

Report_01

Jan van Rongen

2021-10-07

Introduction

We analyse data from the State of Israel related to COVID vaccines and infections. This analysis started as a reanalysis of a document circulating on Internet [1]. The previous technical document explored the data, while this document focusses on proper modelling that was not found in [1].

The document itself is an RMarkdown worksheet that, when formatted in pdf, hides the code.

Data sources and cleaning

As before We use three data sources: two from the Israeli government site and one from a spreadsheet with population data. The data was imported manually on 2021-10-02.

The file `vaccinated-per-day-2021-09-28` is aggregated to a per week file. Numeric fields with `<5` or `<15` were converted to 3 and 8 resp. Weeks are identified by their first day. Four missing records for the 90+ age category were added to `cases-among-vaccinated-134.csv`

Using the `population` table we construct cumulative totals of fully vaccinated, single vaccinated and not vaccinated.

Goal and methods.

To analyse as best as possible two groups: fully vaccinated vs non vaccinated on rate of infection. Previous analysis (`tech_doc_3`) shows that a detailed breakdown in age groups does not help. Adults were vaccinated immediately, under 20 much later. So we can restrict ourselves to these two groups.

Fully versus unvaccinated.

Fully vaccinated in week x : those who received exactly two doses and their last dose in or before week $x-2$. It is highly unlikely that they were infected twice while fully vaccinated so we assume that did not happen.

Unvaccinated: again here we assume that we measure the first infection.

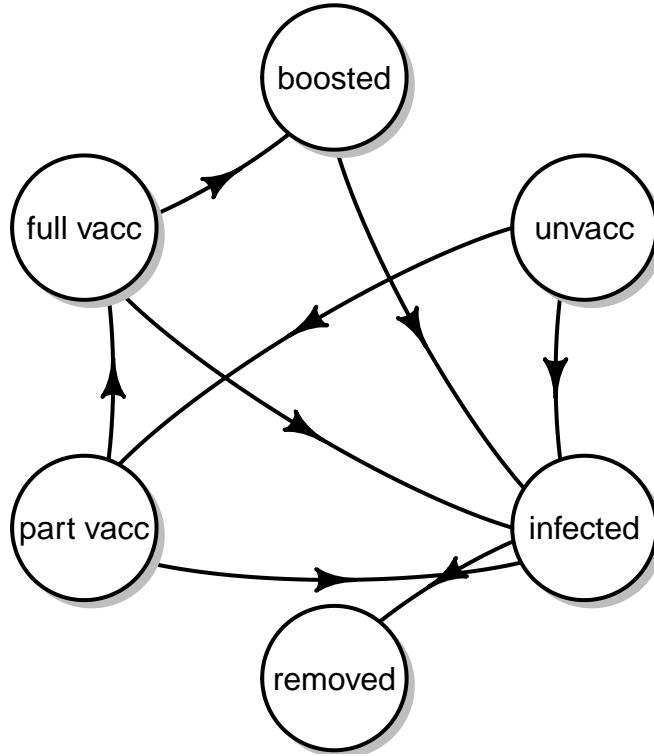
Disease prior to start of 2020-12-20 vaccination: we assume that the population that already had at least one infection is $W = 0.15$ of the total population. That should be a parameter that can be changed,

State Transition diagram

At the start of the period under investigation, there are only unvaccinated people. After that, there are three vaccination states: partial, full and boosted. In any of those 4 states people can become infected. After infection they might not recover, so we have nine states in total. However that is an unworkable model as many transitions cannot be calculated.

Thus we use the following transitions where every infected person is removed from the data, corresponding to the common practice in cohort studies where there is an end state after the event under consideration. So our model removes the infected people after 1 week.

State Transition Diagram



It is possible that infected people re-enter the system to get vaccinated. There is no way we can tell.

Calculating Susceptible population.

For this diagram we can derive the number of people in each state, even though we do not know exactly which state transitions occur from week to week.

Algorithm

The assumption is that the effectiveness of the vaccine decreases over time. From the SIR model we can learn that for unvaccinated population S_n , the amount of new incidents is a linear function of the effective reproduction number $R_e(t)$. A vaccine with initial effectiveness $0 < \beta < 1$ will have decaying effectiveness with a rate of the shape $(\beta \cdot e^{-\alpha \cdot t})$ so the reproduction number for this group is one minus that times the $R_e(t)$. Taking two groups, the risk ratio is the quotient of those reproduction numbers, so $RR_{f/n} = 1 - \beta \cdot e^{-\alpha \cdot t}$.

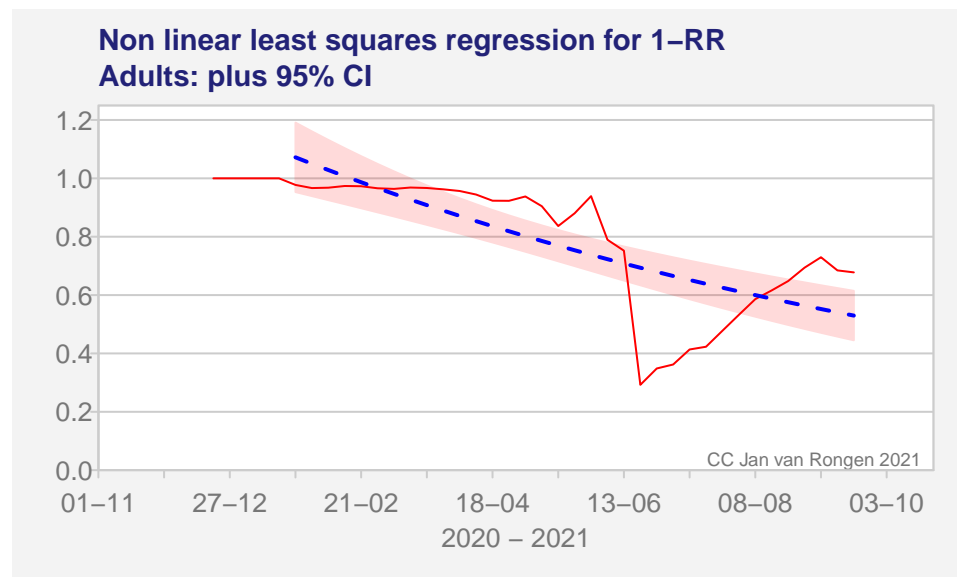
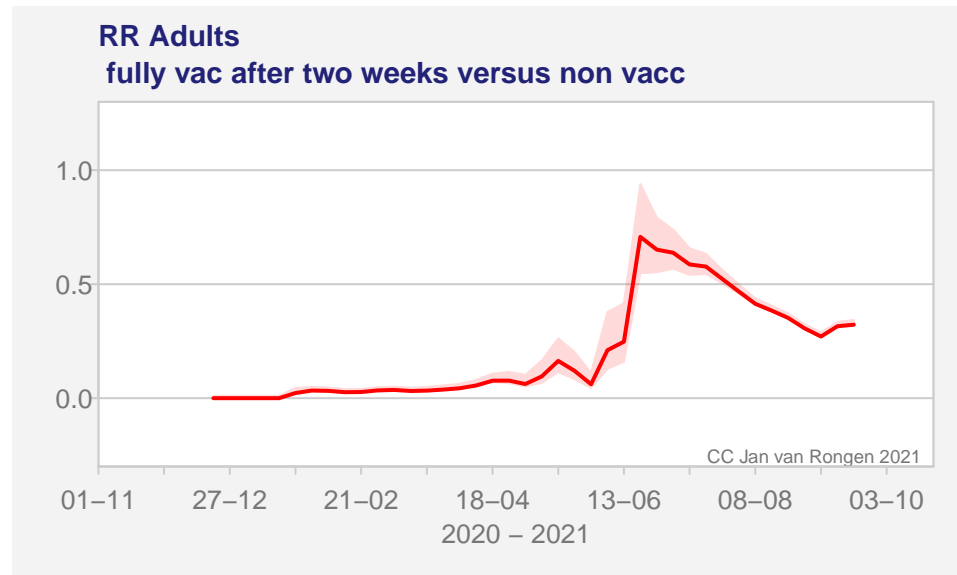
This is in fact an exponential regression of $RR_{f/n} - 1$ against the function $-\beta \cdot e^{-\alpha \cdot t}$. This looks familiar: it models the Vaccine Effectiveness as an exponential function. That is a Cox regression.

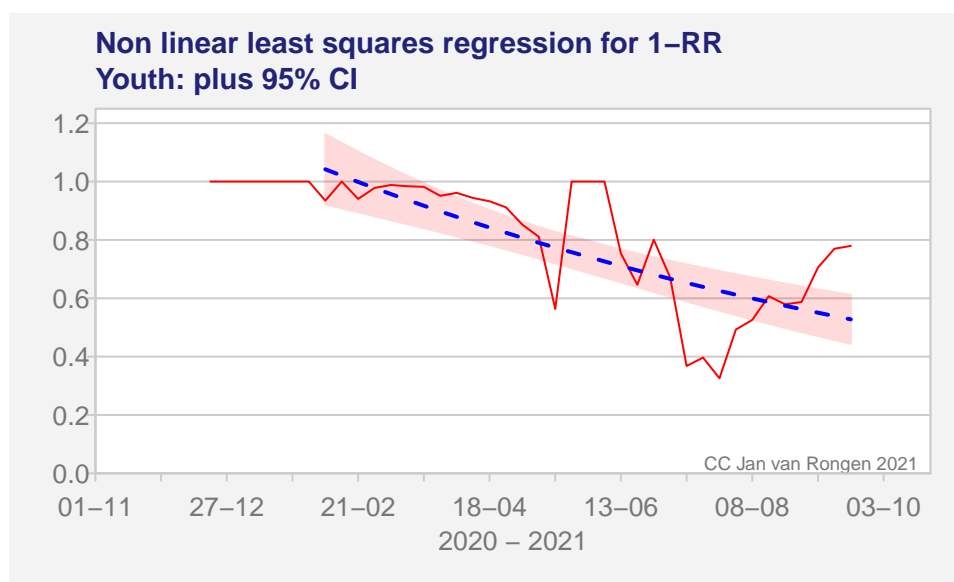
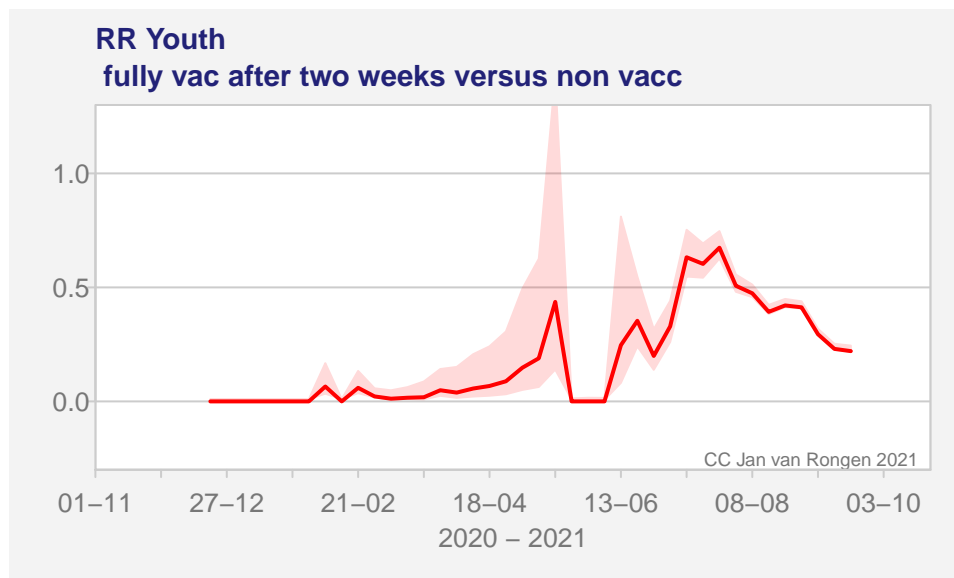
This is a so called non linear least squares; we use the R function `nls` from the `base` library and we estimate the confidence interval using the `invertR` package.

Get the data ready

We construct two data-frames for the Adults and the Youth Age groups with the relevant information. We assume that 15% of the population was infected before start of vaccination and so is excluded from the system.

And now the estimates.





Discussion

The adjustments on the input data are more realistic than previous attempts. The effect is that the RR decrease by the adjustments for positives, but increases when the proportion of positives in 2010 is increased. The overall picture is of unreliable and missing data in the summer period.

The regression shows the uncertainty in the estimate of the curve given the data. So the uncertainty in the data is not shown together with the regression. No need to do that as the picture is clear enough: the result is inconclusive.

From here we might conclude that the decay has a half-time of 6-10 months, but the data clearly does not fit the exponential model very well. In fact you would never think of such a regression if you would only see the original curve.

As the above model is a Cox regression in disguise, we must conclude that any survival-like model does not model this data.