

Technical Document vs 0.2

Jan van Rongen

2021-10-03

Introduction

We analyse data from the State of Israel related to COVID vaccines and infections. This is the technical document that contains all code used. This is a reanalysis of a document circulating on Internet.

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

Data sources and cleaning

We use four data sources: three 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.

One file not yet used.

Definitions and methods

Fully vaccinated are people from the day of their second dose. *Single vaccinated* are people that had one dose but not two (from the date of first dose). All others are *unvaccinated*. *Infected* are people that (on a certain day) tested positive. All others are *not infected* on that day.

A *cross table* is a 2x2 table with two (0,1) categories. The entries are the number of people in that combination of categories. When the matrix is $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then the *relative risk* AKA *risk ratio* AKA *RR* is $(a/(a+b))/(c/(c+d))$. The *odds ratio* is also known as *OR* is $a.d/b.c$

In epidemiologic RR and OR are used a lot. Numbers can be quite large, so in programming we have to avoid numerical overflow or incorrect rounding.

Data quality

Coding small numbers as `<5` or `<15` is a bit strange, but let's skip that. The population data has a bit of a problem. The table has 9215400 for the total population, while wikipedia has 9364000, for 2019. That means we will underestimate the number of unvaccinated people by (absolutely) 1 percentpoint.

One more serious problem is the way the Israeli population is defined. All 450K+ colonists in the occupied part are counted, but not the Palestinians that live there. Colonists have access to vaccines, Palestinians not. That means that we actually work with incomplete data (or overcomplete, depending on your point of view).

Another point: I do not know if the most recent data is complete or whether there is some left in the pipeline. That can be done by comparing with an older version of the files, but that is for later.

Finally something that comes out of a later analysis of the RR for Adults. There is a sharp increase around mid june, at the same time that there is a sharp increase in the number of infections. If the date in the `cases` file is the date of report instead of Day Of Onset, or a mix of both, that could explain a fast increase of the RR. In this case it would mean a slower reporting of infections with non vaccinated people.

Data overview

Note that one table has 40 weeks, one has 41, so we can only combine the data of the first 40 weeks.

First we aggregate over the 20+ age groups.

Fully vs non vaccinated Adults

In fig. 1, the last week in this sample, 83% of the Adults is fully vaccinated.

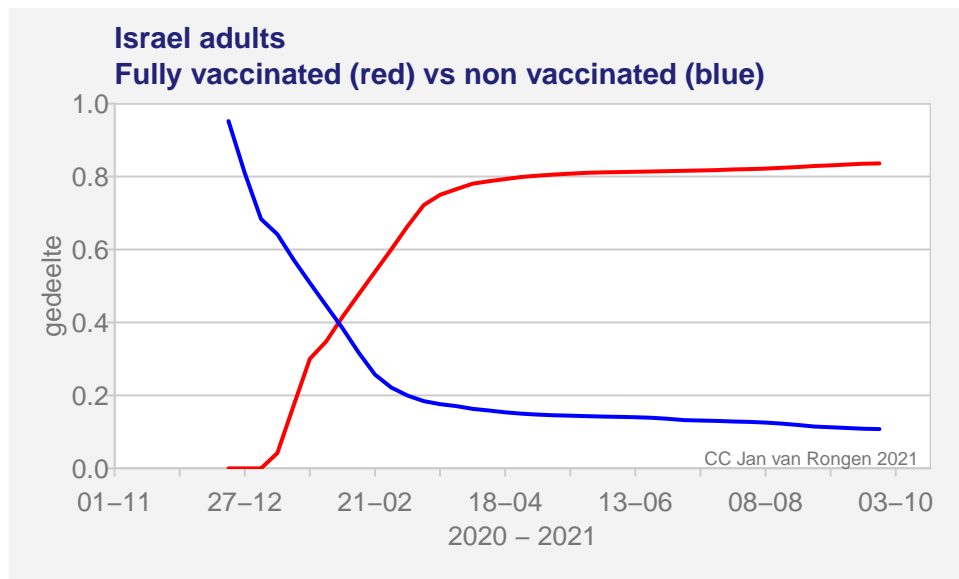


Figure 1: Degree of vaccination

Same for youth

Youth (age group 0-19) has a low percentage fully vaccinated. 75% is not (yet) vaccinated at all. See fig. 2.

Interesting second wave

See fig. 3.

There are two remarkable things here: almost no infections from april to end june and a high wave of infections later.

That begs the question: what is the RR? Did vaccination help? For the april-june period the numbers are probably too low for any accurate estimate, but for that second wave?

Relative Risk for Adults

See fig. 4.

That means that the relative risk for fully vaccinated adults in the week starting 2021-09-19 is 23% in other words not vaccinated people are at least 4 times more likely to get infected.

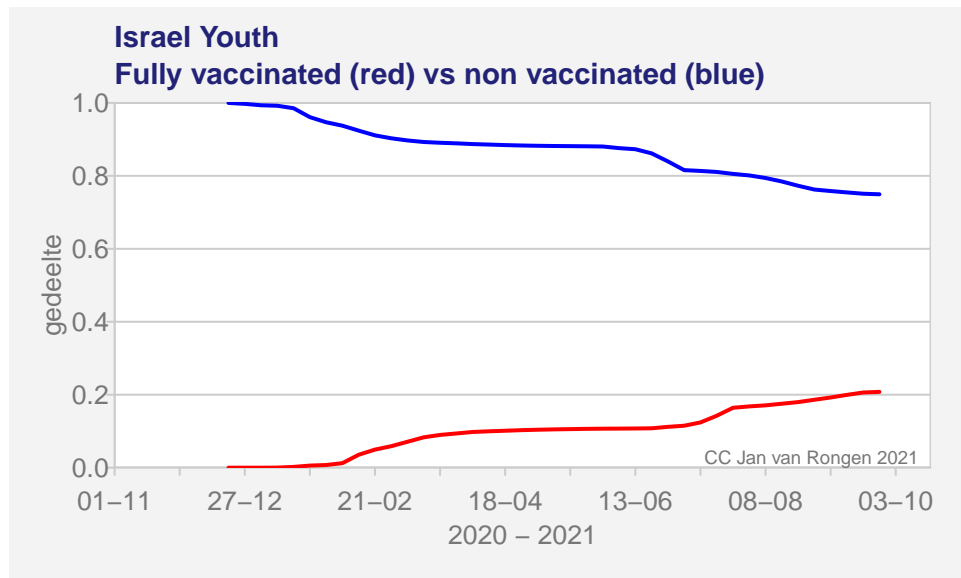


Figure 2: idem, Youth

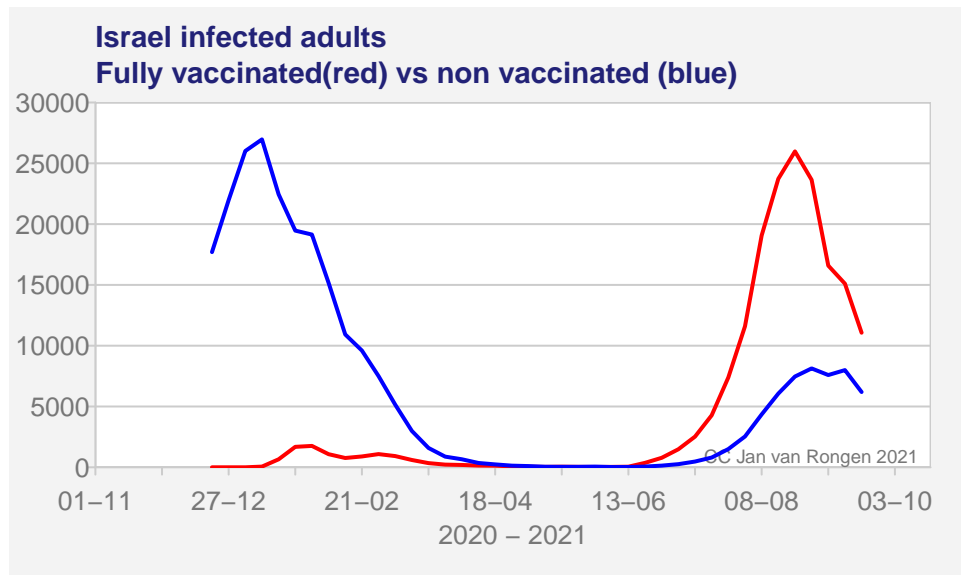


Figure 3: Infection Patterns

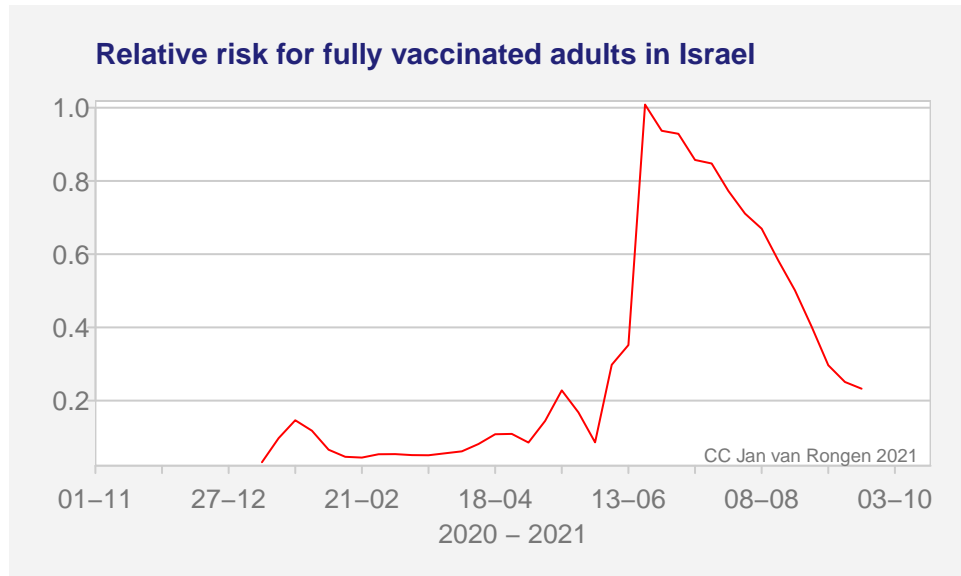


Figure 4: Adults RR

Without the third shot

Maybe the decrease is caused by the start of “boosting”: using a third shot. We repeat the above plot with data from people that had no third one. See fig. 5.

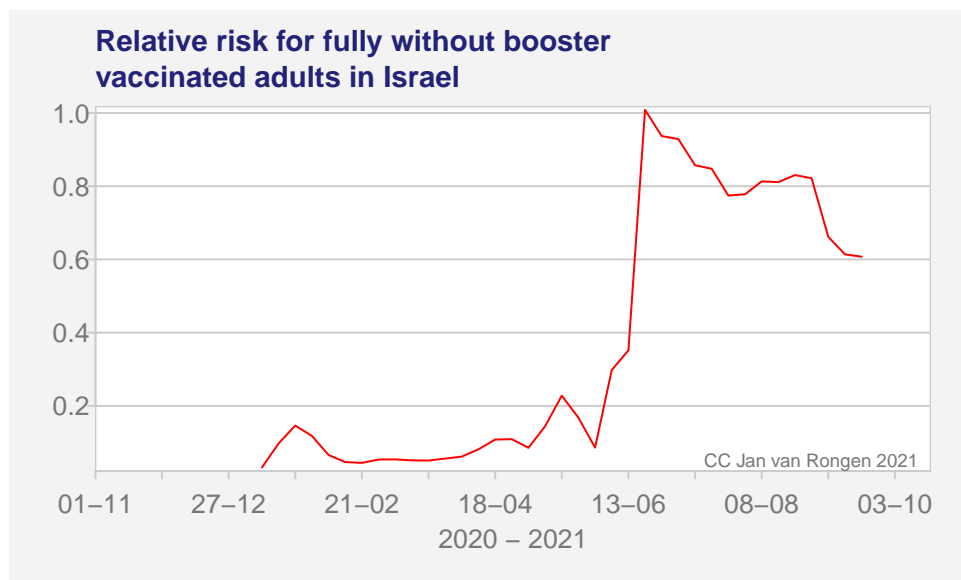


Figure 5: Adults RR (2)

Conclusion

With the above overall data, it does not make much sense to look at the behaviour of smaller (age) groups. We will probably see another example of the ecological fallacy.

Figures 4 and 5 suggest that a third shot (“boosting”) helps. But the RR seems so volatile here, without further analysis we simply do not know.

Finally, in no way we see any added value to introduce a measure of $VE=(1-RR)$. It is not found in the literature of epidemiology AFAIK, see f.i. https://www.wikilectures.eu/w/Attributable_and_Relative_Risks,_Odds_Ratio/

References

This file can be found on github. All referenced documents are in a subdirectory.