size $y = 1, 2, \ldots, 10$ generated using equations (**??**) and (**??**) conditioned on the probability of extinction, $z_\infty = 0.285$.}
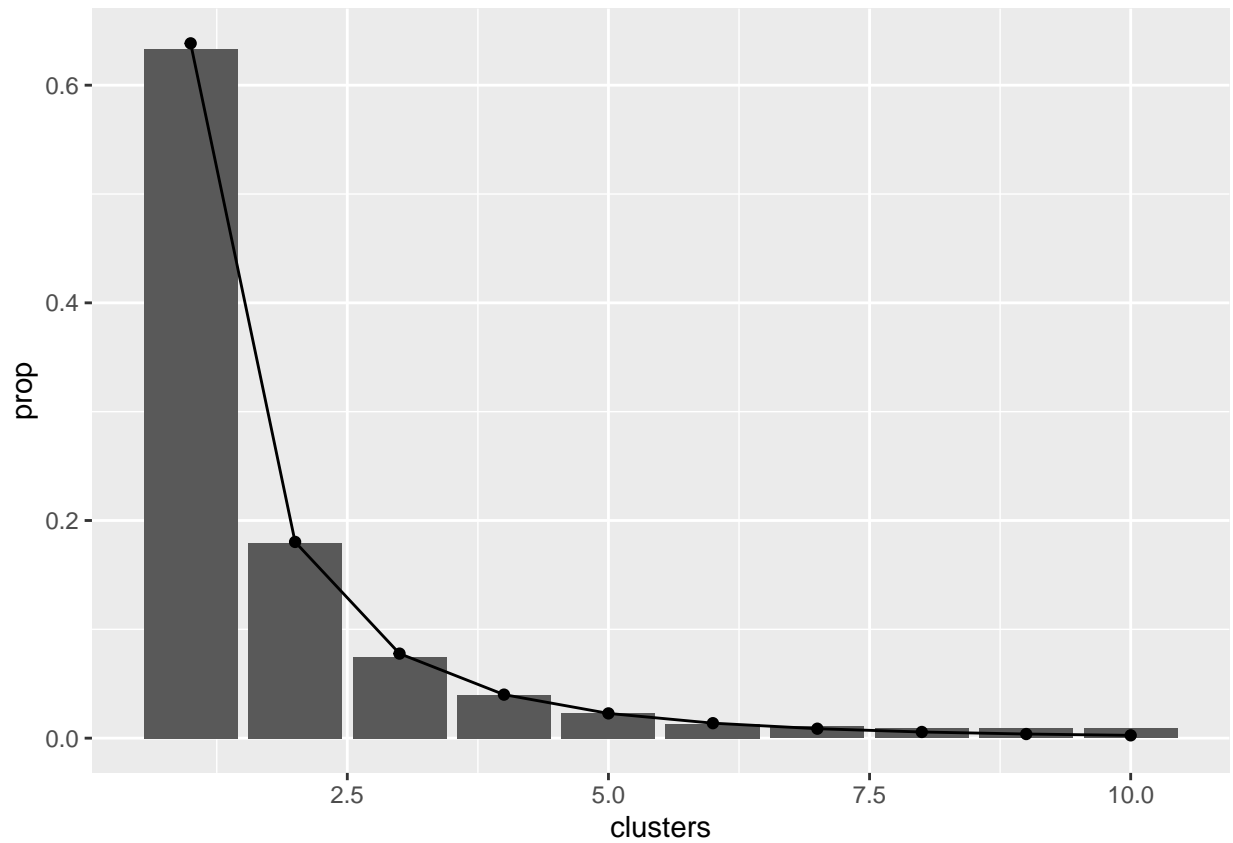
```
## [1] 28596       1
```

Figure. Cluster size distribution obtained from simulation (gray bars) compared to the theoretical prediction (points) using the actuar package.
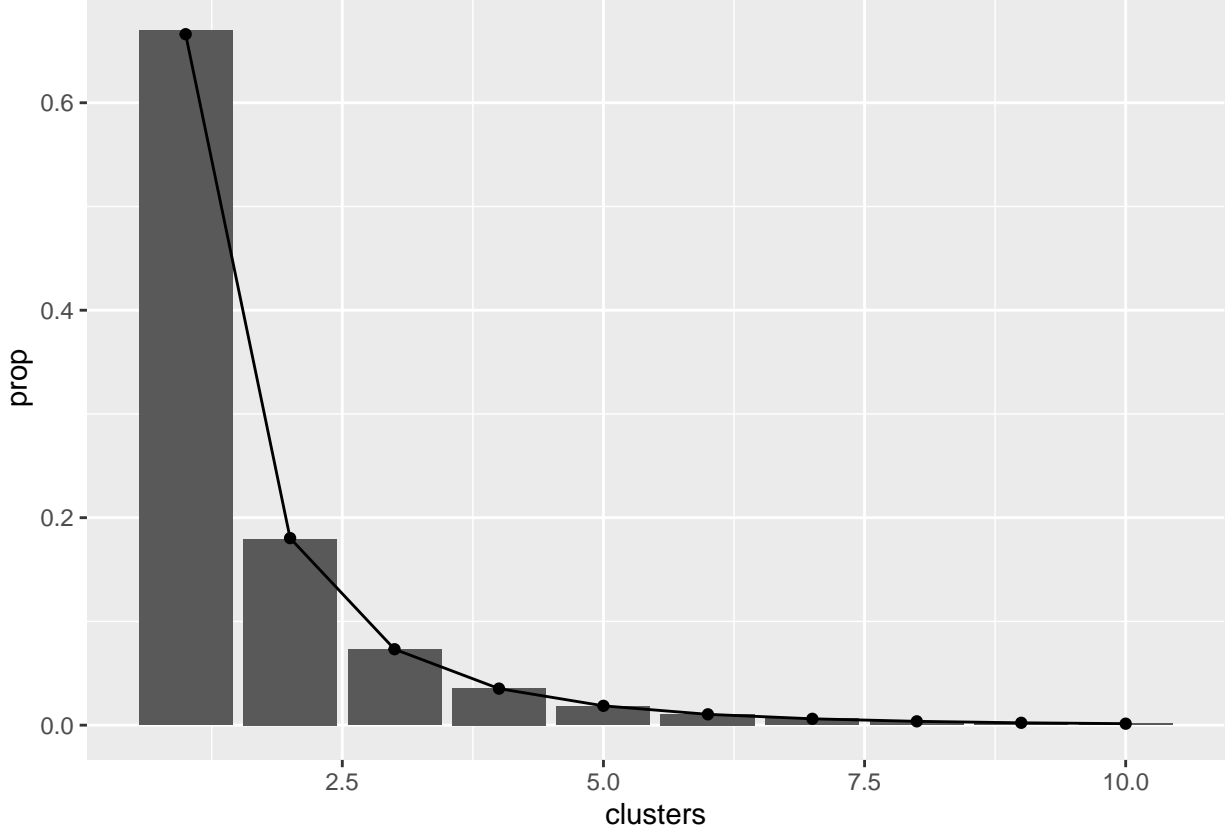
Figure. Cluster size distribution obtained from simulation (gray bars) compared to the theoretical prediction of a Borel distribution (points). We generated 100000 simulations of a Poisson branching process with $R_0 = 2$ retaining those that went extinct within 10 generations ($n = 20155$). Points are the theoretical predictions for the probability of observing a cluster of size $y = 1, 2, \ldots, 10$ generated using equations (**??**) and (**??**) conditioned on the probability of extinction, $z_\infty = 0.203$.

- We will compare the chain size distribution for the Poisson mixture with the chain size distribution obtained from a Poisson offspring distribution. Chain sizes arising from a Poisson branching process follow the Borel-Tanner distribution (Mode and Sleeman 2000) conditioned on a minor outbreak (Yan 2008),

$$P(Y = y) = \frac{1}{z_\infty} \frac{e^{-R_0 y}(R_0 y)^{y-1}}{y!} \tag{5}$$

where z_∞ is the probability of extinction of a Poisson branching process, which is not equal to one when $R_0 > 1$.

- To derive the chain size distribution for the Poisson mixture, we use the result from Blumberg and Lloyd-Smith (2014) and therefore require the derivatives of powers of the generating function (**??**). Let

$$T_y(z) = \frac{1}{y}(Q_Y(z))^y, \quad y = 1, 2, \ldots$$

Then the probability of a chain having size y is

$$P(Y = y) = \frac{1}{(y-1)!} T_y^{(y-1)}\Big|_{z=0} \tag{6}$$

- To evaluate the derivatives of

$$(Q_Y(z))^y = (pe^{R_0^D(z-1)} + (1-p)e^{R_0^A(z-1)})^y, \tag{7}$$

21

i.e., the yth power of the generating function (**??**), we need to apply the chain rule $y - 1$ times. The kth derivative of the inner function of equation (7) evaluated at $z = 0$ is

$$g_k = p(R_0^D)^k e^{-R_0^D} + (1-p)(R_0^A)^k e^{-R_0^A}, \quad k = 1, 2, \ldots, y-1. \tag{8}$$

The kth derivative of the outer function of equation (7) evaluated at $z = 0$ is

$$f_k = \frac{y!}{(y-k)!} \left(pe^{-R_0^D} + (1-p)e^{-R_0^A} \right)^{y-k}, \quad k = 1, 2, \ldots, y-1. \tag{9}$$

The generalized chain rule or Faa di Bruno's formula (citation) yields

$$T_y^{(y-1)}\Big|_{z=0} = \sum_{k=1}^{y-1} f_k B_{y,k}(g_1, g_2, \ldots, g_{y-1-k}) \tag{10}$$

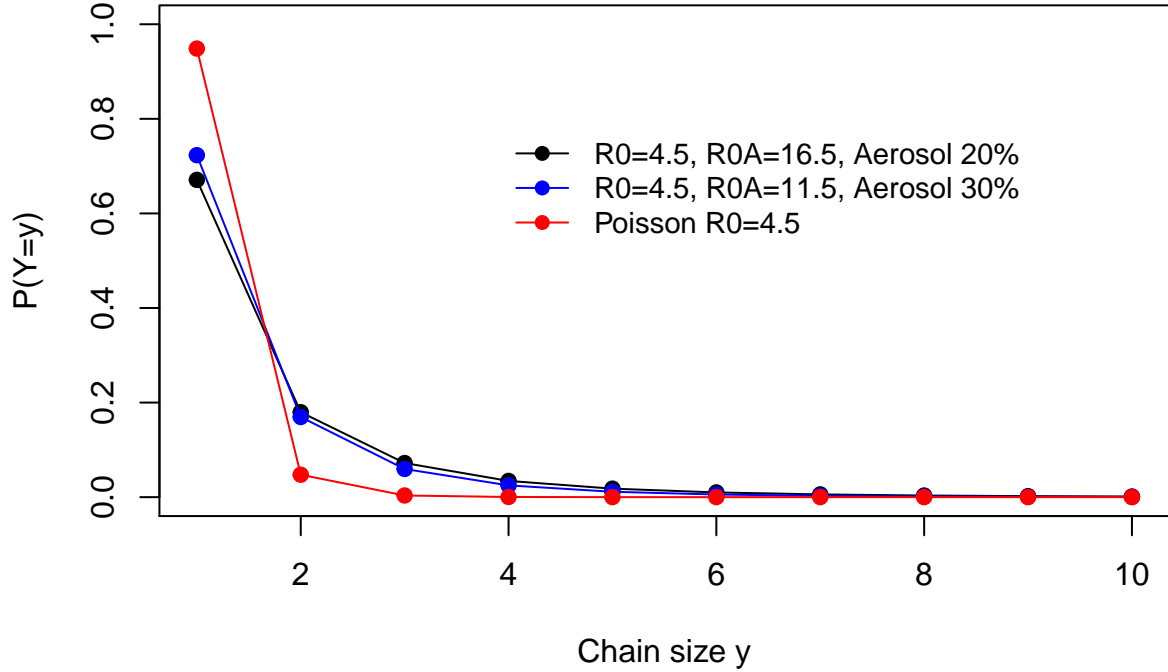where $B_{y,k}$ are Bell polynomials of the g_k.

Figure. Chain size distributions conditional on extinction for a Poisson mixture are more fat-tailed than the corresponding Poisson distribution with the same mean ($R_0 = 4.5$). Here $R_0^D = 1.5$.
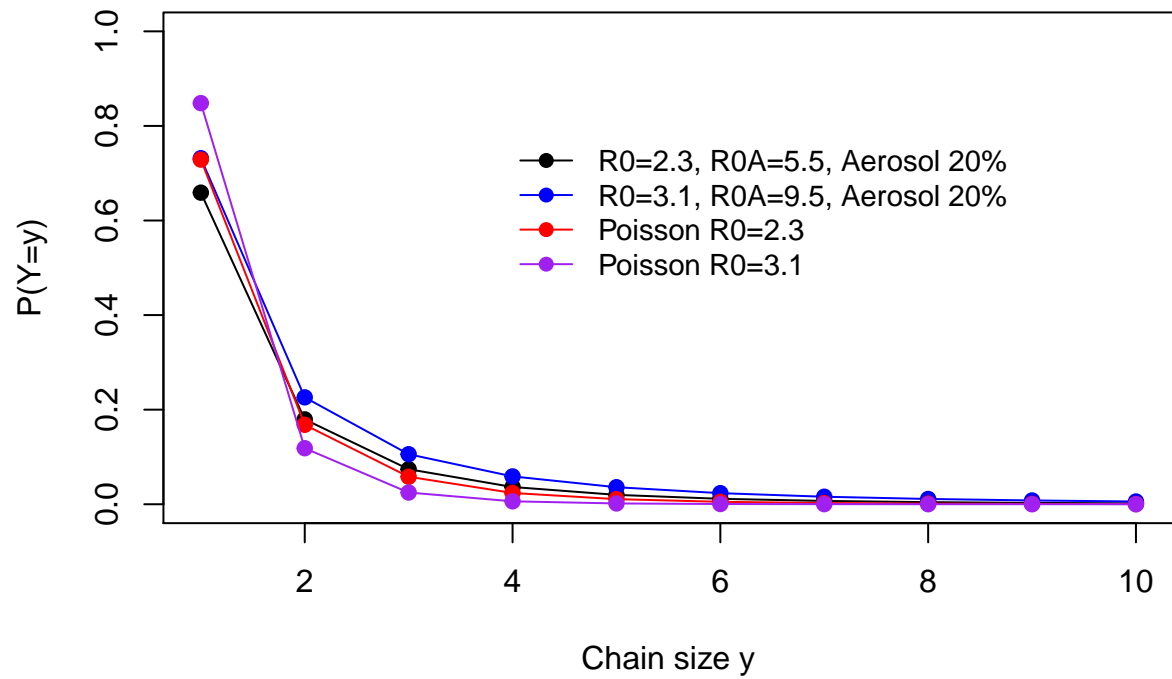
Figure. Chain size distributions conditional on extinction for a Poisson mixture vary little with increasing δ. Here $R_0^D = 1.5$.
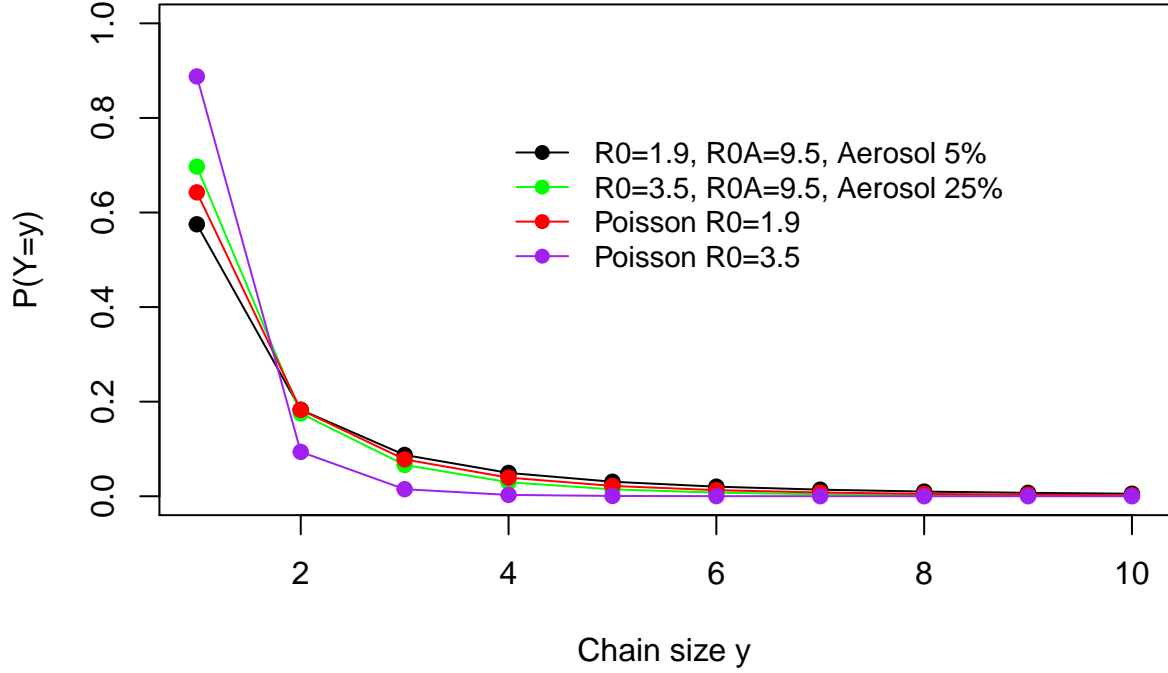
Figure. The burstiness of the Poisson mixture offspring distributions do not manifest in the corresponding chain size distributions conditioned on extinction. The 5 percent aerosol model has a fatter tail than the 25 percent aerosol model, but R_0 is smaller. As R_0 decreases, minor outbreaks occur with greater probability in the mixture models. The Poisson model with $R_0 = 1.9$ looks fairly similar to the mixture model with $R_0 = 1.9$. As R_0 increases the difference between the mixture chain size distribution and the Poisson distribution becomes larger. Note that the distribution is conditional on extinction, i.e., we have divided it by the probability of extinction. The analytical expression for the chain size distribution will not capture burstiness. Major outbreaks occur with probability $1 - z_\infty$. To compare the analytical chain size distribution with a distribution obtained via simulation, we must not divide it by z_∞.
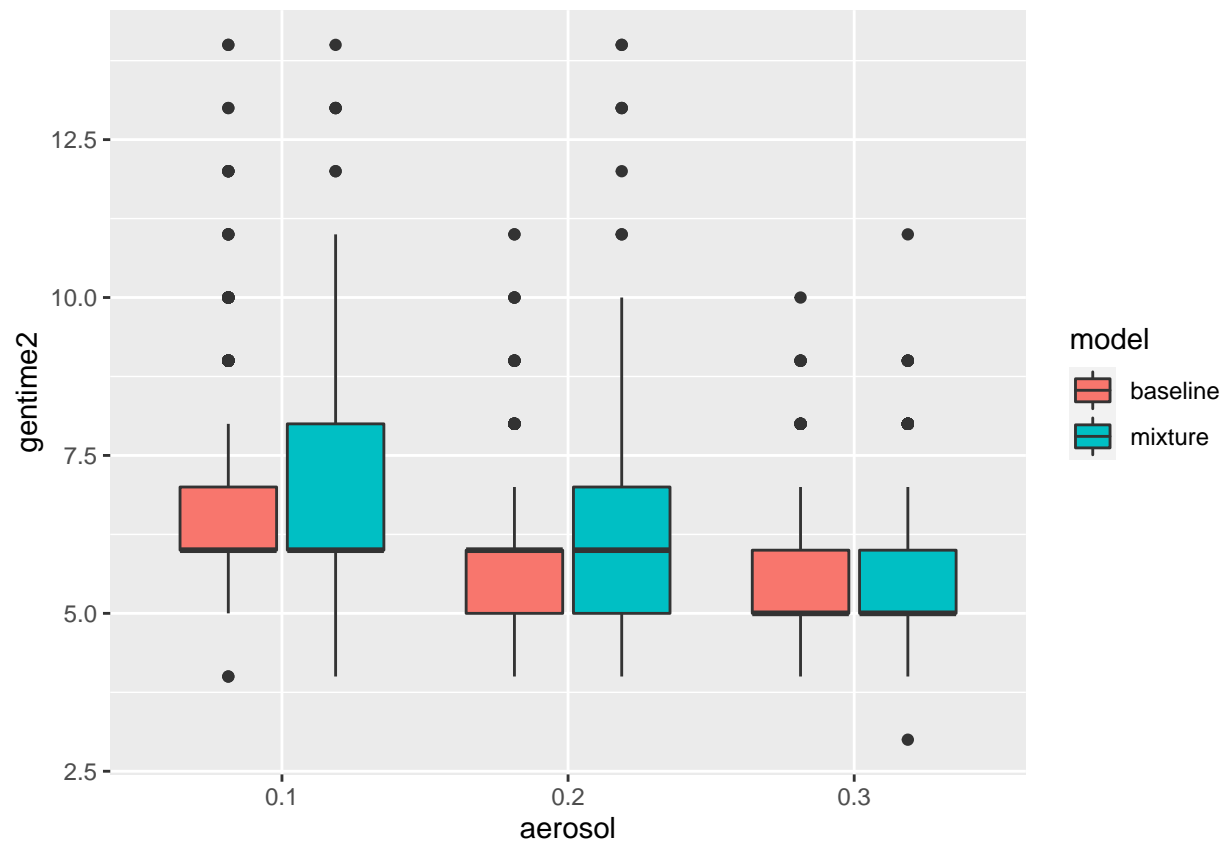
Simulating branching processes

Figure. As $1 - p$ increases, so does R_0. The Poisson mixture generally produces epidemics that take off more slowly than the equivalent Poisson model.
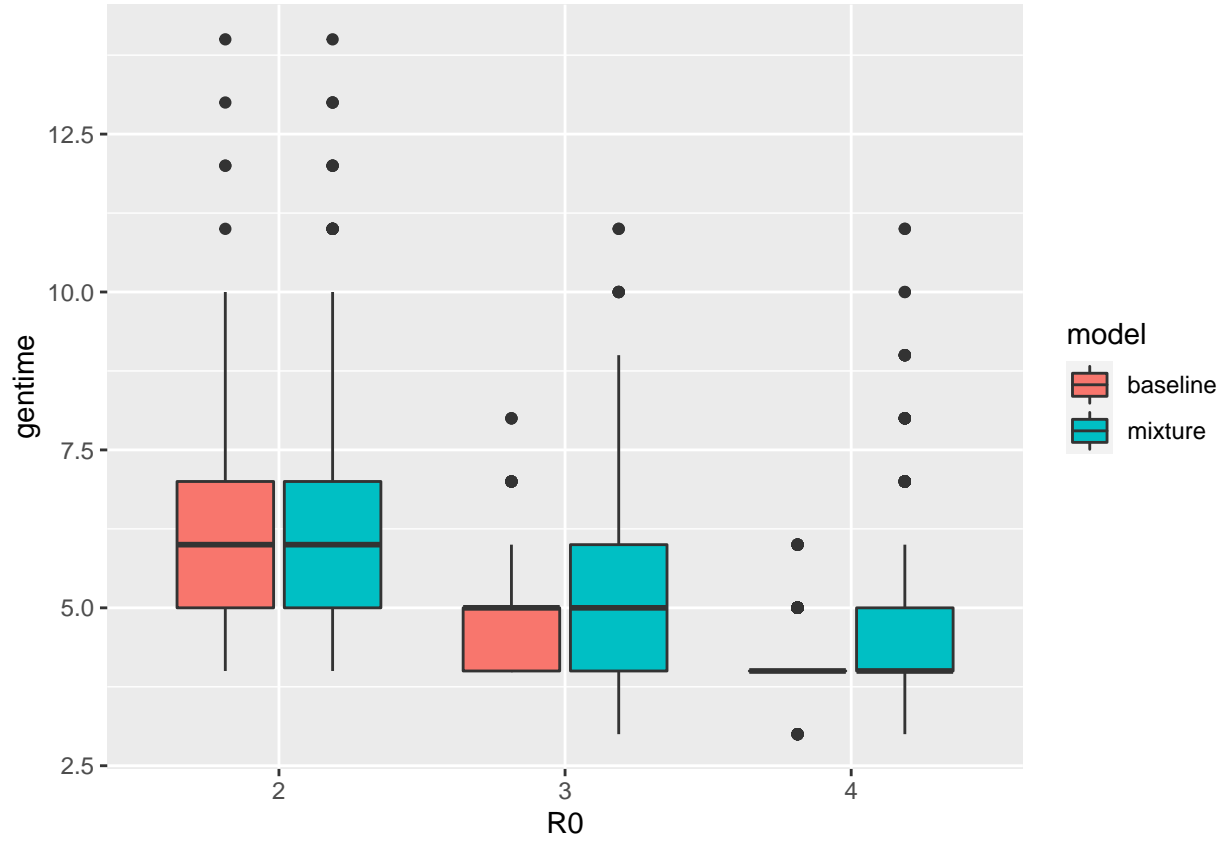
Figure. As R_0 increases (resulting from increasing R_0^A), the median first generation to have 50 cases declines for the baseline and mixture model, indicating outbreaks become more explosive. The generation time is much more variable for the mixture models than the baseline models. In a given epidemic, direct contact effects may dominate (leading to slower epidemics) or aerosol transmission effects may dominate (leading to faster epidemics). For example if $R_0 = 4$ about 21% of epidemics reach 50 cases within three generations but additionally about 21% of epidemics take 6 generations or more to reach 50 cases.
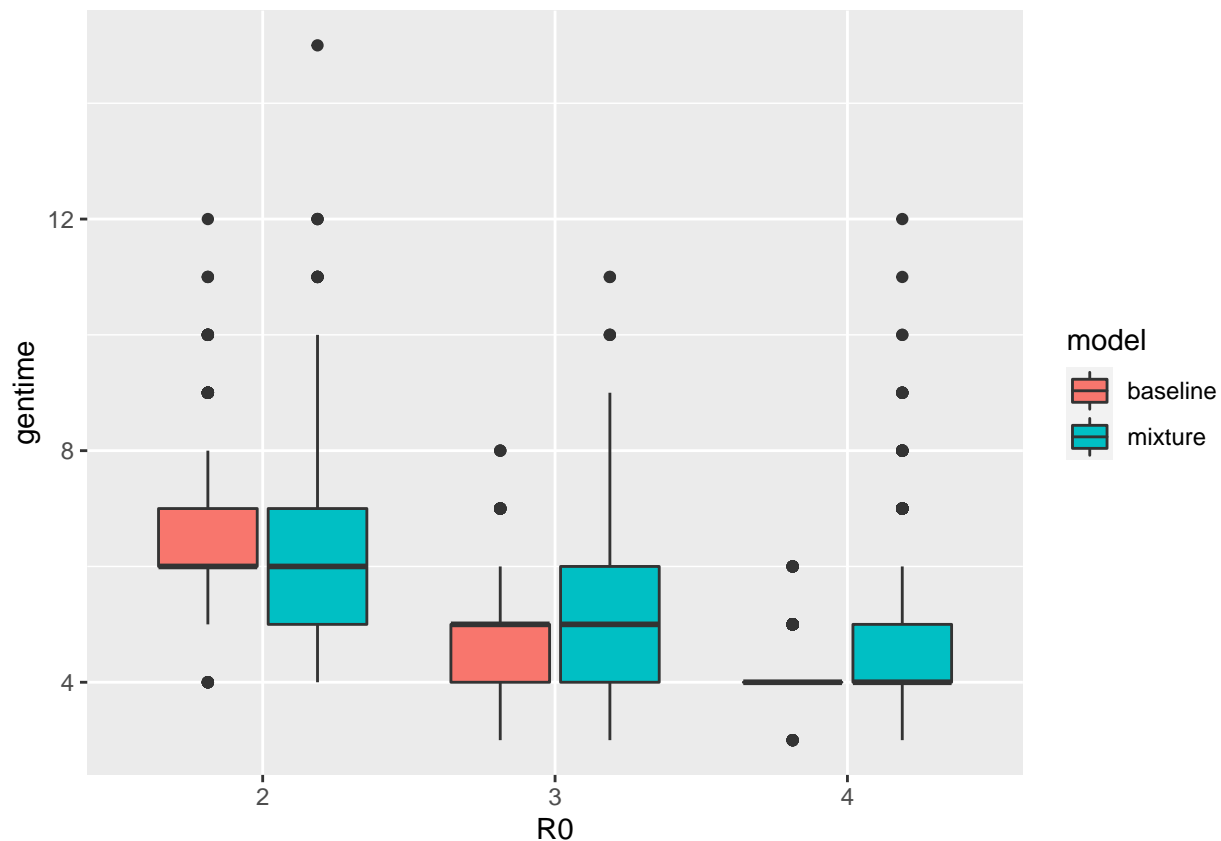
Figure. As R_0 increases (resulting from increasing R_0^A), the median first generation to have 50 cases declines for the baseline and mixture model, indicating outbreaks become more explosive. The generation time is much more variable for the mixture models than the baseline models. In a given epidemic, direct contact effects may dominate (leading to slower epidemics) or aerosol transmission effects may dominate (leading to faster epidemics). For example if $R_0 = 4$ about 21% of epidemics reach 50 cases within three generations but additionally about 21% of epidemics take 6 generations or more to reach 50 cases. This figure was generated using the actuar package.