The COVID-19 herd immunity threshold is not low: A re-analysis of European data from spring of 2020

Spencer J. Fox¹, Pratyush Potu², Michael Lachmann³, Ravi Srinivasan⁴, Lauren Ancel Meyers¹,³

1 - Department of Integrative Biology, The University of Texas at Austin
2 - Department of Mathematics, The University of Texas at Austin
3 - The Santa Fe Institute
4 - Applied Research Laboratories, The University of Texas at Austin

Commentary

The time course and burden of the COVID-19 pandemic will depend on the herd immunity threshold (HIT) of the virus, which is the fraction of the population that needs to be immunized for an epidemic to slow in the absence of mitigation efforts. Estimates for the COVID-19 HIT range from 6% to over 60%¹,². Given that roughly 10% of the global population has been infected³, the low end of this range implies that the pandemic should soon burn out on its own, while the high end paints a grim picture of future morbidity and mortality, in the absence of pervasive non-pharmacological interventions, efficacious vaccines, or life-saving drugs.

The recent publication of the Great Barrington Declaration (GBD), which calls for relaxing all public health interventions on young, healthy individuals, has brought the question of herd immunity to the forefront of COVID-19 policy discussions⁴,⁵. The authors state that “immunity in the population is playing a substantial role in controlling the spread,” tacitly referencing preprints by multiple GBD authors that posit HITs of 10-20%²,⁶. Evidence against this claim is mounting, including pandemic resurgences throughout Europe and the US and attack rates exceeding 50% in the hardest hit regions and congregate living settings⁷–⁹.
Given that this unpublished work is fundamentally shaping public discourse and global policy, reconciling its claims with the rapidly evolving state of the pandemic is paramount. To this end, we reevaluated the core model from the study and have identified a fundamental flaw that leads to underestimation of the COVID-19 HIT. The authors sought to identify the cause of the summer slowdown in four European countries by fitting an SEIR-like model of COVID-19 transmission to case count data up to July of 2020. The analysis is structured so that one of two explanations are possible. Either the pandemic is self-limiting (i.e., the HIT is low) or social distancing and other community mitigation efforts slowed transmission. However, teasing apart the contributions of these factors from the case data alone is statistically impossible. In other words, one cannot estimate the HIT without making assumptions about the efficacy of community mitigation, and vice versa (See supplement).

So the researchers made a strong assumption about community mitigation efforts in Europe in the spring and summer of 2020. Roughly, they assume that Europe locked down throughout April and then returned to normal (linearly) by the end of August (Figure 1A - Blue). By assuming that interventions disappear steeply, the model concludes that the pandemic must be fading due to immunity buildup, and thus estimates low HITs. As it turns out, the derived HIT is highly sensitive to the assumed timeline of mitigation (Figure S2-S3) and we have good reason to believe their assumption is flawed. The authors use mobility traces to justify their pattern\(^2\), but other precautionary policies like school closures, wearing of face coverings, and social distancing have likely kept transmission repressed far below the pre-April baseline (Figure 1A - Black). When we plug this plausible scenario into the Aguas et al. model\(^2\) (Figure 1A - Green), the COVID-19 HIT estimate increases six-fold for Belgium, three-fold for England, ten-fold for Portugal, and six-fold for Spain (Figure 1B). A range of alternative scenarios produce similar estimates (Figure S2-S3). If policymakers were to adopt a herd immunity strategy, in which the virus is allowed to spread relatively unimpeded, we project that cumulative COVID-19 deaths would total almost 650,000 (95% CI: 500,000 - 780,000) across all four countries.
through the end of the pandemic under the revised HIT estimates, roughly five-fold higher that
projected under the original low HIT estimates (Figure S4).

The fragility of the Aguas et al. study\textsuperscript{2} undermines a key premise of GBD and other
recent calls for “herd immunity” strategies. To their credit, the authors clearly demonstrate that
population heterogeneity in susceptibility to infection can dramatically lower the herd immunity
threshold\textsuperscript{10}. However, their model can only disentangle the impacts of heterogeneity versus
interventions on COVID-19 transmission when approached with sufficient data and validated
assumptions. Our rough, but arguably more plausible, re-estimates of the COVID-19 HIT
corroborate strong signals in the data and compelling arguments that most of the globe remains
far from herd immunity. Moreover, abandoning community mitigation efforts would jeopardize
the welfare of communities and integrity of healthcare systems.

\textbf{Figure 1: Re-estimation of the COVID-19 herd immunity thresholds (HIT) in four European
countries, using the model of Aguas et al.\textsuperscript{2}.} (A) Strength of non-pharmacological
interventions indicated by a government response index\textsuperscript{11} (black) compared to the trends
assumed by Aguas et al.\textsuperscript{2} (blue) and \textit{plausible} alternatives (green) derived to match the
observed indices. The time-series are scaled for comparison, where values of zero and one
correspond to lack of mitigation efforts and the average level of community mitigation in April, respectively. (B) Estimated HIT using the Aguas et al.\(^2\) approach, under the original assumption that non-pharmacological efforts rapidly decrease to baseline (blue) and alternative assumption that measures mirror the government response index (green).

**Acknowledgements**

The authors thank the authors of Aguas et al.\(^2\), for sharing their research code and data, which allowed us to replicate and extend their analyses. We acknowledge the UT COVID-19 Modeling Consortium for helpful discussion and comments, and the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper. URL: [http://www.tacc.utexas.edu](http://www.tacc.utexas.edu). We also acknowledge support from a CDC COVID-19 supplementary grant (U01IP001136-01-01).

**Author Contributions**

SJF, ML, and LAM conceived of the initial idea. PP carried out all analyses. RS and ML developed the analytical framework for statistical identifiability. SJF and LAM supervised the research. SJF and LAM wrote the initial draft. All authors significantly contributed to the manuscript and approved the final draft. PP, SJF, and RS verified the underlying data.

**Declaration of interests**

The authors declare no conflicts of interest.

**References**


Keaten J. WHO: 10% of world’s people may have been infected with virus. Associated Press. 2020; published online Oct 5. https://apnews.com/article/virus-outbreak-archive-united-nations-54a3a5869c9ae4ee623497691e796083 (accessed Nov 10, 2020).


Gomes MGM, Corder RM, King JG, et al. Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. medRxiv 2020; published online May 2. DOI:10.1101/2020.04.27.20081893.

Supplementary information for “The COVID-19 herd immunity threshold is not low: A re-analysis of European data from spring of 2020”

Spencer J. Fox Pratyush Potu Michael Lachmann
Ravi Srinivasan Lauren Ancel Meyers

November 11, 2020

1 Overview
The low COVID-19 herd immunity thresholds (HITs) of 6-21% estimated in Aguas et al. [1] are inconsistent with other model-derived estimates [5,9,2] and seroprevalence-based estimates from some of the hardest hit regions around the world [3]. To explain this apparent discrepancy, we conducted a thorough review of their methods and code. We identified key assumptions about the timing and extent of community mitigation efforts that shift the COVID-19 HIT estimates downward. As their code was made openly available (https://github.com/mmgomes1/covid), we apply their exact model structure and fitting procedure to evaluate the sensitivity of the HIT estimates to these assumptions.

Below, we detail our: (1) slight modification to their model fitting procedure, (2) sensitivity analyses with respect to the assumed mitigation curves (i.e., timing and magnitude of transmission reduction via non-pharmacological interventions), (3) derivation of alternate plausible mitigation curves from the Oxford COVID-19 Government Response Tracker response index data, (4) long-range COVID-19 mortality projections depending on the estimate HIT, and, finally (5) mathematical argument regarding the statistical non-identifiability the model (i.e., inability to simultaneously estimate the impact of community mitigation and population heterogeneity). The code used for the primary analyses along with the data can be accessed here: https://github.com/pratyush16/VariationalSusceptibility

2 Modification to the Aguas et al. model
We slightly modified the modeling framework of Aguas et al. [1] to include an additional month of COVID-19 incidence data (through August 7th, 2020) for Belgium (https://epistat.wiv-isp.be/covid/) and England (https://coronavirus.data.gov.uk/details/about-data). For Portugal and Spain we
analyzed data through July 10th, 2020 as in [1]. We made this adjustment to address a strong tendency of the model to estimate a low HIT without evidence of increased transmission following intense mitigation during the initial stay at home orders.

3 Sensitivity of COVID-19 HIT estimates to mitigation curves

We analyzed the sensitivity of the HIT estimates from Aguas et al. to variations in the assumed temporal progression of the mitigation curves within their modeling framework. Specifically they assume a general shape of mitigation where mitigation begins, increases until it reaches a maximum level, remains at the maximum for some time, and then slowly returns to a baseline of no mitigation. In this shape there are five key parameters governing the progression: (1) the time that mitigation begins, (2) the time it takes for mitigation to take full effect, (3) the maximum impact of mitigation, (4) the duration that mitigation remains at maximum, and (5) the time for mitigation to be completely removed.

The modeling framework fits the timing that mitigation begins and the maximum impact it will have, but makes assumptions about the remaining parameters. Their framework assumes that once mitigation begins it will take 21 days to reach a maximum level, will remain there for 30 days, and will subsequently return to the original baseline level after 120 days. They test sensitivity to slight deviations of the time to return to baseline and find minimal change in HIT estimates looking at 150 or 180 days. As noted in the manuscript and seen in Figure 1, we believed their assumptions about mitigation progression were driving their low HIT estimates, so we tested the sensitivity of the low HIT estimates to a wide-range of values for the maximum mitigation duration (Figure S2) and times to return back to baseline (Figure S3). It’s clear that the model is sensitive to the assumed shape of mitigation, with HIT estimates ranging from almost 5% to 85% depending on the country and assumed shape.

It is clear that the estimated HITs are extremely sensitive to these assumptions, and that there are many combinations of heterogeneity and mitigation progression that can give similar fits (described in the next section), so it is extremely important that the assumed mitigation curves match reality. While Aguas et al. matched their mitigation shapes to mobility data, these data are not an accurate picture for total transmission mitigation, as decoupling of mobility and transmission has been previously noted as populations adopt precautionary behavior like mask wearing and social distancing [12].

Instead, we focused on the government response index developed by researchers at the University of Oxford [7]. This index captures 18 policy indicators widely implemented around the world, and produces an overall value that captures the strength of government responses through time as shown in Figure 1A. We found that the Aguas et al. baseline assumptions did not match the actual progression of mitigation policies, and instead chose to fit the esti-
Figure S1: Country-specific case counts used for fitting alongside daily estimates of the effective reproduction number. Case counts in black indicate data used that are the same as Aguas et al, and light grey indicates the additional data we included in our fitting procedure as described in the modifications section of the supplement.
Figure S2: Sensitivity of herd immunity threshold (HIT) estimates to assumptions about the duration at maximum mitigation. Vertical black dashed line indicates the assumed value in Aguas et al. All estimates are made assuming the baseline value of time for mitigation to be completely removed of 120 days.

Response index. We assumed the baseline assumptions of 21 days to implement mitigation and 30 days at maximum mitigation impact, and fit the timing to completely remove mitigation based on the trends in the government response index. Fitted estimates for the time to return to normal can be found in Figure S3 for each country, and are almost an order of magnitude larger than assumed in Aguas et al (120): Belgium (828 days), England (1,130 days), Portugal (981 days), and Spain (1,033 days). While the government response index also likely doesn’t capture the full progression of community mitigation, because it doesn’t capture adherence to policies, we feel that it likely captures general countrywide trends better than mobility data alone.

4 Projecting mortality under a herd immunity policy

We calculated the final epidemic size for an uncontrolled epidemic based on the method described in [6] for each country. We carried out the same fitting
Figure S3: Sensitivity of herd immunity threshold (HIT) estimates to assumptions about the time for mitigation to be completely removed. Colored solid lines show sensitivity across all four European countries. Vertical black dashed line shows the assumed value in Aguas et al, and vertical colored lines show the fitted value from government response index used in our main analysis for each of the specific countries. All estimates are made assuming the baseline value of maximum mitigation duration of 30 days.

procedure as Aguas et al. using either their assumed mitigation curve or our revised version, we then calculated the mortality cost of following a “herd immunity strategy” using the estimated parameters from each European country. For each European country we used the estimated final epidemic size estimated for that scenario as the total infection count and an infection fatality rate of 0.68% (95% CI: 0.53%-0.82%) [10]. For the US, we assumed average final epidemic sizes based on the average of the estimates for the four countries for each assumed mitigation curve scenario, which suggests final epidemic sizes of 13.9% and 79.2% for the Aguas et al. and alternative scenarios respectively. Estimates for both scenarios can be seen in Figure S4. Country population sizes were assumed to be 56,286,961 (England), 11,607,113 (Belgium), 10,186,314 (Portugal), 46,761,086 (Spain), and 331,674,530 (United States) as estimated from UN population statistics made available through https://www.worldometers.info/.
5 Non-identifiability of the model

The model fitting procedure from Aguas et al. explains transmission dynamics between mitigation and population heterogeneity, which impacts herd immunity thresholds. In Aguas et al., they begin with assumed mitigation curves and estimate the herd immunity threshold. Here we describe mathematically why their estimation procedure is only identifiable with strong assumptions about the shape of the mitigation curves. We first break down the model structure into component parts to make clear the tension between mitigation and herd immunity. Then we use this analytical framework to flip their estimation procedure around to show how one can estimate mitigation curves that fit epidemic trajectories for any assumed herd immunity threshold.

To start, consider an SEIR model with variable susceptibility $S_t(x)$ and time-dependent mitigation $M_t$ due to non-pharmaceutical interventions (NPIs), as given in Aguas et al.:
\[
\begin{align*}
\dot{S}(x) &= -x\lambda S(x)M \\
\dot{E}(x) &= x\lambda S(x)M - \delta E(x) \\
\dot{I}(x) &= \delta E(x) - \gamma I(x) \\
\dot{R}(x) &= \gamma I(x)
\end{align*}
\]

Then, \( R_0 = \frac{\beta}{\gamma}, \quad R_t = \frac{\beta}{\gamma} \bar{S}_t M_t = R_0 \bar{S}_t M_t, \)

where \( R_0 \) is the basic reproduction number, \( R_t = R_{\text{eff}}(t) \) is the effective reproduction number at time \( t \), and \( \bar{S}_t = \int x' S_t(x') dx' \) is the mean susceptibility at time \( t \). More generally, we can consider time-dependent effective reproduction numbers of the form

\[
R_t = R_0 F_t(S) M_t, \tag{1}
\]

where \( F_t(S) \) is some functional of the variable susceptibility \( S_t(x) \) as a function of \( x \) (notice that the functional itself does not change over time, but that the resulting function has a time dependence since \( S_t(x) \) does). The model of Aguas et al. corresponds to the case when \( F_t(S) = \bar{S}_t \). From this equation it is already clear that given a time series \( R_t \), we can for every \( t \) only infer the combined impact of heterogeneity and mitigation (the product \( F_t(S) M_t \)), but not each separately.

Now consider \( \tilde{R}_t = R_t / M_t = R_0 F_t(S) \), the effective reproduction number if there was no intervention at all (i.e., if \( M_t \equiv 1 \)). Herd immunity is reached at \( T = \min \left\{ t' : \tilde{R}_t \leq 1 \text{ for all } t \geq t' \right\} \) (i.e., the first time after which \( \tilde{R}_t \) remains at or below 1), and the corresponding herd immunity threshold is

\[
\text{HIT} = 1 - S_T,
\]

where \( S_t = \int S_t(x') dx' \).

**Epidemic dynamics with Gamma-distributed variable susceptibility**

Following a similar derivation as that in Montalban et al. [11], it can be shown that

\[
S_t(x) = S_0(x) e^{-x \cdot k_t}
\]

where \( k_t = \int_0^t \lambda_s M_s ds \) and \( \lambda_t \) is the force of infection at time \( t \). Assuming \( S_0(x) \) is a Gamma\((a,a)\) density (where \( a \) is related to the coefficient of variation by \( \text{CV} = 1/\sqrt{a} \)), we have that

\[
S_t(x) = \left( \frac{a}{a + k_t} \right)^a \Gamma_{a,a+k_t}(x).
\]

7
Taking integrals, it can be shown that

\[ S_t = \left( \frac{a}{a + k_t} \right)^a \]

and

\[ \bar{S}_t = \left( \frac{a}{a + k_t} \right)^{a+1} = (S_t)^{\frac{a+1}{a}}. \]

Using the previous formula, we have that \( S_t = (\bar{S}_t)^{\frac{a}{a+1}}. \) Since \( \bar{R}_t = R_t / M_t = 1, \) we have that \( R_0 \bar{S}_T = 1 \) and

\[ \text{HIT} = 1 - (\bar{S}_T)^{\frac{a}{a+1}} = 1 - R_0^{-\frac{a}{a+1}}. \]

### Deriving mitigation curves from herd immunity thresholds

We now show how it is possible to derive a mitigation curve \( M_t \) that is consistent with the dynamics under Gamma-distributed variable susceptibility for a given herd immunity threshold.

To begin, let us reparameterize time in terms of the proportion of susceptibles \( S_t \) to make our calculations simpler. First, Note that \( S_t \) is strictly monotone decreasing so long as \( \lambda_t M_t > 0 \) for all \( t \) (i.e., both the force of infection and mitigation curve are strictly positive), and that \( S_0 = 1 \) and \( S_t \to 0 \) as \( t \to \infty \).

Let \( \sigma_t = 1 - S_t \). Then, we can reparameterize time by \( t \mapsto \sigma \) (i.e., \([0, \infty) \mapsto [0, 1]) \). That is, Equation (1) becomes

\[ R_{\sigma} = R_0 F_\sigma M_{\sigma}. \]

Assume that \( F_t \) and \( R_t \) are given. Then \( M_{\sigma} \) can be calculated as

\[ M_{\sigma} = \frac{R_{\sigma}}{R_0 F_\sigma}. \]

Let us write script letters for log of the values:

\[ \mathcal{M}_{\sigma} = \log(M_{\sigma}) = \log(R_{\sigma}) - \log(R_0) - \log(F_\sigma) = R_{\sigma} - R_0 - \mathcal{F}_{\sigma}. \]

Since \( F_t(S) = (S_t)^{(a+1)/a} \), we have that \( F_\sigma = (1 - \sigma)^{(a+1)/a} \). Therefore, \( \mathcal{F}_{\sigma} = \frac{a+1}{a} \log(1 - \sigma) \) and

\[ \mathcal{M}_{\sigma} = R_{\sigma} - R_0 - \frac{a + 1}{a} \log(1 - \sigma). \]

As before, herd immunity is reached at the value \( S_* \) of \( S \) at which no growth occurs in the absence of intervention, i.e., \( R_t / M_t = R_* / M_* = 1 \). That is,

\[ 1 = R_0 F_*(S_*), \]

and therefore \( F_*(S_*) = R_0^{-1} \). Let \( \sigma_* = 1 - S_* \). Hit. Since \( F(S) = S^{(a+1)/a} \), this implies

\[ (1 - \sigma_*)^{\frac{a+1}{a}} = \frac{1}{R_0}. \]
Taking logarithms, we find that
\[
\frac{a + 1}{a} = \frac{-R_0}{\log(1 - \sigma_s)}.
\]
Finally, substituting this into the expression above we get
\[
M_\sigma = R_\sigma - R_0 + \frac{R_0}{\log(1 - \sigma_s)} \log(1 - \sigma)
= R_0 \left( \frac{\log(1 - \sigma)}{\log(1 - \sigma_s)} - \left(1 - \frac{R_\sigma}{R_0} \right) \right).
\]
Inverting the mapping \( t \mapsto \sigma \), we can again write this as a function of time \( t \):
\[
M_t = R_0 \left( \frac{\log(S_t)}{\log(S_s)} - \left(1 - \frac{R_t}{R_0} \right) \right).
\]
Exponentiating both sides, we finally arrive at the expression
\[
M_t = \exp \left[ \log(R_0) \left( \frac{\log(S_t)}{\log(S_s)} - \left(1 - \frac{\log(R_t)}{\log(R_0)} \right) \right) \right] = R_t \cdot R_0^{\frac{-\log(R_t)}{\log(R_0)}}.
\]
This gives us a formula for calculating a mitigation curve \( M_t \) that is consistent with the underlying variable susceptibility model of Aguas et al., for any chosen value of a herd immunity threshold.

**Computing mitigation curves as a function of herd immunity threshold**

We can see from the previous section that the model is attempting to match the dynamics seen in the data through a combination of mitigation and herd immunity as impacted through population heterogeneity. To further show why strong mitigation assumptions are needed to estimate the herd immunity thresholds, we show how one can craft mitigation curves to match dynamics for a wide range of herd immunity thresholds for each of the four countries investigated (Belgium, England, Portugal, and Spain).

Using Equation (2), we can estimate daily mitigation impacts for a given herd immunity threshold (HIT) given the necessary data. To do so, one only needs estimates for the basic reproduction number \( R_0 \), daily effective reproduction number \( R_t \), and the proportion of the population still susceptible to the disease \( S_t \). For each of the four countries we first estimated \( R_t \) using the available case data and a common method for measuring the instantaneous reproduction number assuming a mean serial interval of 5.8 days and standard deviation of 4.48 days [4, 8]. We took the largest value of \( R_t \) as our rough estimate for the \( R_0 \). Finally, we used the case data and the estimated country-specific reporting
rates estimated in Aguas et al. to estimate the true incidence of disease each day, which we converted to $S_t$ using $S_t = S_{t-1} - C_t/\rho$. Here, $C_t$ is the number of the reported cases in that country at time $t$, and $\rho$ is the country-specific reporting rate from Aguas et al: Belgium (0.06), Portugal (0.09), Spain (0.059), England (0.024).

In Figure S5 we show the calculated mitigation curves for each country across a range of potential HITs from 20% to 70%.

References


Figure S5: Using cases data and $R_t$ estimation we calculate $M_t$, the NPI transmission multiplier timelines that would be inferred to have a predetermined HIT.