Rapid Reviews COVID-19

Review 2: "Herd immunity thresholds for SARS-CoV-2 estimated from unfolding epidemics"

Pieter Trapman¹

¹Senior Lecturer, Mathematics, Stockholm University: Stockholms Universitet, Sweden

Published on: Nov 17, 2020

DOI: 10.1162/2e3983f5.5f18b7ab

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

RR:C19 Evidence Scale rating by reviewer:

Not informative. *The flaws in the data and methods in this study are sufficiently serious that they do not substantially justify the claims made. It is not possible to say whether the results and conclusions would match that of the hypothetical ideal study. The study should not be considered as evidence by decision-makers.*

Review:

In this manuscript the impact of heterogeneity in susceptibility and connectivity of people in a population on the herd immunity threshold (HIT) for SARS-CoV-2 is investigated. The main conclusion is that, because of this heterogeneity, herd immunity can already be reached at levels as low as 10% on a country level.

I think the manuscript is not informative, because I do not believe the quantitative results and think that the claims made may be dangerously misleading, for reasons I will explain those issues below. However, I think that the paper has clear value in showing the qualitative impact of variation in susceptibility and connectivity on the HIT.

The main modeling issue

The main issue I have with the mathematics of the paper is the choice of creating heterogeneity in susceptibility by assuming that this susceptibility is gamma distributed. It is true that threshold parameters, and real time or “generation based” growth rate for epidemics on (nice enough) random networks depend on the mean and the variance of the degree distribution (which is what is meant with connectivity in this manuscript). However, the final size of an epidemic, and also whether the epidemic will go up again when control measures are lifted is very much dependent on how many individuals there are with low degrees or with very low susceptibility.

What I think explains the results of this paper, is that by increasing the Coefficient of Variation (CV), more and more individuals will have very low connectivity or susceptibility and because of that will not get infected. This problem can be illustrated by looking at Extended Data Table 1 for Portugal as a whole. The gamma distribution used has expectation 1 (by design) and variance $(4.26)^2$. That means that already a fraction of over 68% has a Susceptibility below $1/100$. If $R_0 = 4.26$, even if the entire

population is infected apart from one individual with susceptibility $1/100$, that individual has probability $e^{-4.26/100} \approx 0.96$ of escaping infection.

With the above parameters even if everybody in the population is exposed to an infectious pressure which corresponds to the entire population being infected, then still only $21\% = \int_0^\infty g(x)(1 - e^{-R_0 x}) dx$ of the population will get infected. Here $g(x)$ is the density function of a gamma distributed random variable with expectation equal to 1 and CV equal to 4.26.

That such a large part of the population is a-priori almost immune is implicit in the manuscript and should be analyzed in further detail. In particular, one needs to know whether the CV of the distribution of susceptibility is important or the fraction of the population which is immune for all practical purposes? To me the results of the manuscript become less surprising considering explicitly the fraction of the population which is almost immune. If we ignore the 68% of the population with susceptibility below $1/100$, then the paper states that of the remaining 32%, a fraction $0.073/0.32 \approx .23$ has to be immunized. Where we should note that of that 32% still many have susceptibility below 10% and therefore still a very low chance of getting infected.

It is true that the gamma distributions are fitted to data from different countries and the modelled curve is pretty good on visual inspection. This, however, might possibly be explained by heterogeneities in the population which are due to geographical location or because of the way people respond to the pandemic, and contain no information on how the epidemic will spread if people go back to normal contact patterns: if there are geographical subregions in a country or region which largely escaped the pandemic, then many people in those subregions will escape infection because they did not get exposed. By the nature of the model in the manuscript this escape of infection has to be ascribed to low susceptibility (or connectivity), which is misleading. Similarly, people who use efficient ways of social distancing will be treated as people of low susceptibility or connectivity, which in the model will be maintained if measures are lifted.

Observations from the past's future

The predictions provided in the main figures are based on data until the end of June. Although the model predicts second waves in Belgium, Portugal, and Spain, I think that recent data show that the size and the duration of those waves are in reality larger than predicted in the manuscript. England definitely has a larger outbreak than predicted.

Modeling the impact of Non Pharmaceutical Interventions (NPIs)

In all epidemiological models, assumptions which are not justified by data have to be made. Often, this is not a problem if one wants to obtain qualitative insight in the model, but it might lead to wrong quantitative predictions. The assumptions on the shape of impact of interventions is quite arbitrary of having three weeks linear increase followed by constant impact for thirty days and then linear decrease of careful behavior back until baseline. It is claimed that there is an excellent agreement with observed mobility patterns (ref 6), but this is not really shown in the manuscript. Furthermore, it is questionable whether the mobility patterns give a perfect proxy for contact behavior. In addition, it is assumed that the impact of NPIs is independent of susceptibility levels and that there are no confounders in e.g. age, occupation etc. I agree that those assumptions are reasonable choices (you have to assume something, and every choice would be arbitrary), but they are unlikely to be realistic, and therefore quantitative predictions should not be trusted.

Remarks on the assumptions underlying the Markov SEIR model

The "Markov" SEIR model in which there is a constant rate of going from Exposed to Infectious and from Infectious to Recovered is not justified in the paper. It is known that the generation interval is very important to find a relation between R_0 and the real time growth rate. I would not have much problem with this assumption for obtaining qualitative results, but for quantitative results further sensitivity analysis regarding this assumption is necessary.

In addition, the model would gain some realism if instead of having a fraction of the exposed people being able to infect, creating an extra compartment of being "pre-symptomatic infectious". It seems likely that people who are in the E class are not able to infect anybody when they were just infected themselves, while they might be infectious in the few days before they start to show symptoms.