

1 ON THE UNCERTAINTY ABOUT HERD IMMUNITY 2 LEVELS REQUIRED TO STOP COVID-19 EPIDEMICS

3 Daniel Gianola^{a,b,c}

^a Department of Animal Sciences, University of Wisconsin-Madison, USA;

^b Department of Dairy Science, University of Wisconsin-Madison, USA;

^c Institut Pasteur de Montevideo, URUGUAY

¹Corresponding author. Email: gianola@ansci.wisc.edu

4 May 31, 2020

5 Abstract

6 COVID-19 evolved into a pandemic in 2020 affecting more than 150 countries. Given
7 the absence of a vaccine, discussion has taken place on the strategy of allowing the
8 virus to spread in a population, to increase population "herd immunity". Knowledge
9 of the minimum proportion of a population required to have recovered from COVID-
10 19 infection in order to attain "herd" immunity, P_{crit} , is important for formulating
11 epidemiological policy. A method for measuring uncertainty about P_{crit} based on a
12 widely used package, EpiEstim, is derived. The procedure is illustrated using data
13 from twelve countries at two early times during the COVID-19 epidemic. It is shown
14 that simple plug-in measures of confidence on estimates of P_{crit} are misleading, but that
15 a full characterization of statistical uncertainty can be derived from EpiEstim, which
16 reports percentiles only. Because of the important levels of uncertainty, it is risky to
17 design epidemiological policy based on guidance provided by a single point estimate.

18 1 Introduction

19 COVID-19 evolved into a pandemic in 2020 affecting more than 150 countries that vary in size,
20 population density, economic resources and public health systems. Given the absence of a vaccine,
21 considerable discussion (e.g., Dowdy and D'Souza, 2020) at various society levels has taken place on

22 a strategy of allowing the virus to spread in a population, to increase population "herd immunity".
23 Sweden's adoption of the approach has been at the center of the debate.

24 Kwok et al. (2020) estimated the minimum proportion of a population required to have re-
25 covered from COVID-19 infection in order to attain "herd" immunity, P_{crit} in various countries.
26 Knowing P_{crit} is important for formulating epidemiological policies tailored to a specific country.
27 The parameter P_{crit} is inferred from an estimate of R_t , the viral reproductive, defined as the num-
28 ber of infections per infector at a given time of the epidemic at time t . The formula relating these
29 two quantities is $P_{crit} = 1 - R_t^{-1}$ (Anderson and May, 1992). Since P_{crit} is a fraction, it must take
30 values between 0 and 1, implying that R_t must be larger than 1 for the expression to be meaning-
31 ful. As R_t increases beyond 1, so does the minimum fraction of the population that needs to be
32 infected. Since R_t must be estimated soon after the epidemic starts, often there is large statistical
33 uncertainty associated with the estimate, especially in countries where there are a few cases at the
34 onset. Because the uncertainty about R_t propagates into P_{crit} it is important to consider it.

35 The objective here is to show how to measure uncertainty about P_{crit} . The procedure is
36 illustrated using data from twelve countries at two early times during the COVID-19 epidemic.

37 **2 Inference of P_{crit}**

38 R_t is estimated from a time series containing the number of infections I_0, I_1, \dots, I_t (t denotes time)
39 and from information on the statistical distribution of the virus serial interval, defined as the
40 number of days mediating between the appearance of symptoms in pairs of infectors and infected
41 persons. Nishiura et al. (2020) used data on 28 pairs of patients in the context of SARS-CoV-2
42 infections and found that a log-normal distribution with mean and standard deviation of 4.7 and
43 2.9 days, respectively, provided the best fit.

44 Table 1 of Kwok et al. (2020) displays estimates of R_t and of P_{crit} obtained at early stages of
45 the COVID-19 epidemic for various countries. It does not report statistical uncertainty associated
46 with the latter, although confidence intervals are provided for R_t . The confidence intervals can be
47 wide, especially for countries with smaller number of cases in the time series. For example, in the
48 table, the width of the 95% confidence interval on R_t for the USA is 0.28 whereas for Slovenia it
49 is 3.47. A rough confidence interval for P_{crit} can be constructed by a plug-in method using values
50 of the bands for R_t , but this way of proceeding may be misleading.

51 A suitable measure of the statistical uncertainty associated with inferring P_{crit} is obtained using
52 a Bayesian approach such as the one employed in the R package **EpiEstim** described in Cori et
53 al. (2013), but with some minor external (to the program) modifications. Briefly, the package
54 employs a model that assumes that the number of infections observed at time t follows a Poisson
55 distribution with parameter $R_t \Lambda_t$. Here, Λ_t is a weighted average of past infections; weights are
56 calculated from knowledge (or estimation) of the serial interval distribution. If, R_t is assigned a

57 $\text{Gamma}(a, b)$ prior distribution, the posterior is also a $\text{Gamma}(a_t, b_t)$ distribution where a_t and b_t
 58 depend on a and b , on the number of past infections and on the distribution of the serial interval
 59 (Cori et al., 2013) .

60 Since R_t has a $\text{Gamma}(a_t, b_t)$ posterior distribution, then R_t^{-1} has an inverse Gamma distrib-
 61 ution with parameters (a_t, b_t) (Sorensen and Gianola, 2002; Gelman et al., 2014) with density

$$62 \quad f(R_t^{-1} | I_0, I_1, \dots, I_t, \text{serial interval distribution}) \propto (R_t^{-1})^{-a_t-1} \exp\left(-\frac{b_t}{R_t^{-1}}\right). \quad (1)$$

63 Since $R_t^{-1} = 1 - P_{crit}$ is a linear relationship, the posterior distribution of P_{crit} has density

$$64 \quad \begin{aligned} & f(P_{crit} | I_0, I_1, \dots, I_t, \text{serial interval distribution}) \\ 65 \quad &= \frac{(1 - P_{crit})^{-a_t-1} \exp\left(-\frac{b_t}{1 - P_{crit}}\right)}{\int_0^1 (1 - P_{crit})^{-a_t-1} \exp\left(-\frac{b_t}{1 - P_{crit}}\right) dP_{crit}}; 0 < P_{crit} < 1. \end{aligned} \quad (2)$$

66 For instance, if $a_t = 164$ and $b_t = 65$, the posterior density of P_{crit} is

$$67 \quad f(P_{crit} | I_0, I_1, \dots, I_t, \text{serial interval distribution}) = \frac{(1 - P_{crit})^{-(164+1)} \exp\left(-\frac{65}{1 - P_{crit}}\right)}{9.6426 \times 10^{-7}}, \quad (3)$$

68 and takes the form depicted in Figure 1. The posterior expectation of R_t is $\frac{164}{65} \approx 2.52$ so a plug-in
 69 estimate would give $1 - \frac{1}{2.52} \approx 0.60$ as an approximation to the posterior mean of P_{crit} , implying
 70 that a minimum 60% of the population needs to be infected in order to attain herd immunity. The
 71 posterior mean and standard deviation of P_{crit} are 0.60 and 0.03, respectively, and the coefficient of
 72 variation is 5.2%. The posterior probability that P_{crit} is larger than 0.60 is 53.6%.; the probability
 73 that it is smaller than 0.55 is 5.9%, and the probability that it takes values between 0.55 and
 74 0.60 is about 40.5%. Assuming the case fatality rate is 1%, in a country with a population of
 75 10 million persons (e.g., Azerbaijan or Sweden), the "road" to immunity would produce about
 76 55,000 expected deaths if $P_{crit} = 0.55$, versus 62,000 if it were $P_{crit} = 0.62$, a difference of 7,000
 77 person dying. The larger the uncertainty about P_{crit} , the larger the risk associated with choosing
 78 a strategy for control of the epidemic.

79 **3 Application to the COVID-19 pandemic**

80 Data were downloaded from the European Centre for Disease Prevention and Control, an European
 81 Union organization based in Sweden. Twelve countries were chosen and organized into two groups

82 as shown in Table 1. Estimates of R_t (posterior median and 2.5 and 97.5 percentiles of the posterior
83 distribution) were obtained using **EpiEstim** at times $t = 8, 14$ from the first case. Simple "plug-in"
84 estimates of P_{crit} and of percentiles were derived from the relationship between the latter and R_t .
85 In some cases (Brasil, Uruguay, New Zealand and USA) some percentiles could not be calculated
86 because R_t was smaller than 1, producing "plug-in" estimates of P_{crit} outside of its parameter
87 space. Inter-percentile ranges were wide in many instances. For example, at $t = 8$ (the earliest
88 time at which **EpiEstim** produced an estimate of R_t , the range was 35% in Argentina, 31% in
89 Cuba, 37% in Denmark and 22% in Israel; at $t = 14$ the range was still wide for Israel, at 35%.
90 For Sweden, the ranges were 21% and 8% at times $t = 8$ and $t = 14$, respectively. The uncertainty
91 about P_{crit} was large for many countries.

92 Instead of using the crude "plug-in" method, which can yield estimates outside of the $0 - 1$
93 space for a proportion, **EpiEstim** can be "tricked" to produce the entire posterior distribution.
94 At any t , **EpiEstim** returns the mean and standard deviation (sd) of the posterior distribution.
95 If the parameterization is $Gamma(shape = a, rate = b)$, then $a = \left(\frac{mean}{sd}\right)^2$ and $b = \frac{mean}{sd^2}$ from
96 which (2) can be numerically evaluated. Alternatively, the posterior distribution of P_{crit} can be
97 estimated by drawing random numbers from a computer as follows: 1) sample a large number of
98 $Gamma(a, b)$ deviates. 2) For each draw, form $1 - \frac{1}{sampled\ value}$, and accept it as draw from
99 the posterior distribution of P_{crit} if it resides in the $(0, 1)$ interval; discard it otherwise.

100 The posterior densities for Group A and B countries are in Figures 2 and 3, respectively. The
101 "plug-in" measures of uncertainty for P_{crit} derived from the percentiles of the posterior distribution
102 of R_t were often misleading or useless. For example, for Argentina (Figure 2, curve in red) the
103 naive lower bounds for $t = 8$ does not inform that P_{crit} should probably be at least 50% to attain
104 herd immunity. For Israel (Figure 3) the naive bounds exaggerate uncertainty as the posterior
105 densities are reasonably sharp, both at $t = 8$ and $t = 14$. In Uruguay (Figure 2), the low number
106 of cases observed translated into low values of R_t , especially at $t = 14$ but suggested that herd
107 immunity could be reached with a low proportion of the population infected, which would perhaps
108 take a considerable amount of time but at a low cost in terms of number of dead people. However,
109 at $t = 8$ the analysis suggested that P_{crit} would need to be around 50 – 60%. For Sweden (Figure
110 3), it appears that P_{crit} would need to be at least 55%; for the USA (Figure 3), the uncertainty
111 was enormous, as suggested by the analyses conducted at $t = 8$ and $t = 14$.

112 4 Conclusion

113 Attaining herd immunity requires a minimum proportion of the population to be infected. Such
114 proportion can be inferred from estimates of the virus reproducing number calculated at early
115 stage of the epidemic. Plug-in measures of confidence on estimates were misleading, but a full
116 characterization of statistical uncertainty was derived from publicly available software. Because of

117 the important levels of uncertainty, it is extremely risky to design epidemiological policy based on
118 the guidance provided by a single point estimate, as in Kwok et al. (2020).

119 5 References

120 Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-
121 varying reproduction numbers during epidemics. *American Journal of Epidemiology* 2013 178:
122 1505-1512.

123 Dowdy D, D'Souza G. Early Herd Immunity against COVID-19: a dangerous misconcep-
124 tion. 2020. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
125 <https://coronavirus.jhu.edu/from-our-experts/early-herd-immunity-against-covid-19-a-dangerous-misconception>

126 Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*
127 (3rd Ed) 2014. CRC Press, Boca Raton.

128 Kwok Ko, Lai F, Wei WI, Wong SYS, Tang JWT. Herd immunity- estimating the level required
129 to halt the COVI-19 epidemics in affected countries. *J Infect* 2020 doi.org/10.1016/j.inf.2020.03.027

130 Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19)
131 infections. *International Journal of Infectious Diseases* 2020 93 284-286.

132 Sorensen D, Gianola D. *Likelihood, Bayesian and MCMC methods in quantitative genetics* 2002.
133 Springer, New York.

Table 1. Estimates of the median, 2.5 and 97.5 percentiles of the posterior distribution of $R_{[t]}$ (virus reproducing number) at days $t=8, 14$ of the COVID-19 pandemic in twelve countries. The “plug-in” P_{crit} (minimum proportion of the population that needs to be infected in order to reach herd immunity) estimates are directly obtained from the relationship $P_{crit} = 1 - R_{[t]}^{-1}$.

	Median $R_{[8]}$ (2.5%_[8], 97.5%_[8])	$P_{crit[8]}$ (2.5%_[8], 97.5%_[8])	Median $R_{[14]}$ (2.5%_[14], 97.5%_[14])	$P_{crit[14]}$ (2.5%_[14], 97.5%_[14])
GROUP A				
<i>Argentina</i>	3.31 (2.31,4.56)	0.70 (0.43,0.78)	2.11 (1.72,2.56)	0.53 (0.42,0.61)
<i>Brasil</i>	1.14 (0.16,3.79)	0.12 (<i>ND</i> ¹ ,0.74)	3.18 {2.06,4.63}	0.69 (0.54,0.78)
<i>Chile</i>	3.27 (2.11,4.81)	0.69 (0.53,0.79)	4.18 (3.61,4.82)	0.76 (0.72,0.79)
<i>Cuba</i>	2.70 (1.74,3.96)	0.63 (0.43,0.74)	1.81 (1.39,2.31)	0.45 (0.28,0.57)
<i>Peru</i>	3.31 (2.41,4.40)	0.70 (0.58,0.77)	2.73 (2.39,3.10)	0.63 (0.58,0.68)
<i>Uruguay</i>	2.58 (2.16,3.05)	0.61 (0.54,0.67)	1.03 (0.86,1.22)	0.03 (<i>ND</i> ¹ ,0.18)
GROUP B				
<i>Denmark</i>	3.07 (1.52,3.43)	0.67 (0.34,0.71)	10.30 (9.09,11.61)	0.90 (0.89,0.91)
<i>Germany</i>	3.13 (2.96,3.61)	0.68 (0.66,0.72)	2.39 (2.25,2.55)	0.58 (0.56,0.61)
<i>Israel</i>	3.69 (1.56,7.24)	0.73 (0.36,0.86)	2.46 (1.59,3.62)	0.59 (0.37,0.72)
<i>N. Zealand</i>	2.65 (0.92,5.81)	0.62 (<i>ND</i> ¹ ,0.83)	3.83 (2.86,5.00)	0.74 (0.65,0.80)
<i>Sweden</i>	3.30 (2.31,4.54)	0.70 (0.57,0.78)	2.84 (2.53,3.17)	0.65 (0.60,0.68)
<i>USA</i>	2.85 (0.99,6.24)	0.65 (<i>ND</i> ¹ ,0.84)	1.53 (0.65,3.00)	0.35 (<i>ND</i> ¹ ,0.67)

¹*ND*: not defined because $R < 1$.

Figure 1. Posterior density of $P(\text{crit})$ for a hypothetical example. $E[P(\text{crit})]=0.60$ and the posterior coefficient of variation is 5.2%.

Figure 2. Posterior densities of $P(\text{crit})$ at times=8 and 14 of the COVID19 pandemic for Argentina, Brasil, Chile, Cuba, Peru and Uruguay

Figure 3. Posterior densities of $P(\text{crit})$ at times=8 and 14 of the COVID19 pandemic for Denmark, Germany, Israel, New Zealand, Sweden and USA

Density





