

The distribution of outbreak sizes (challenge problem)

Immune escape and transmission of equine influenza*

John M. Drake & Pejman Rohani

1 Introduction

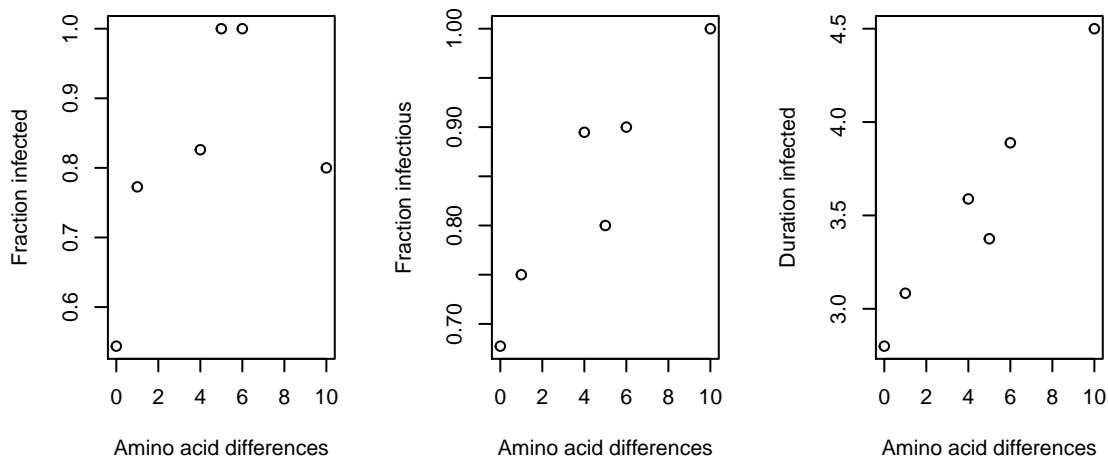
This exercise is presented as a challenge problem. It integrates information and tools taught in the first two laboratory sessions. The goal of the problem is to investigate the effect of immune escape on the probable outbreak size of recurrent influenza epidemics.

The problem is derived from a recent paper (Park *et al.* 2009. Quantifying the impact of immune escape on transmission dynamics of influenza *Science* 198:14-26). This paper reported results from experimental infections of horses with homologous and heterologous vaccines. The paper showed that the number of amino acid differences in hemagglutinin epitopes between vaccine and challenge strains of influenza correlated with three key parameters: p , the probability that a challenged subject would be infected; q , the probability an infected subject would become infectious; and d , the duration of the infectious period.

The raw data are available in the file `equine.RData`.

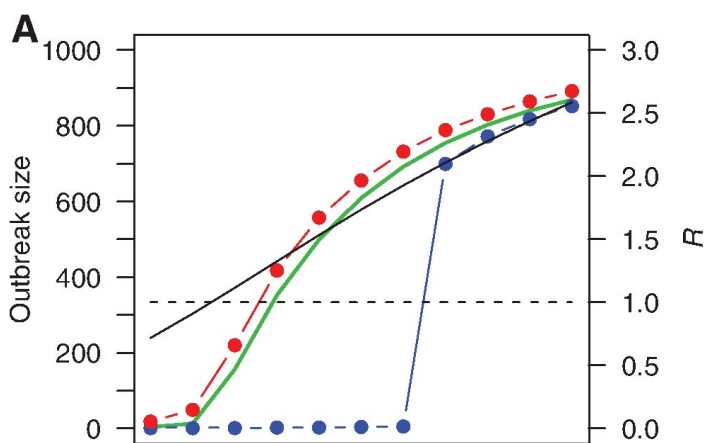
```
> load('equine.RData')      #load the data and plot
> summarize.p<-aggregate(equine$p,list(equine$a),FUN=mean)
> summarize.q<-aggregate(equine$q[!is.na(equine$q)],list(equine$a[!is.na(equine$q)]),FUN=mean)
> summarize.d<-aggregate(equine$d[!is.na(equine$d)],list(equine$a[!is.na(equine$d)]),FUN=mean)
> par(mfrow=c(1,3))
> plot(summarize.p, xlab='Amino acid differences', ylab='Fraction infected')
> plot(summarize.q, xlab='Amino acid differences', ylab='Fraction infectious')
> plot(summarize.d, xlab='Amino acid differences', ylab='Duration infected')
```

*Licensed under the Creative Commons attribution-noncommercial license, <http://creativecommons.org/licenses/by-nc/3.0/>. Please share and remix noncommercially, mentioning its origin.



The model is a modified version of the *SEIR* equations where the parameter q (which depends on the number of amino acid differences between two strains) determines the fraction of exposed individuals that move directly into the recovered class. The model equations are given in Section 4 of the supplementary material. Park *et al.* also derive the expression for R appropriate for this model: $R = \beta pqd$

An interesting finding made by Park *et al.* is that the distribution of outbreak size in a vaccinated population of 1000 individuals had extremely high variance in the vicinity of $R = 1$. Particularly, very small outbreaks were common even when the reproductive ratio was exceeded. In fact, it wasn't until $R \approx 2$ that the median outbreak size approached the size predicted from the deterministic model. This information was summarized by Park *et al.* in the following figure, which is taken directly from their paper.



***Exercise 1.** Drawing on your knowledge of stochastic and deterministic *SIR* epidemics (and the code introduced in this course), reproduce the results of Park *et al.*. Further, extend these results. At what value of R does the 5th percentile in outbreak size match the deterministic prediction?