Analysis of virulence dependence on age at infection

Petr Kouba

November 30, 2018

This document summarizes what I have learned from the structuring with respect to age at infection and its effect on virulence.

1   Fitting the deathrates to survival data

I have used the same approach as before to estimate the deathrates. With the difference, that I have only considered the years **after exposure** for fitting (Notice the shift of the beggining of the age agis in plots in Figure 1. This time, I did the fitting for 6 groups of individuals: Those who got infected at age 5, 15 and 30 and those who were exposed, but did not get infected at age 5, 15 and 30. I used the latter three classes to estimate the natural deathrate for the former 3 groups.

TODO: Estimate the natural deathrate from the control group and use it for all three compartments. See how much it differs.
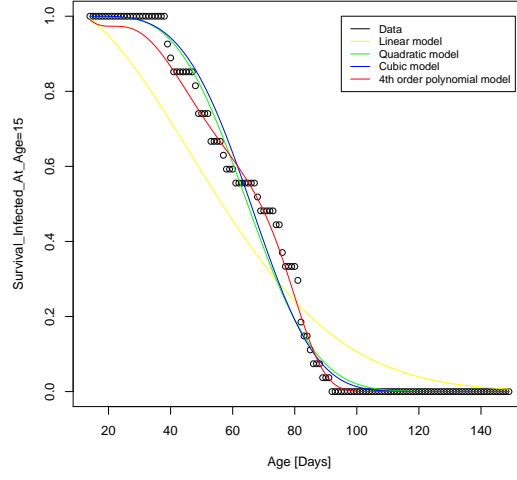
In the following, I present the plots of the survival data and the models of the survival given our various fits for delta.
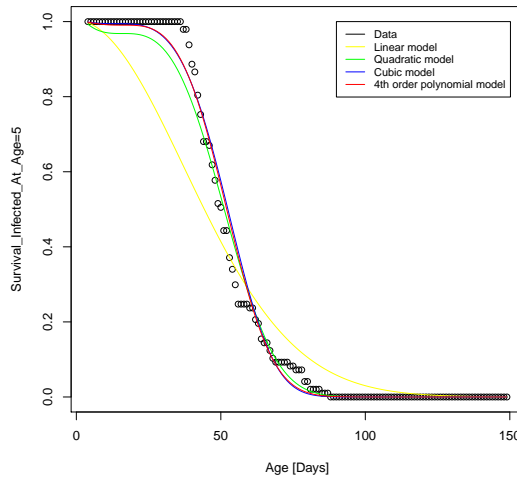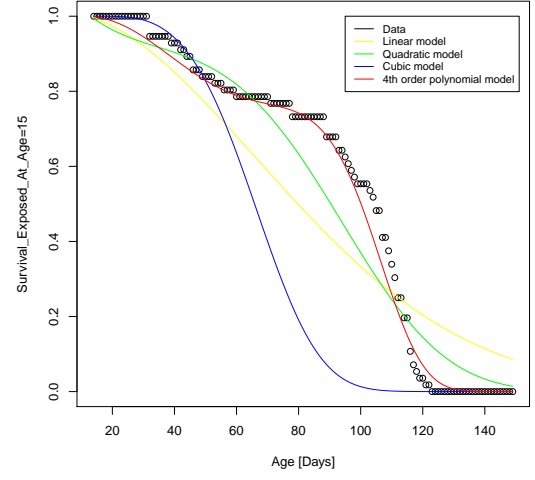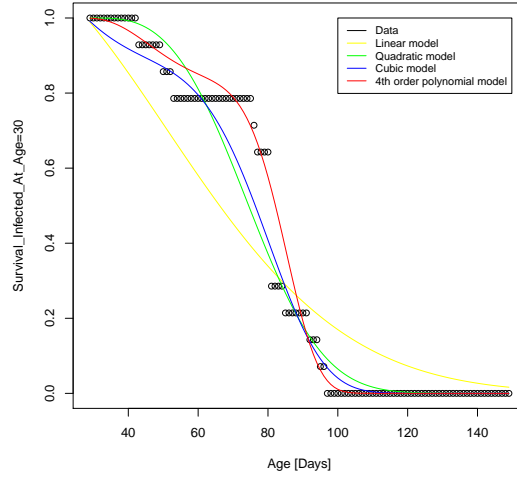

(a) Infected, age=5
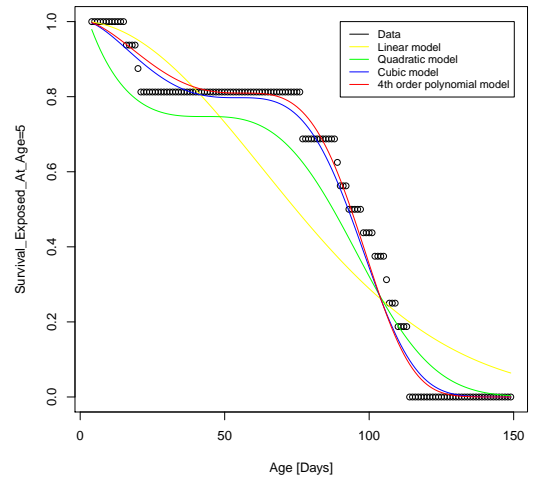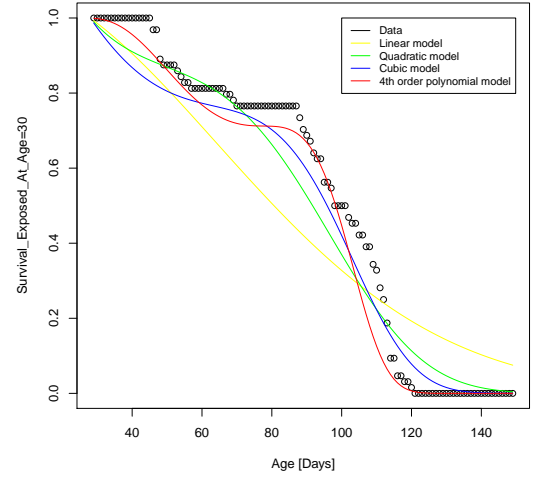
(b) Exposed but uninfected, age=5

(c) Infected, age=15

(d) Exposed, age 15

(e) Infected, age 30

(f) Exposed, age 30

Figure 1: Comparison of the survival patterns and corresponding models of survival based on different fits for delta

3

2 Obtained virulances

TODO: Death rate plots comparing the deathrates of different exposed compartments and the control population can also be put in this section

2.1 Choosing the best fit for each compartment

TODO: I will do the likelihood ratio tests for each fit of compartment to which I fitted the delta, but first I need to get the absolute values of the likelihood including the binomial coefficients

Question: Supposed I have the LRTs for all the fits, but they yield polynomials of different orders as the best fits for different age-at-infection groups. Is it then legitimate to use the best fit for each compartment? Or should we stick with consistency, in that sense that we claim te order of the polynomial to be the same within a group of the same infection status (infected/exposed)?

Sofar I used our previous results, obtained from the fitting of the total control/exposed/infected populations (before structuring wrt. age at infection) and I picked the quadratic fit for the infected individuals and 4th order polynomial for the exposed individuals.

2.2 Virulence with respect to the age of infection

I have expressed the virulences for given age-of-infection groups as the difference of the death rates of infected and exposed populations:

$$V_i(a) = \delta_i^{infected}(a) - \delta_i^{exposed}(a) \tag{1}$$

Where V stands for virulence, a for age, δ for death rate and $i \in \{5, 15, 30\}$ is the age at infection.

Taking these virulences, I have transformed them into functions of the age of infection, by shifting the independent variable (formerly age) by the age at infection $a \to (a_{infection} = a - i)$. The transformer virulences (i.e. virulences as a function of the age of infection) are plotted in the Figure 2.

From Figure 2, we can see that the virulence tends to be higher in the individuals who were infected at younger age. We now need to verify this conjecture.
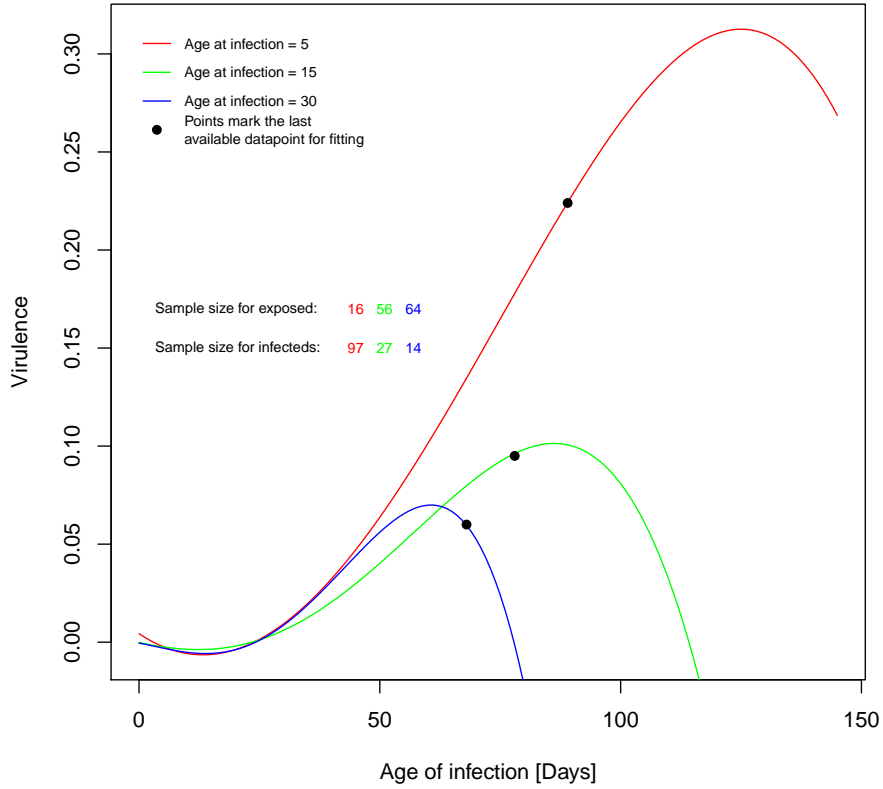
4

Figure 2: Virulences as functions of the age of infection, for different ages at infection. The fits might not be reliable beyond the black dots, since the dots mark the highest age of infection, where there was a datapoint serving as a support for our fit.

2.3 Veryfing the significance of the influence of the age at infection on virulence

In order to try to verify the significance of the effect of the age at infection, I came up with the following method, which seems reasonable, but I am not yet sure if it does not cause too big error amplification. We could plot the distribution of deaths due to virulence along the age of infection, for different age-at-infection classes. And our null hypothesis would be that the distribution does not depend on age at infection and therefore it should be the same for all of the 3 age-at-infection classes.

Lets have a look at it step by step. First, we provide the plots of the distribution of deaths along the age of infection axis. To see the normalized histograms representing the distributions of deaths, have a look at Figure 3.

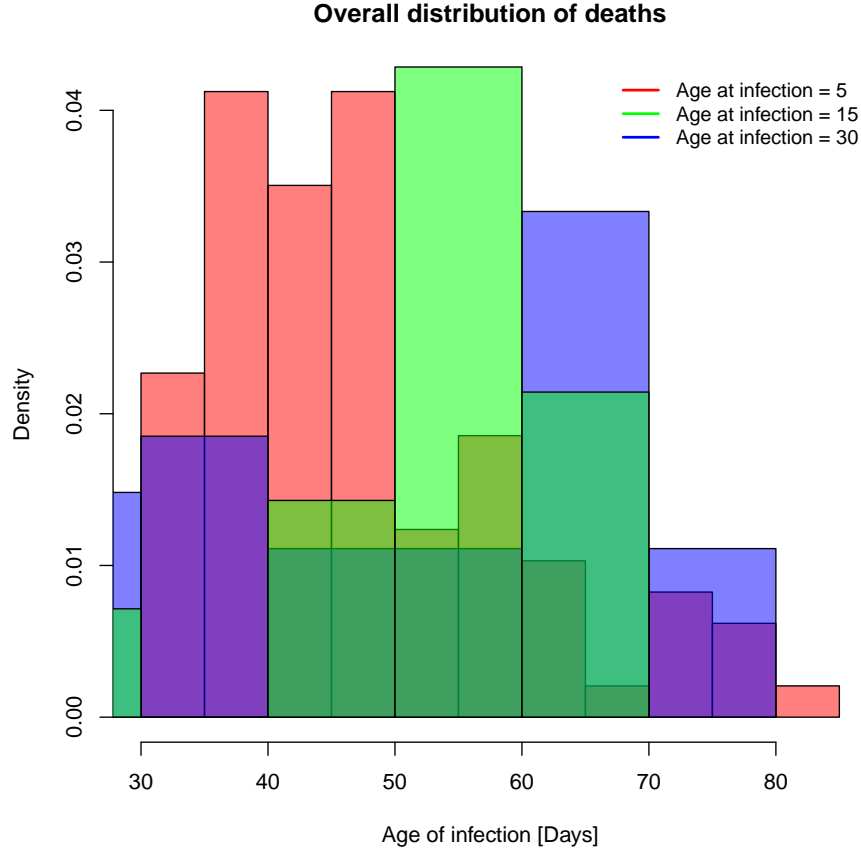**Overall distribution of deaths**

Figure 3: The overall distribution of deaths, for different age-at-infection classes

However these distribution contain lot of signal from the deaths due to natural causes. To suppress this, I decided to weight the contribution of each death to the histogram, by the ratio of virulence to the total deathrate for a given age.

For example: if an individual, infected at age 15, dies at age of infection 50 (that is at age 65), and the virulence at age 65 is 0.038, whereas the total deathrate at age 65 is 0.041, we will count it towards our histogram with the weight $0.038/0.041 = 0.93$.

Weighting our histograms as described above yields the histograms provided in Figure 4
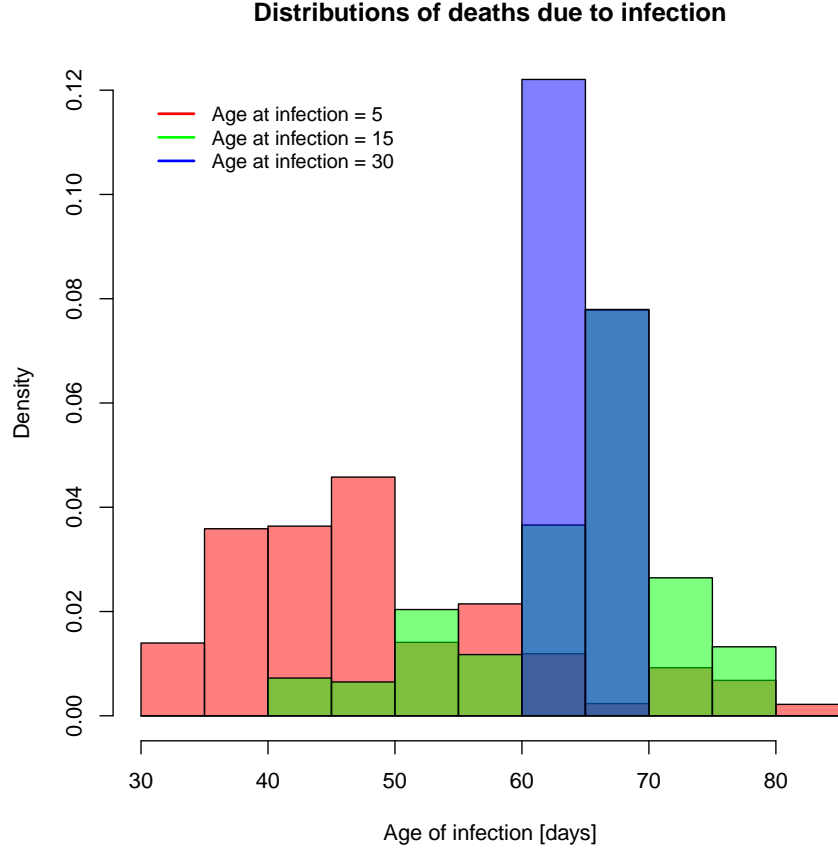
Distributions of deaths due to infection

Figure 4: The distribution of deaths due to virulence only, for different age-at-infection classes

For some ages, there was a negative virulence, which does not make any sense and it is only an imperfection, which can be understood from the equation 1. It simply happened that for some ages, we had higher death rate in the exposed population then in the infected population. We explain this by being in the stage of the infection when it does not kill the host and due to some fluctuations we even had a higher death rate in the exposed population. Therefore, we interpret the virulence in those years to be 0, this fixed our problem with potentially negative weights in the histogram.

From the Figure 4 we can once again see, that the hosts who were infected at younger age tend to die at earlier age than the hosts who got infected later, which again supports the conjecture that the infection is more virulent in younger hosts.

I think by the above procedure I have transformed the problem into a more suitable form, since this now seem better fit for hypothesis testing. I pick a suitable distribution and test the probability that all of those three distributions in Figure 4 were distributed according to it.

TODO: Think about how to choose the distribution for our null hypotheses. Perform the hypothesis testing, to potentially rule out the hypothesis that virulence is independent of the age at infection.