Negative binomial mixture branching process model of transmission

Key questions

- Does the mechanistic addition of population structure induce qualitatively different outbreak patterns from a standard superspreading model?
- How does decreasing the level of superspreading by a) changing the population structure e.g., by shifting the contact structure away from opportunistic encounters/aerosol transmission and towards regular contacts/direct contact transmission, and b) decreasing the average number of successful contacts in the superspreading cohort affect heterogeneity in outbreak patterns, and what are the implications for containment?

Model Assumptions

We assume that infected individuals can be divided into two disjoint groups - a fraction p that contribute to transmission via superspreading, and the remaining fraction of the population $1-p$ that that do not contribute to superspreading transmission. In the superspreading cohort, the mean cumulative number of contacts leading to transmission of infection per infected individual per unit time is high at $\beta_1 = \beta_D + \delta$, whereas in the non-superspreading group, it is low $\beta_2 = \beta_D < \beta_1$. In both groups the contact process follows a Poisson distribution with mean β_i $i, = 1, 2$. Then the contact process for the entire population is a finite Poisson mixture with random variates,

$$\text{number of cumulative contacts per infectious individual per unit time} \sim p\text{Poisson}(\beta_1) + (1-p)\text{Poisson}(\beta_2). \tag{1}$$

In both groups,we assume the infectious period is gamma distributed with mean $1/\gamma$ and coefficient of variation $1/\sqrt{k}$ with probability density function

$$f(x) = \frac{(\gamma k)^k}{\Gamma(k)} x^{k-1} e^{-k\gamma x}$$

The gamma distribution is flexible in that allows for right-skewed distributions (i.e., $k < 1$) and distributions with a central tendency ($k > 1$), with $k = 1$ leading to the exponential distribution. The probability generating function for the mixture follows

$$\begin{aligned} h(s) &= \int_0^\infty \left(p e^{\beta_1 x(s-1)} + (1-p) e^{\beta_2 x(s-1)} \right) \frac{(\gamma k)^k}{\Gamma(k)} x^{k-1} e^{-k\gamma x} dx \\ &= p \frac{(\gamma k)^k}{(\gamma k + \beta_1(1-s))^k} + (1-p) \frac{(\gamma k)^k}{(\gamma k + \beta_2(1-s))^k} \\ &= p \left(\frac{1}{(1 + \frac{\beta_1}{\gamma k}(1-s))} \right)^k + (1-p) \left(\frac{1}{(1 + \frac{\beta_2}{\gamma k}(1-s))} \right)^k \\ &= \frac{p}{(1 + \frac{R_0^A}{k}(1-s))^k} + \frac{(1-p)}{(1 + \frac{R_0^D}{k}(1-s))^k}, \end{aligned} \tag{2}$$

where $R_0^D = \beta_2/\gamma = \beta_D/\gamma$ and $R_0^A = \beta_1/\gamma = (\beta_D + \delta)/\gamma$. Therefore equation (2) shows that a finite mixture of negative binomial distributions models a combination of close contact transmission and superspreading. The number of secondary infections per generation is obtained from

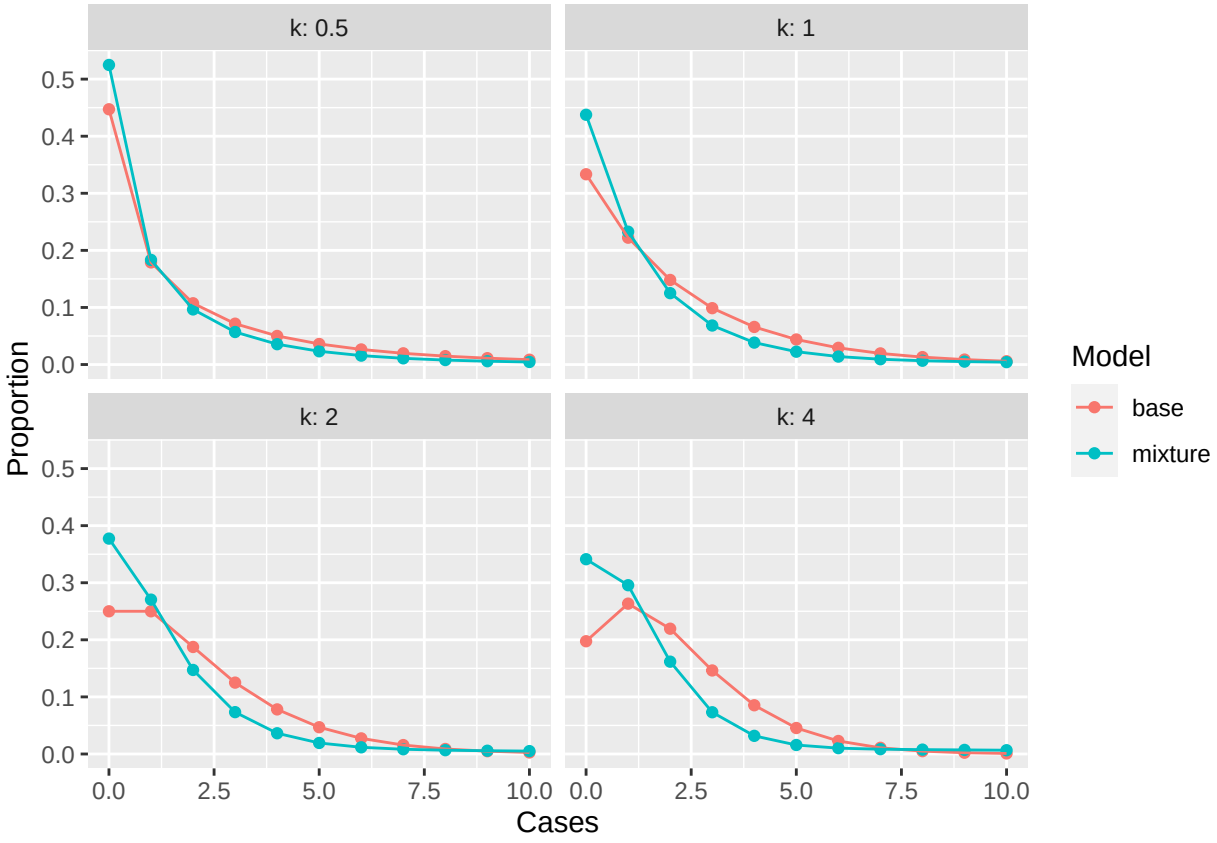
$$\text{number of secondary infections} \sim p\text{Negative Binomial}(R_0^A, k) + (1-p)\text{Negative Binomial}(R_0^D, k). \quad (3)$$

The mean number of secondary infections is $R_0 = pR_0^A + (1-p)R_0^D = R_0^D + p\delta$.

We compare the mixture model with a baseline negative binomial model with the same R_0 and dispersion parameter k.

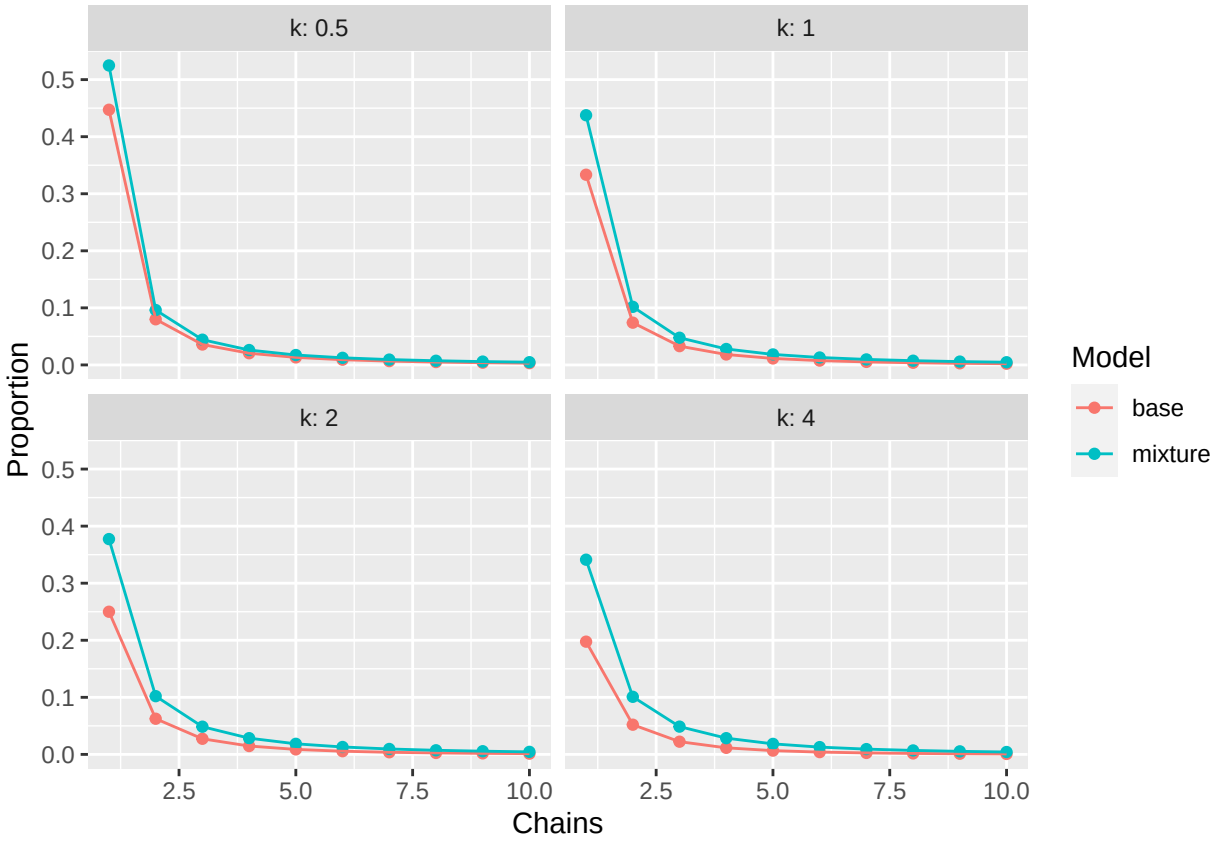
Probability mass functions for baseline and mixture models

Here we compare the probability mass functions of the mixture model ($R_0^D = 1.1$, $p = 0.1$, additional contacts $\delta = 9$) with the base model for various values of k. The mean number of secondary infections for both models is $R_0 = 2$. For the mixture models, the probability of no secondary infections is always greater than the negative binomial model with the same R_0 and k. As k increases there is a greater central tendency in the number of secondary infections in the base model.



Corresponding chain size distributions

Here we compare the chain size distributions of the mixture model ($R_0^D = 1.1$, $p = 0.1$, additional contacts $\delta = 9$) with the base model for various values of k. The mean number of secondary infections for both models is $R_0 = 2$. The chain size distribution is fatter tailed for the mixture models compared to the corresponding base models.



Statistics that show hallmarks of transmission heterogeneity

Hallmarks of heterogeneous transmission include:

- Greater variability in the number of secondary infections (fat tailed)
- Smaller probability of major epidemics
- Greater variability in chain sizes
- Larger probability of observing no secondary infections and of observing small chains that go extinct

Here we study the coefficient of variation of the number of secondary infections, the probability of a major outbreak, the probability of observing a small transmission chain of less than or equal to 10 cases, and the mean and coefficient of variation of small chain sizes (conditioned on extinction).

In each of the following, p and δ are varied but $R_0 = R_0^D + p\delta$ is fixed at $R_0 = 2$. The following figures show that smaller values of p (and larger values of δ) lead to more heterogeneous epidemics, even if the dispersion parameter $k > 1$.

Numerical studies (assuming $R_0 > 1$)

To compare output of the standard negative binomial model and finite mixture negative binomial model, we calculated various summary statistics. We set the mean number of secondary infections per individual R_0 to 2 in both models. To explore the impact of variability in infectious period distributions in the standard

and mixture models, the dispersion parameter k was varied between 1/2 and 4. To study the impact of transmission from a combination of regular and superspreading cohorts, we varied p and δ in the mixture model while keeping the basic reproduction number fixed at $R_0 = R_0^D + p\delta = 2$. The superspreading proportion p was varied between 0.01 and 1 and the number of additional contacts $\delta = (R_0 - R_0^D)/p$ was simultaneously adjusted to retain $R_0 = 2$. As p increases, the number of additional contacts declines, i.e., a low proportion of superspreaders need to have a high additional contact rate.

Outbreaks that have hallmarks of superspreading include high variability in the number of secondary infections per infected individual, small probability of major epidemics, high variability in transmission chain sizes, high probability of observing no secondary infections per infected individual and high probability of observing small transmission chains. To compare variability in the number of secondary infections per infected individual in the standard negative binomial model and finite mixture negative binomial model, we computed the coefficient of variation of seceondary infections. To compare variability in transmission chain sizes in the standard negative binomial model and finite mixture negative binomial model, we computed the mean and coefficient of variation of chain sizes. To compare probabilities of observing small transmission chains in the standard negative binomial model and finite mixture negative binomial model, we used the chain size distributions to compute the probability of observing a chain consisting of less than or equal to 10 cases. To compare probabilities of a major epidemic, we used the probability generating functions to compute the probability of extinction numerically.

CV offspring distribution

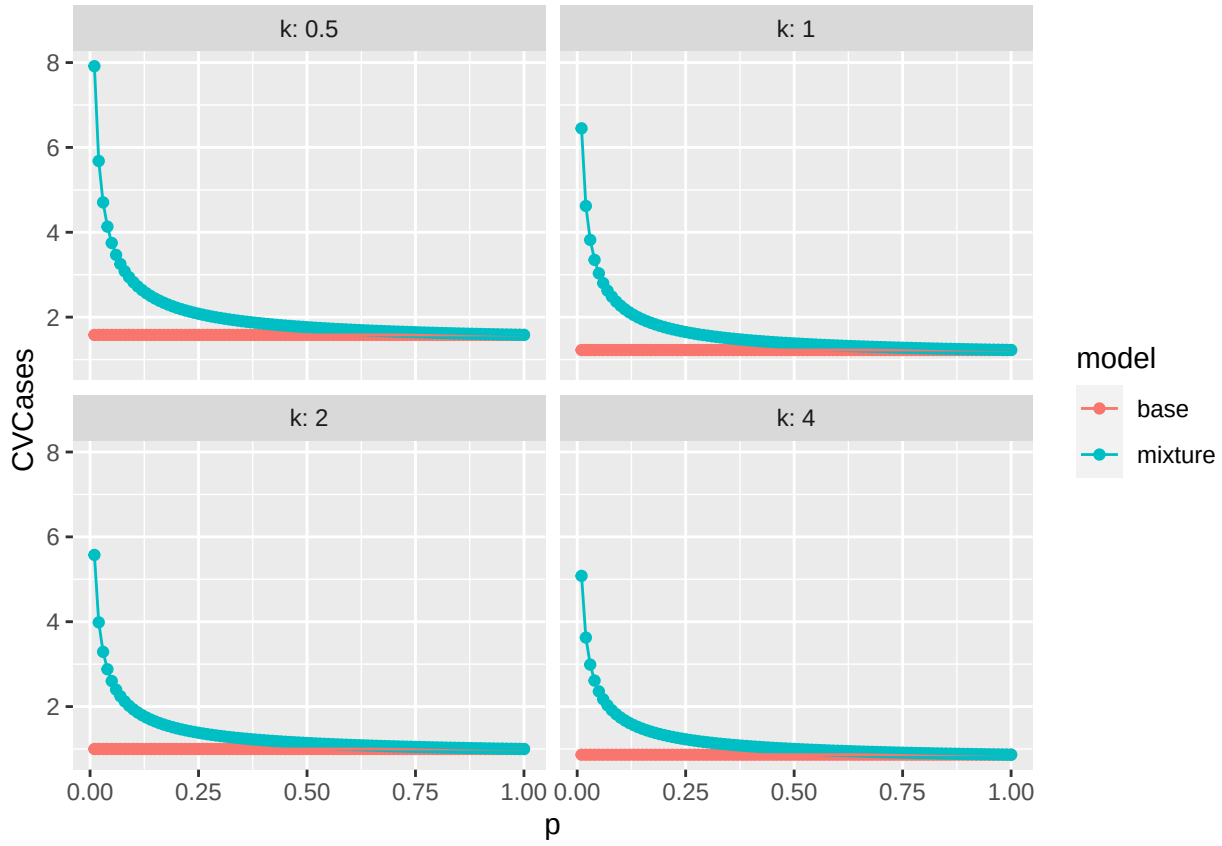


Figure 3. Coefficient of variation of secondary infections in the mixture model is highest for small dispersion parameter k, small p and large number of additional contacts. The coefficient of variation of secondary infections in the mixture model decreases as p increases and approaches the value of the standard model as p approaches 1. There is greater variability in the number of secondary infections in the mixture

model compared to the base model, even if $k > 1$, with the highest variability for small dispersion parameter k, small p and large number of additional contacts.

Probability of major outbreak

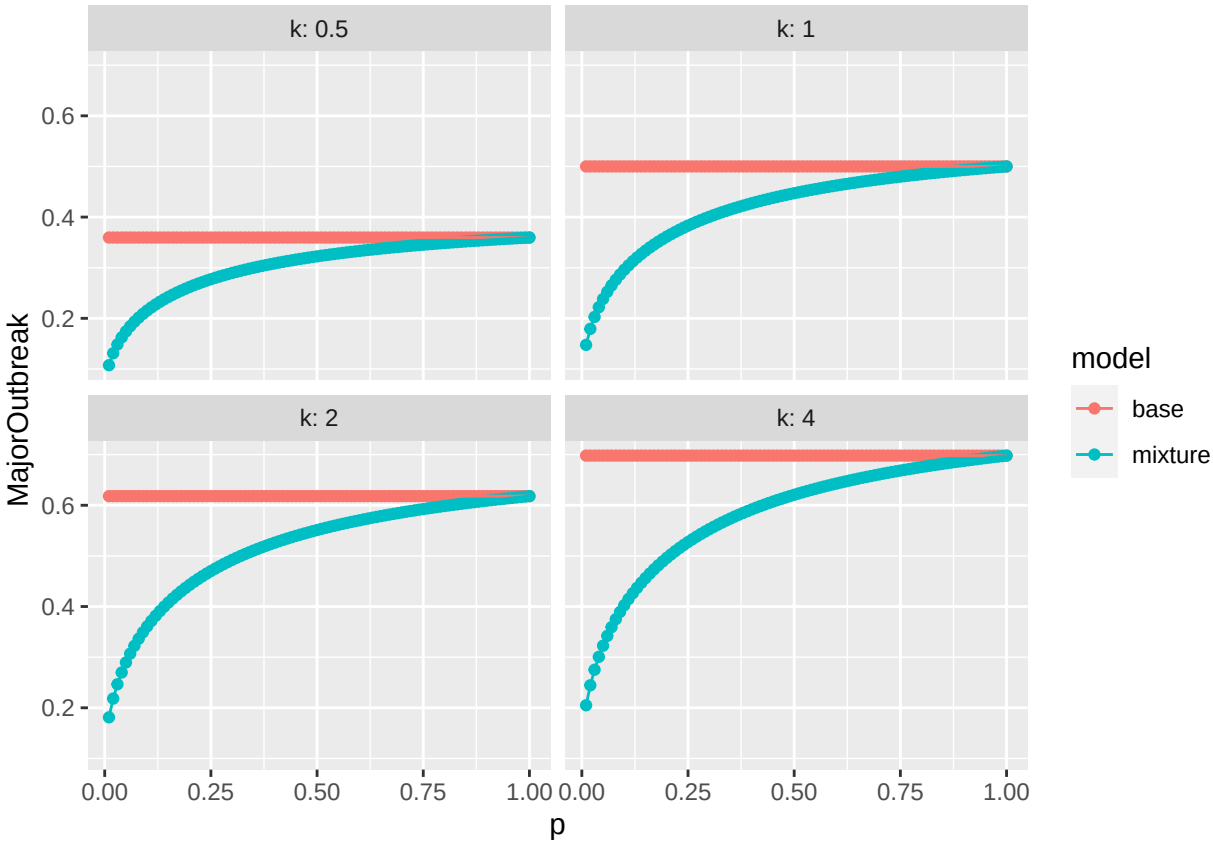


Figure 4. There is smaller probability of major epidemics in the mixture model compared to the base model, even if $k > 1$, with the lowest probabilities for small dispersion parameter k, small p and large number of additional contacts.

Probability of observing a transmission chain of size <= 10

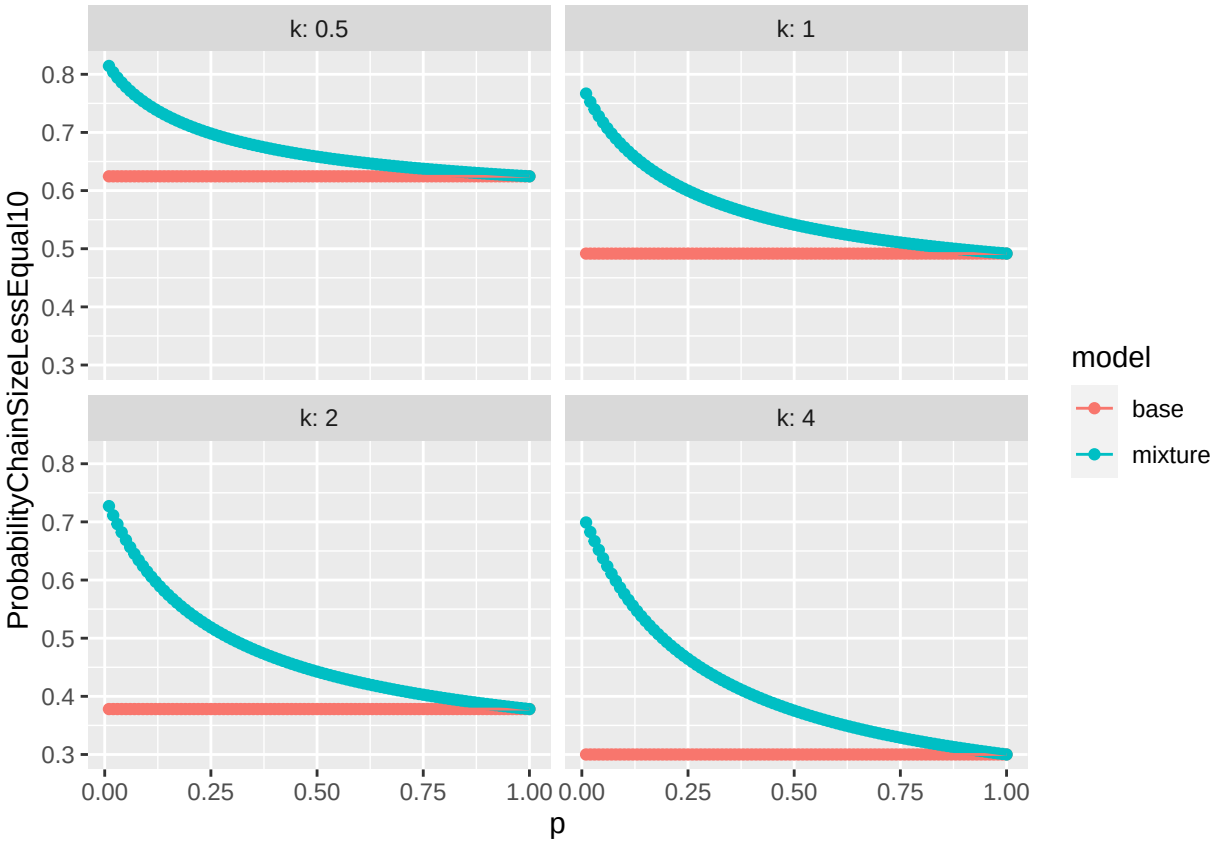


Figure 5. There is larger probability of observing small chains that go extinct in the mixture model compared to the base model, with the highest probabilities for small dispersion parameter k, small p and large number of additional contacts.

Mean chain size

Fatter tail in the chain size distribution conditioned on extinction (i.e. restricting to small transmission chains that go extinct) will make the mean chain size conditioned on extinction larger. Perhaps not a true hallmark of superspreading but it is indicative of fatter tailed chain size distributions.

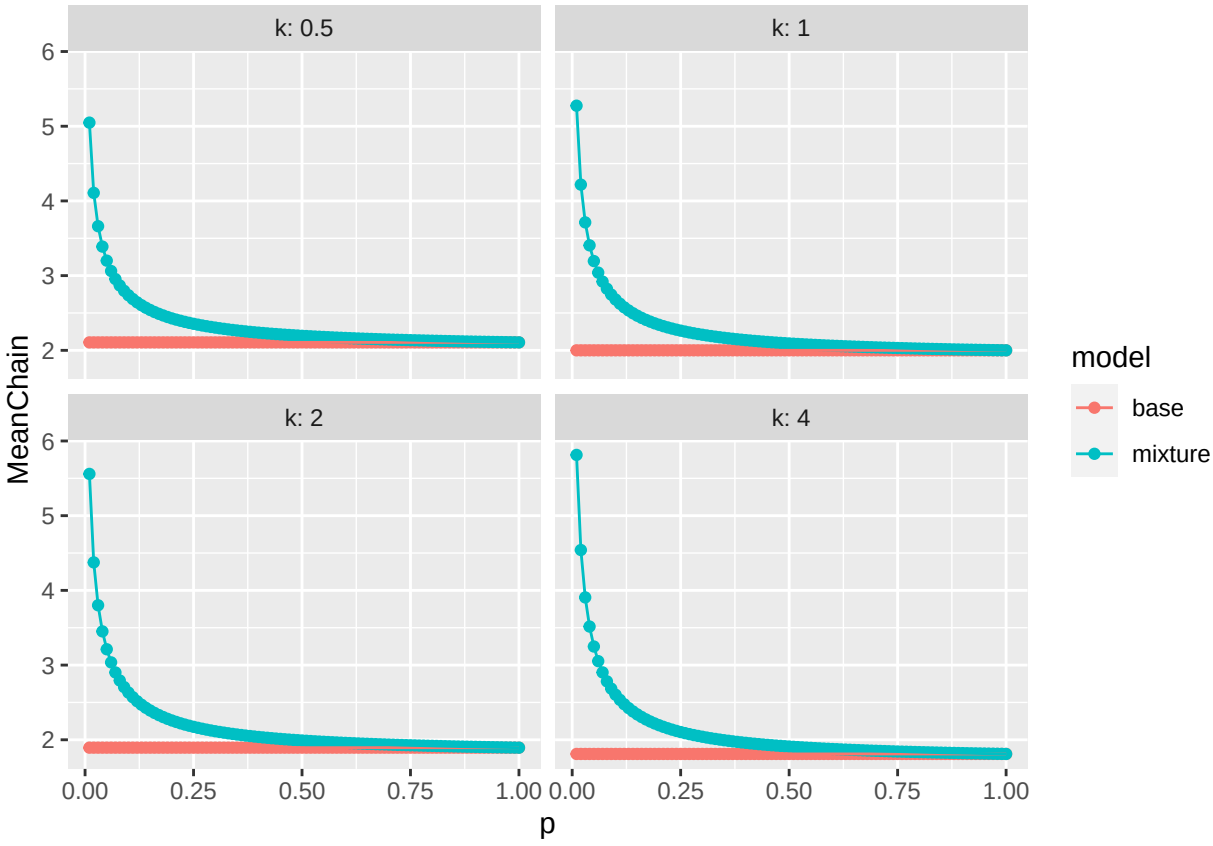


Figure 6. Mean chain sizes are larger in mixture models compared to the base model, even if $k > 1$.

CV chain size

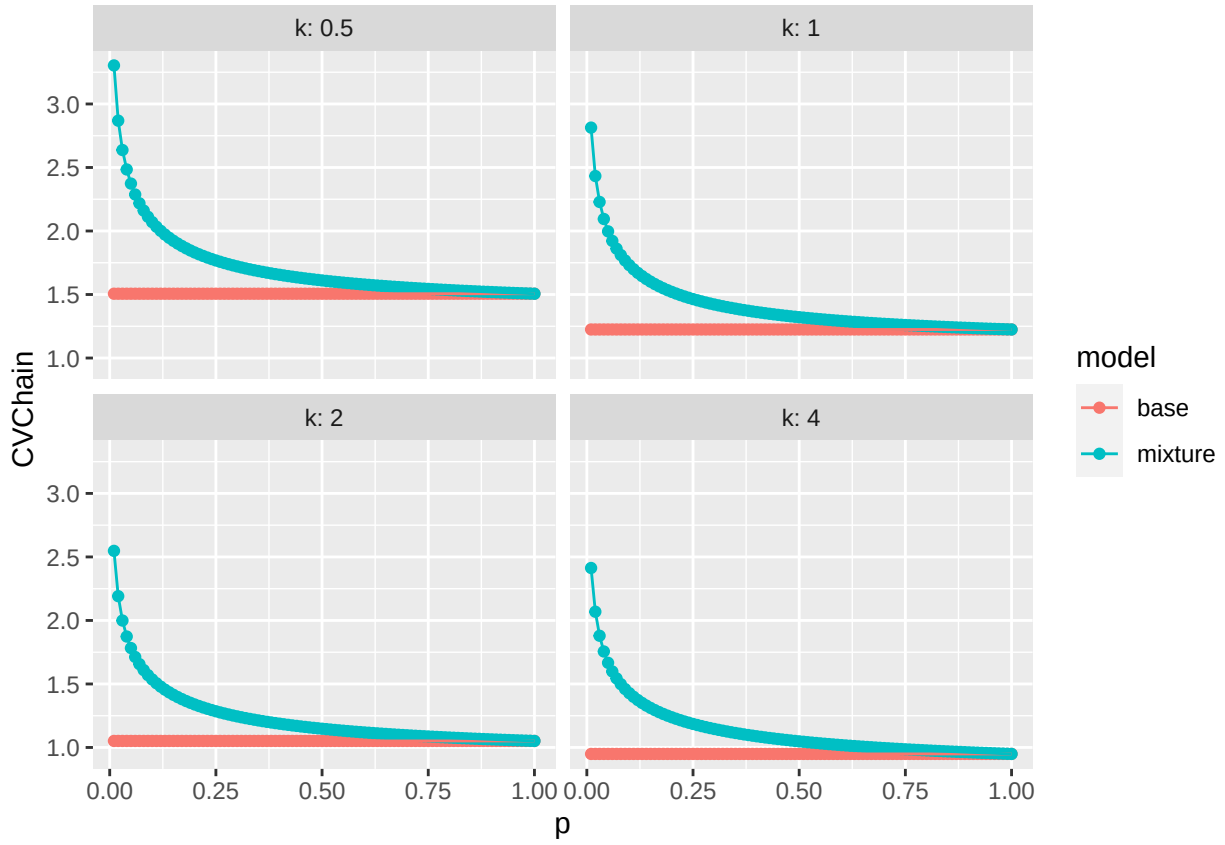


Figure 7. There is greater variability in chain sizes in the mixture model compared to the base model, even if $k > 1$, with the highest coefficients of variation observed for small dispersion parameter k, small p and large number of additional contacts.

Control activities

We study the effect of targeted control activities on the superspreading cohort. Control effort is denoted by c, $0 \leq c \leq 1$ where $c = 0$ implies the application of no control effort and $c = 1$ indicates full control of superspreading transmission. We firstly alter population structure by reducing p (thereby increasing $1 - p$) by a factor $1 - c$ while keeping all other parameters fixed. Secondly, we reduce the individual reproduction number by decreasing the number of additional contacts δ by a factor $1-c$ while keeping all other parameters fixed.. Both strategies lead the same effective R_0,

$$R_0 = R_0^D + (1 - c)p\delta.$$

When $c = 1$, effective R_0 is the same as R_0^d, the basic reproduction number of the pathogen in the regular transmission cohort. To answer the questions of which control activity induces the greatest probability of extinction and whether patterns become more heterogeneous as epidemic control c is applied, we start with $R_0 = R_0^D + p\delta = 2$, decrease each of p and δ by a factor $1 - c$ in increments of 0.01 and examine their effect on the coefficient of variation of secondary infections and the probability of extinction.

Here we will study the effect of two ways of reducing R_0:

(a) decreasing the proportion p of individuals in the population with high contact rate , which may be considered to be the same as increasing the proportion of the population that self-isolate when sick or

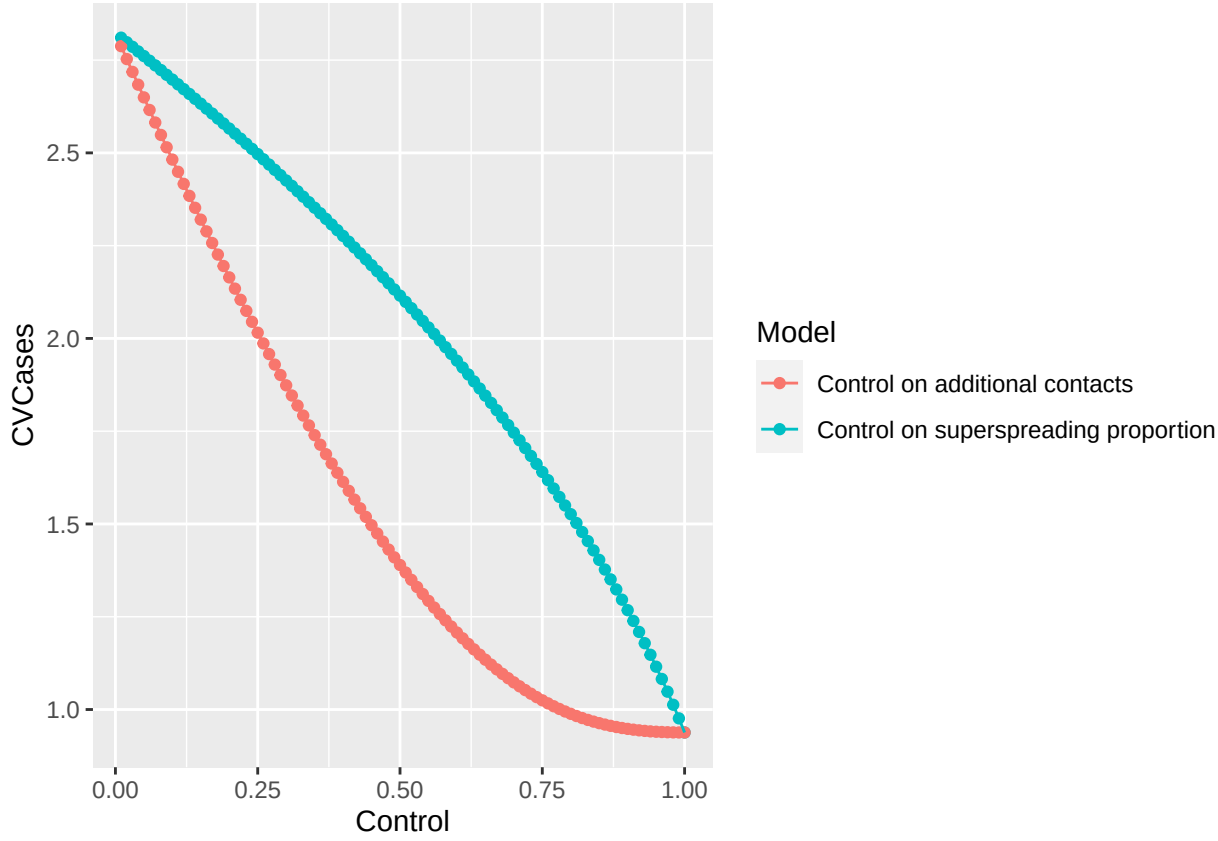
comply with stay-at-home orders, comply with face covering mandates or other measures that reduce the chance of transmission;

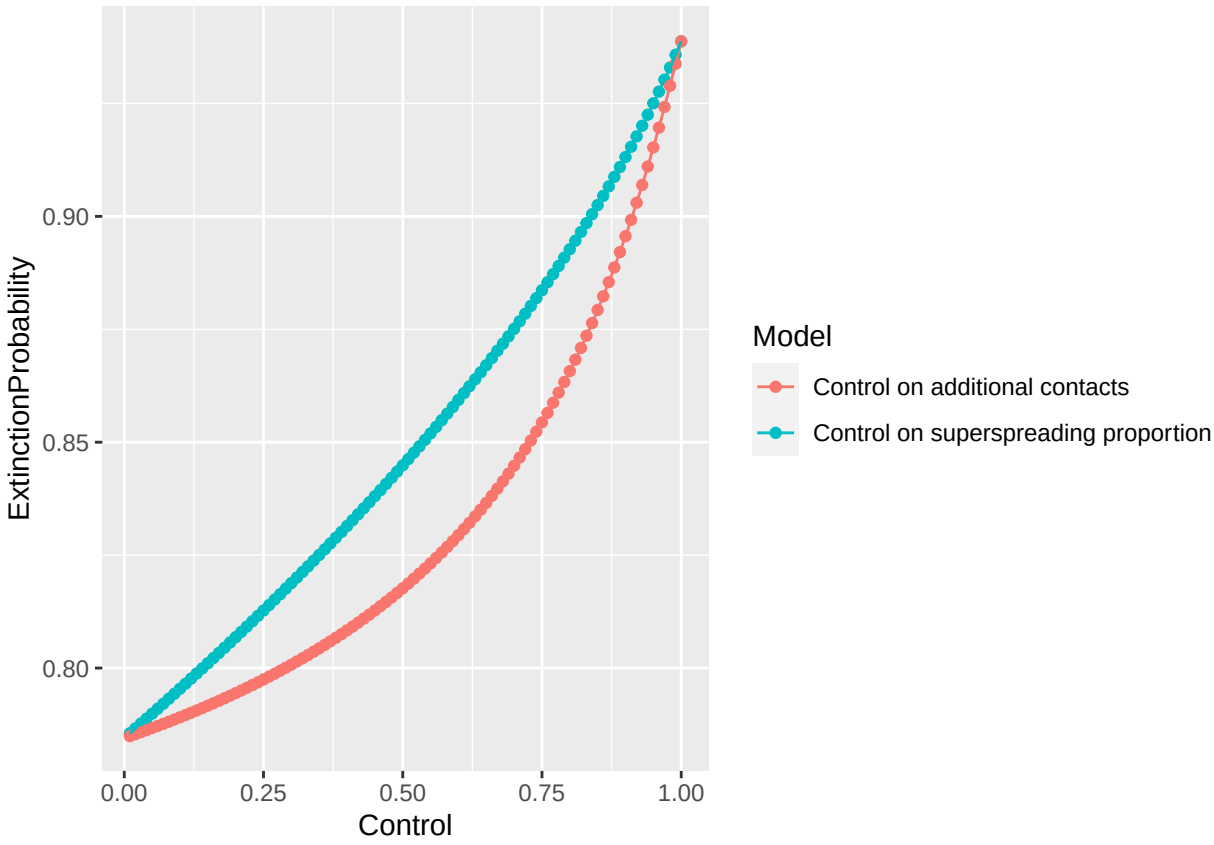
(b) decreasing the number of additional contacts per individual in the superspreading cohort.

In scenario (a), p is reduced by multiplying it by by a factor $1-c$ (which in turn increasing the proportion in the non-superspreading cohort $1-p$ by a factor $1-c$) and in scenario (b), δ is reduced by multiplying it by by a factor $1-c$. In both control scenarios, effective R_0 is the same:

$$R_0^{\text{effective}} = R_0^D + (1-c)p\delta. \tag{4}$$

Here we show that control actions that act to reduce the proportion of superspreaders p in the population lead to greater heterogeneity in outbreaks (i.e., higher coefficient of variation in secondary infections and higher extinction probability) than reducing the number of additional contacts δ. While reducing p increases the probability of extinction, it also has the less desirable effect of increasing stochastic variation in outbreaks (e.g., coefficient of variation of secondary infections).




ExtinctionProbability
0.90
0.85
0.80
0.00
0.25
0.50
0.75
1.00
Control
Model
Control on additional contacts
Control on superspreading proportion