

Estimating population immunity without serological testing

Andrew Lesniewski
Department of Mathematics
Baruch College
55 Lexington Avenue
New York, NY 10010

Draft of April 23, 2020
First draft: April 20, 2020

Abstract

We propose an approximate methodology for estimating the overall level of immunity against COVID-19 in a population that has been affected by the recent epidemic. The methodology relies on the currently available mortality data and utilizes the properties of the SIR model. We illustrate the application of the method by estimating the recent levels of immunity in 10 US states with highest case numbers of COVID-19.

The purpose of this brief note is to propose an approximate methodology for determining the level of immunity in a population affected by (an infectious disease such as) COVID-19. In the absence of reliable, large scale testing, paucity and spotty character of case data, our approach relies on the mortality data.

The proposed methodology is predicated on the following simplifying assumptions.

- (i) The deterministic SIR model [1] is an adequate mathematical specification of the COVID-19 epidemic. In particular, we assume that individuals who have recovered from an infection develop a long lasting immunity against reinfection.
- (ii) The basic reproduction ratio has the value given by

$$\mathcal{R}_0 = 2.28. \quad (1)$$

This value was found to be the midpoint of the range estimated by means of MLE in [8], see also [5].

- (iii) The infection rate parameter of the SIR model is given by

$$\beta = 25. \quad (2)$$

This choice is motivated by the fact that $\tau = 1/\beta$ is the characteristic time scale (in the units of a year) of the spread of the epidemic. Given that the length

of the full cycle of COVID-19 appears to be about 6 months, we believe that any choice in the range of 20 to 30 would be appropriate. The choice above is probably conservative. The above values of \mathcal{R}_0 and β dictate the following value of the recovery rate parameter $\gamma = \beta/\mathcal{R}_0$ of the SIR model:

$$\gamma = 10.96. \quad (3)$$

- (iv) The reliance on the mortality data introduces a lag of l days between the date when an individual contracts the disease and the date they die. For the estimates below we assume

$$l = 20. \quad (4)$$

This choice is justified by a number of studies, see e.g. [2] and references therein: “The symptoms of COVID-19 infection appear after an incubation period of approximately 5.2 days. The period from the onset of COVID-19 symptoms to death ranged from 6 to 41 days with a median of 14 days”.

- (v) The infection fatality rate (IFR) is given by the value

$$\mu = 0.005. \quad (5)$$

There is a wide spectrum of estimates of μ in the literature, see e.g. [3], [4], [6], based on complete tests conducted on small closed populations exposed to COVID-19, but there appears to be a consensus that the actual range should be between 0.1% and 1.0%. The difficulty lies of course in the fact that the IFR is based not only on the observed cases, but also on undiagnosed and asymptomatic cases. It also shows a strong dependence on other factors such as the patient’s age.

Notice that all parameter values selected above are population averages. Assuming distributions of values across risk factors such as incubation length, age group, geography, etc. could conceivably lead to more refined forecasts.

Our algorithm relies on the following properties of the SIR model. Let $S(t)$, $I(t)$, and $R(t)$ denote the population fractions of susceptible, infected, and removed at time t , respectively. Note that, in this model, $R(t)$ includes the individuals who have left the population through mortality at the rate μ . Then:

- (i) $S(t)$, $I(t)$, and $R(t)$ add up to 1, for all t :

$$S(t) + I(t) + R(t) = 1. \quad (6)$$

- (ii) The following conservation law holds for all t :

$$S(t) + I(t) - \mathcal{R}_0^{-1} \log S(t) = 1. \quad (7)$$

- (iii) The following relation holds:

$$\int_{T_0}^T I(t) dt = -\frac{1}{\beta} \log \frac{S(T)}{S(T_0)}, \quad (8)$$

where $T_0 < T$.

We now proceed as follows. We replace the continuous time variable t with daily samplings $t_n = n\delta$, $n = 0, 1, \dots, n$, where $\delta = 1/365$, and denote $R_n \equiv R(t_n)$. Here, $t_n = T$ corresponds to the latest observation, and $t_{n-l} = T_l$ corresponds to the lagged date $T_l = T - l$ days. We estimate R_{n-l} , the removed fraction on the date $n - l$, as

$$R_{n-l} = \frac{F_n}{\mu N}, \quad (9)$$

where N is the population size, and F_n is the total number of fatalities up to date n . Properties (i) and (ii) above imply that $-\mathcal{R}_0^{-1} \log S_n = R_n$, i.e.

$$S_{n-l} = e^{-\mathcal{R}_0 R_{n-l}}. \quad (10)$$

This implies that

$$I_{n-l} = 1 - S_{n-l} - R_{n-l}. \quad (11)$$

Now, from the third property we conclude that

$$\begin{aligned} \log S(T) &= \log S(T_l) - \beta \int_{T_l}^T I(s) ds \\ &\approx \log S(T_l) - \beta l \delta I_{n-l}, \end{aligned}$$

i.e.

$$S_n \approx S_{n-l} \exp(-\beta l \delta I_{n-l}). \quad (12)$$

Similarly,

$$R(T) = R(T_l) + \gamma \int_{T_l}^T I(s) ds,$$

i.e.

$$R_n \approx R_{n-l} + \gamma l \delta I_{n-l}. \quad (13)$$

Also,

$$I_n \approx 1 - S_n - R_n. \quad (14)$$

The approximations used in the calculation above are reasonable, but it is impossible to give an *a priori* estimate of the error made: they may equally well overshoot or undershoot the actual value of the integral $\int_{T_l}^T I(s) ds$. We should also point out that the algorithm is extremely easy to implement in computer code.

Assuming that all individuals who have recovered from the infection develop a long lasting immunity to the disease, we are thus led to the following estimator of the percentage of individuals with immunity:

$$\text{Percentage of individuals with immunity} = (1 - \mu) R_n. \quad (15)$$

We have used the estimator (15) for the recent data of the 10 US states with the largest numbers on COVID-19 cases. The results, using the data taken from [7] as of April 20, April 21, and April 22 of 2020, are summarized in Table 1. Each of the columns contains the estimated percentages of immune individuals for the date indicated.

State	% Imm 04-20	% Imm 04-21	% Imm 04-22
NY	29.16	30.15	31.00
NJ	15.95	17.20	18.22
MA	8.76	9.46	10.47
PA	3.64	4.34	4.60
CA	1.08	1.15	1.26
MI	8.33	9.08	9.44
IL	3.68	4.00	4.26
FL	1.34	1.41	1.51
LA	9.57	10.09	10.56
TX	0.61	0.64	0.67

Table 1: Estimated percentages of populations immune against COVID-19

It is possible that our estimates, due to the conservative assumptions made, might actually somewhat underestimate the actual levels of immunity. They do, however, show enough variability among the states that they might be of interest as rough estimates of the current levels of immunity against COVID-19.

Acknowledgement. I would like to thank Margaret MacNeil and Nicholas Lesniewski for discussions.

References

- [1] Kermack, W. O., and McKendrick, A. G.: A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. Lond.* **A115**, 700 - 721 (1927).
- [2] Rothan, H. A., and Byrareddy, S. N.: The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, *J. Autoimmunity*, **109**, <https://doi.org/10.1016/j.jaut.2020.102433> (2020).
- [3] Roques, L. et al: Using early data to estimate the actual infection fatality ratio from COVID-19 in France, <https://doi.org/10.1101/2020.03.22.20040915> (2020).
- [4] Russell, T. W. et al: Estimating the infection and case fatality ratio for COVID-19 using age-adjusted data from the outbreak on the Diamond Princess cruise ship, <https://doi.org/10.1101/2020.03.05.20031773> (2020).
- [5] Sanche, S. et al: High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2, *EID Journal*, **26**, <https://doi.org/10.3201/eid2607.200282> (2020).
- [6] Verity, R. et al: Estimates of the severity of COVID-19 disease, <https://doi.org/10.1101/2020.03.09.20033357> (2020).

- [7] Worldometer <https://www.worldometers.info/coronavirus/country/us/>.
- [8] Zhang S. et al: Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis, *Int. J. Infect. Dis.*, **93**, 201- 204 (2020).