Mathematical modeling suggests pre-existing immunity to SARS-CoV-2.

Marcus Carlsson, Gad Hatem, Cecilia Söderberg-Nauclér

Abstract

Mathematical models have largely failed to predict the unfolding of the COVID-19 pandemic. We revisit several variants of the SEIR-model and investigate various adjustments to the model in order to achieve output consistent with measured data in Manaus, India and Stockholm. In particular, Stockholm is interesting due to the almost constant NPI's, which substantially simplifies the mathematical modeling. Analyzing mobility data for Stockholm, we argue that neither behavioral changes, age and activity stratification nor NPI's alone are sufficient to explain the observed pandemic progression. We find that the most plausible hypothesis is that a large portion of the population, between 40 to 65 percent, have some protection against infection with the original variant of SARS-CoV-2.

1 Why do SEIR-models fail, or do they?

For most countries or regions, when observing the time series of new cases, hospitalizations or deaths by COVID-19, it is clear that they go up and down in a manner that can only be explained by Non-Pharmaceutical Interventions (such as lock-downs, school closures etc.), seasonal and behavioral changes. However, in a few places where the spread got out of control, such as Manaus and New York, the curves look very much like the "wave of infections" predicted by standard mathematical models for infectious diseases, called SEIR. In particular this applies to Stockholm where the authorities have been unwilling to enforce any major NPI's, and the recommendations (as well as other key parameters such as season) have been virtually constant, especially during the second wave.

Despite this, the measured sero-prevalence in all to us known locations (also towns such as Bergamo [49]) is well below the mathematically predicted herd-immunity threshold, which is estimated to somewhere above 70%. Manaus reached a maximum sero-prevalence of 52% in June that waned to below 30% by October [13], yet despite the society being relatively open it was not hit by a second wave until late December, most likely caused by a mutant virus variation [4]. The sero-prevalence in New York after the first wave was 23% [6] and a recent measurement from Stockholm indicates that 22% of the population have had the virus, past the second wave and one year into the pandemic.

Is it possible that this discrepancy between the SEIR model and reality is only due to variations in the effective R-value caused by changing NPI's, or is the SEIR-model simply inapt for modeling COVID-19? Or could we refine the SEIR-model to give more realistic output? Or could it be that the parameters we feed the model are wrong? This is the question we tackle in this paper, whose supplementary material contains a thorough revision of heterogeneous SEIR models and their parameters.

We argue that the first three explanations are less credible, and then revisit the idea of a pre-existing protective immunity to SARS-CoV-2, or pre-immunity for short. We do not address the nature of this protection, which either can stem from pre-existing cross-reactive adaptive immunity or different degrees of innate immunity, but we mention the publications [16, 40] showing different mechanisms where prior exposure to other viruses can prevent infection by SARS-CoV-2. We show that upon taking pre-immunity into account, it is possible to accurately model both the first and second wave of COVID-19 in Manaus and Stockholm. This is by no means a proof that pre-immunity does exist, but it is our hope that this work can help improve the understanding of this new virus and stimulate medical research aiming to further explore potential causes for pre-immunity.

2 A close look at the SEIR-setup

The mathematical models recommended for respiratory diseases like influenza are (refinements of) a basic ordinary differential equation commonly known as SEIR, (see e.g. Ch. 4 and 9, [10]). While these deterministic models do not capture the random nature of real life, it has been shown that more advanced stochastic models behave like these deterministic models when there is a major outbreak in a large population (see e.g. [3] Ch. 5.5 and [7]).

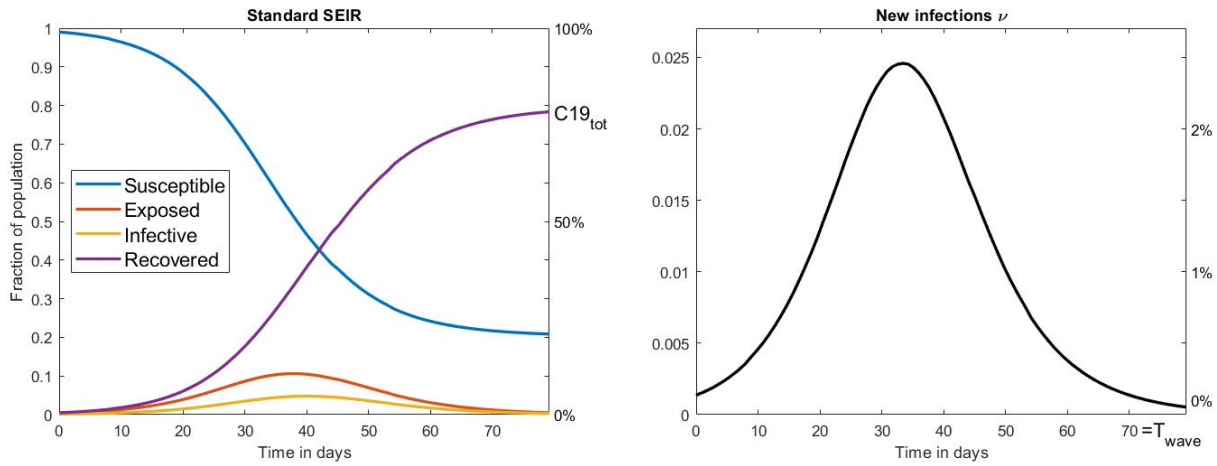

Figure 1: Left: SEIR outcome with $R_0 = 2$. Right: New infections ν.

In the acronym SEIR, $S = S(t)$ stands for the amount of susceptible at time t in a population of N individuals, and $s(t)$ is the corresponding fraction of the population in group S, so $s(t) = S(t)/N$ equals 1 at the beginning of the pandemic (unless there is pre-existing immunity, easily included by simply choosing an initial value less than one). $E(t)$ stands for exposed, to account for the incubation time. $I(t)$ stands for infectious, after which people recover and "appear" in $R(t)$ instead. Like s, the letters e, i, r will denote the respective fractions of the population, so if $i(t) = 0.02$ it means that 2% of the population is infective on the particular day t. The majority of individuals will thus start in s and end in r, but not all. The final value of r is called the final size of the epidemic, usually a bit higher than the herd-immunity threshold because the r-curve overshoots due to the momentum it gets from having a large fraction in i at the same time. The equation system depends on a few additional parameters, most notably the R_0 value that determines the speed of transmission, but also the mean incubation time $T_{incubation}$ and the mean infectious period $T_{infectious}$, explained in depth in the Supplementary Material (SM), Section 5 and 5.1.

An important issue to stress is the difference between the curve i, which models all infectives at a given moment t, and the fraction of *new infections* on a given day, which equals $\nu(t) := s(t-1) - s(t)$. These two are often confused and the latter curve is usually much smaller than the former (since people remain infective for some days but only become infected once). Measured curves for new infections based on PCR-test will scale linearly (due to insufficient testing) with ν (after shifting to account for $T_{incubation}$), not with i.

2.1 Discrepancies between model and reality

Simple models such as SEIR rely on gross simplification of human life, but these may still be useful if their *main characteristics* correlate acceptably well with reality. By main characteristics we mean quantities such as the time T_{wave} it takes for the outbreak to pass, roughly 70 days in Figure 1, which is roughly the time observed in Manaus. Another key quantity is the *final size of the epidemic* i.e. the fraction of people who at some point had COVID-19, which we define as $C19_{tot} = 100r(T_{wave})$. In Figure 1, $C19_{tot}$ is close to 80% and in the right graph we see that a whopping 2.5% will fall ill on the same day! To our knowledge this behavior is not near reality in any hard hit location of the planet (including naval ships and prisons), and yet, this is what many qualified experts and proponents of the mitigation-strategy had predicted would happen in early 2020 (see e.g. [24]). So why does the model fail so drastically?

One answer could of course lay in the (overly simplistic) setup of the equation system, as argued e.g. in [14]. We review the setup of basic SEIR in the Supplementary Material (Sections 5-5.1) with the aim of tuning the model to COVID-19. In short summary; on one hand there is a big gap between the SEIR-model and reality, and parameter choices are somewhat ad hoc. On the other hand, the model is fairly robust and different parameter choices do not affect the overall outcome drastically. Moreover, more accurate equation systems that also take the age of infection into account give almost identical output as the SEIR-equation system (see in particular the blue vs. the yellow curve in SM Figure 9), and hence it seems that these shortcomings can not explain the discrepancy between SEIR-based predictions and observations.

It is important to realize that in theory, these models are valid also under NPI's, as long as these are held constant, as then there will be an artificial fixed R_0 value that is adjusted for the corresponding NPI's, which we denote by R_0^{NPI}. So lets try to lower R_0^{NPI} so that $C19_{tot} \approx 52\%$ to match the observed data from Manaus [13]; $R_0^{NPI} = 1.4$ does the

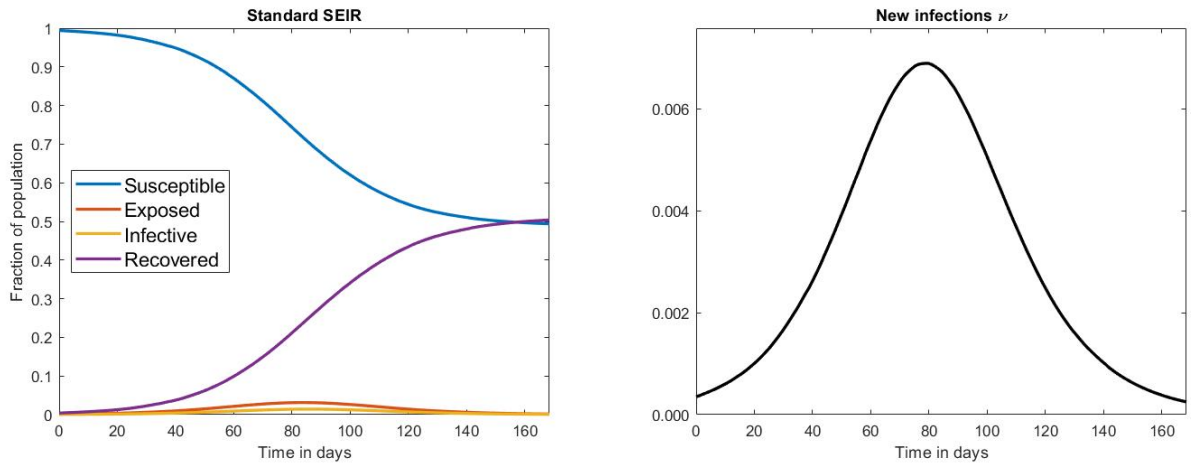

Figure 2: Left: SEIR outcome with $R_0 = 1.4$. Right: New infections.

job, see Figure 2. However, now the wave is expected to be around 150 days, twice the time suggested by real data. The Manaus curve is strikingly similar (in shape and duration) to the one for New York, which landed at $C19_{tot} = 23\%$ [6] as well as the one for Stockholm (that did not go into lock down) and measured a sero-prevalence of $C19_{tot} = 12\%$ in June [21]. With SEIR-modeling these outbreaks should have taken 1 and 2.5 years respectively, and is hence not realistic. This behavior is easy to understand on an intuitive level; in order for the final size of the epidemic to be small the R_0^{NPI} must be near one and then the virus spread becomes unrealistically slow. Thus, the problem with SEIR-models is that at least one of the key output-parameters T_{wave} and $C19_{tot}$ will be totally at odds with reality. *In other words, no matter how you tune the input parameters, the model generates curves which either miscalculates the pace or the magnitude of the epidemic outbreak.*

2.2 Possible explanations

A quick fix to the above dilemma is to include pre-immunity. In the SM Section 5.2 we show that for every solution to the SEIR equation system in the absence of pre-immunity, one can obtain a solution with a level θ of pre-immunity simply by multiplying the solution by the constant factor $(1-\theta)$ and adjusting R_0 accordingly. In other words, if we have found a solution which fits a given time-series well except for the magnitude, we may simply rescale it. To be concrete, if a model has a suitable shape and predicts T_{wave} correctly, but gives $C19_{tot} = 90\%$ whereas measured sero-prevalence indicates 52% [13], this could indicate that 42% are actually immune to infection, since $52 \approx (1-0.42)\cdot 90$. We investigate this possibility in Section 3, but first we will discuss other potential explanations.

It can be argued that the type of stochastic models mentioned above are not apt for modeling of SARS-CoV-2, which indeed is a peculiar virus that spreads in clusters, and it is estimated that 80% of the cases are caused by less than 20% of infected individuals, the so called "super spreaders" [1, 19]. In Section 7.3 we show that, rather surprisingly, adding this complexity to the deterministic SEIR-model does not in any way alter the output. This is no longer true if randomness is taken into account; in [23] a stochastic SEIR-model where R_0 is a random variable (for each infected individual) with a "fat-tailed" distribution, and while this displays an erratic and possibly more realistic behavior, its effect on T_{wave} and $C19_{tot}$ still does not seem to overcome the shortcomings discussed above (see in particular Figure 4 in [23]).

It has been argued that population heterogeneity may explain the gap between model and observation, see e.g. [12] and [50]. In Section 6 we take a closer look at these and show that while heterogeneity certainly plays an important role for more accurate modeling, it is unlikely to be the main explanation of the discrepancy. More precisely, when applied to the data series from Stockholm, the models are in both cases off by a factor of at least 3. To illustrate this, in Figure we see the second wave in Stockholm (blue) along with two attempts to model it using the code by [12] (that was published in Science, red and yellow curves chosen to match $C19_{tot}$ and T_{wave} respectively). The green and purple use similar codes but include pre-immunity, the details of which will be further explained in Section 3.

The remaining explanation that avoids pre-immunity is that *fluctuations* in the effective R-value, denoted R_e, due to NPI's and/or behavioral changes completely controls the way the cases evolve with time, rendering the SEIR-model useless. Yet, the NPI's in Sweden have remained virtually unchanged and SEIR-models are not anywhere near being able to match the development in Sweden (as seen in Figure 3, yellow and red curve) so if this is true it must be caused by

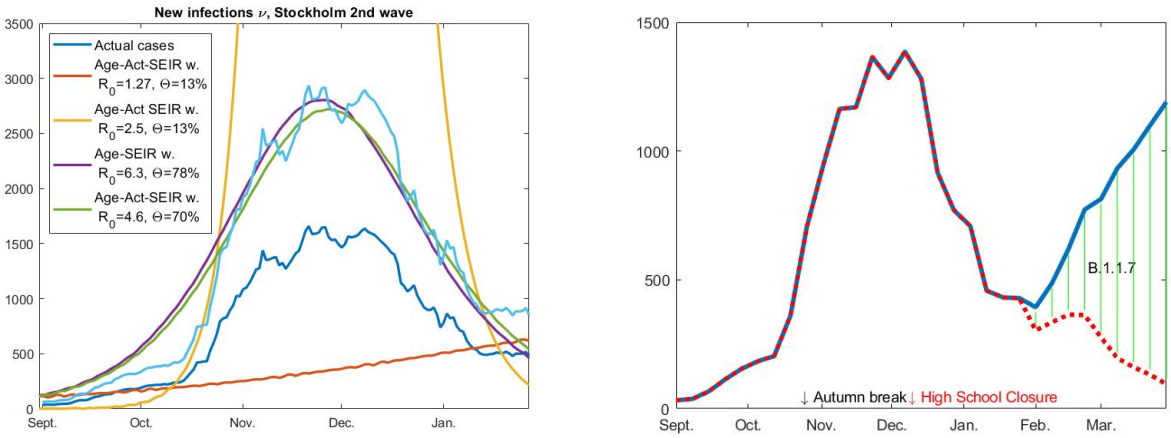

Figure 3: Left: Cases for Stockholm second wave (dark blue) and various attempts to fit with model curves. Light blue shows cases re-scaled to account for under-reporting of actual cases. Right: Cases (for Stockholm) continued into February and March, with and without B.1.1.7.

public awareness and voluntary behavioral changes by the population. In the coming section we take a closer look at this potential explanation.

2.3 The case of Stockholm, Sweden

The second wave hit Sweden much later than many other countries, leading to wide speculation that the "Swedish strategy" had been successful. We now know it did not give us advantages for the second wave, as the Swedish Health authority had predicted. A recent study estimates that the strategy lead to between 26 to 82% unnecessary deaths and no notable improvement for other factors, such as the economy [9].

We focus our analysis on the metropolitan area of Stockholm with 2.4 million inhabitants, which has been hit by 3 distinct waves, spring and autumn 2020 and now a third wave driven by the British mutation B.1.1.7 is estimated to soon reach its peak. In June 2020, 12% were estimated to have been infected in Stockholm, yet the decline of the spread started in early April, a month after onset of the first wave, even though the mitigation strategies were very limited and not well followed by the Swedes. From mid March, high schools and universities were on distance learning, people were expected to work from home and if possible avoid public transportation. Frequent washing of hands, keeping distance of 1.5-2 m, and staying at home if feeling sick, were the main recommendations given to the public from the health authorities and the government. A lock-down was never implemented, family members of confirmed cases were expected to work, grade 1-9 schools remained open and face masks were not recommended. It seemed odd that the soft interventions could have this effect unless the population was protected by some degree of pre-existing immunity. Nevertheless, it is of course possible that these interventions along with public awareness and arrival of spring was enough to keep R_e below one from early April and onwards.

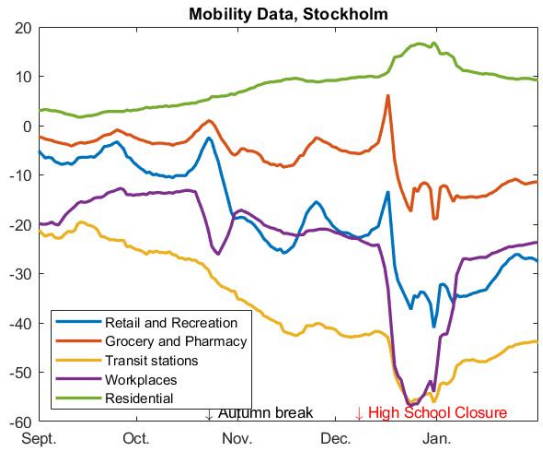

Figure 4: Google Community Mobility Reports

The same argument is however harder to accept considering the second wave (seen in Figure 3, blue curve, left as well as right). The recommendations from the government were the same as described for the first wave and public compliance limited; there were frequent reports of full shopping

malls and crowded transit stations, very few people wearing masks and the weather getting colder, driving people indoors. As is clear to see in Figure 3, the spread started to dampen in late October and peaked in late November, yet the only noteworthy change in recommendations occurred on December 7 when high schools went online. Thus it is unlikely NPI's which makes the steep rise in October to level out and then recede. It is important to realize that in the absence of the British mutation B.1.1.7 the epidemic in Stockholm would most likely have remained calm up until present at least, see the red dashed curve in Figure 3, right. This indicates that a temporary herd-immunity level had been reached (similar to Manaus between its first and second wave) unless, again, it is voluntary changes in behavior that explains why the second wave came to a premature halt.

To investigate this possible explanation we display mobility data for Stockholm obtained from Google Community Mobility Reports, seen in Fig. 4. Clearly there is a decline in activity from around 5% (Grocery and Pharmacy), 15% (Workplaces) to 35% (Transit stations) between October and November/early December, but it is hard to imagine that this has the capacity of reducing the R_e-value to below 1 (from values around 1.5 during October). Notably, the autumn holidays in the end of October as well as Christmas holidays end of December (that in Sweden comes close to a lock-down) have no clear bearing on the development, indicating that behavioral changes are not the main factor for the apparent changes in R_e. In summary, if SEIR-models have any validity at all in describing the spread of SARS-CoV-2, we find that it is very hard to explain the second wave in Stockholm without taking pre-existing immunity into account.

Similar observations are also reported from Lombardy, Italy [52]. The authors write " It appears that a comparatively high cumulative incidence of infection, even if far below theoretical thresholds required for herd immunity, may provide noticeable protection during the second wave" and then argue that these observations can not be attributed to neither NPI's nor behavioral or environmental variations alone.

3 SEIR-modeling with pre-immunity

Several scientists have proposed that a level of pre-immunity within the population could partially explain the unexpected behavior of the SARS-CoV-2 virus spread in society [22, 33, 37], but without providing a thorough mathematical analysis. Similar arguments have also been put forth by Doshi [17] as well as by Sette and Crotty [48], see SM Section 5.2 for a longer discussion. If a pre-existing immunity of substantial relevance would exist, it would significantly affect the model, but in this article we do not speculate in the nature of the pre-immunity, we just provide mathematical support for its existence.

3.1 Incorporating pre-immunity into the SEIR model

The simplest way to incorporate pre-existing immunity in the mathematical model is to select an immunity level of $\Theta\%$ and set $s_{initial} = 1 - \Theta/100$, where $s_{initial}$ is the initial fraction of susceptible individuals in the population, (see SM Section 5.2 for details). The only assumption needed for the SEIR-model to work in theory, is that the NPI's and public behavior remain rather constant during the elapse of the wave, and that seasonal factors have no major impact on the basic reproduction number; R_0^{NPI}. We recall that R_0^{NPI} denotes the amount of people one infected individual transmits the disease to, at the onset of the epidemic with NPI's in place, *and assuming that everyone else is susceptible, i.e.* $s(t) \approx 1$. Under the above assumptions it follows that the effective R-value at time t depends only on $S(t)$ via the simple formula

$$R_e(t) = R_0^{NPI} s(t). \tag{1}$$

In particular, note that $R_e(0) = R_0^{NPI}(1 - \Theta/100)$ where $R_e(0)$ represents the "R_0"-value that initially was estimated in the range 2-3, so this formula entails that these estimates of "R_0" must be adjusted upwards, as a pre-immunity would directly affect the R_0-value. For example, a pre-immunity of 50% would put the above estimates for R_0^{NPI} in the range 4-6.

A more refined way to introduce pre-immunity is by including variable susceptibility, since "being immune" is not a binary variable. On the contrary, neither cross-reactive adaptive nor innate immunity is likely to hinder infection when exposure to the virus is high. We describe the mathematical details in Section 7.8. In particular, we show in Section 7.9 that there is only a slight difference between modeling variable activity levels and variable susceptibility, so in practice it may be hard to distinguish one of these explanations from the other.

To exemplify this, in Figure 5 we plot the output of the basic homogenous SEIR-model (with pre-immunity $\Theta = 35\%$), the age-stratified model by Britton et al. (with pre-immunity $\Theta = 25\%$), the age-activity stratified model (again by Britton et al., now with $\Theta = 0\%$ and "activity parameter" $\eta = 2$, see Section 6.1 or 7.9 for more details), and finally a more advanced age-susceptible stratified model where 25% of the population has no protection at all against the virus, whereas the remaining 75% with pre-existing immunity are divided into three groups and given various levels of protection.

As is plain to see, given any of the above SEIR-models it is possible to chose parameters in any of the other models to get almost identical output. From this we can draw a number of interesting conclusions:

1. It is impossible within this mathematical framework to draw any certain conclusions about the nature of the pandemic, since different real world phenomena have an almost identical effect on the curves. However, this conclusion only holds within a certain parameter interval; as we saw in Figure 3 it is impossible to model the development in Stockholm with the age-activity stratified model and no pre-existing immunity, without choosing absurd values for the parameter η regulating activity variation. This is further elaborated in SM Section 6.1.

2. Immunity is rarely either 100% or 0%, whether a person gets infected will depend on the dose of infectious virus attacking a person with different protective antibody and T-cell immunity levels. Despite this, a more refined model taking various levels of immunity into account, and a simpler model treating the biological reality as binary, give the same output. Hence, if one is careful with the *interpretation* of Θ as the amount of "immune", it is perfectly fine from a mathematical viewpoint to replace the more advanced age-susceptibility stratified model with a more simple one.

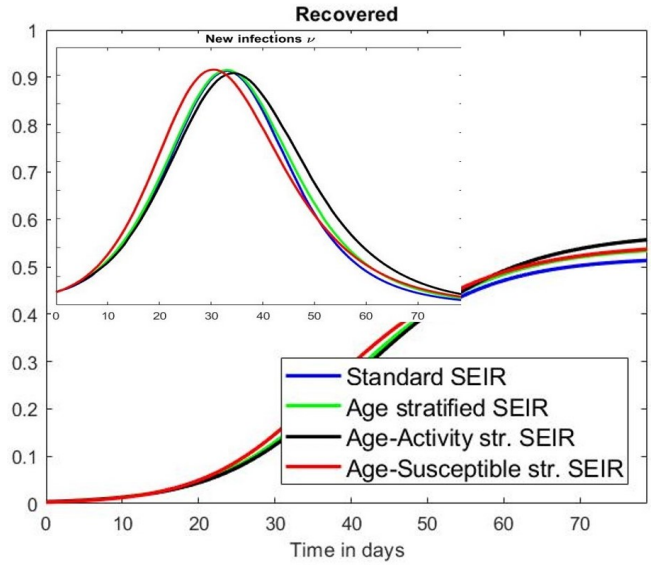

Figure 5: Blue: Standard SEIR (2) with $R_0 = 3.1$ and pre-immunity level $\Theta = 35\%$. Green: Age-stratified SEIR with $R_0 = 2.7$ and $\Theta = 25\%$. Black: Age-activity model with "activity difference factor $\eta = 2$", $R_0 = 2$ and $\Theta = 0$. Red: Age-susceptible stratified SEIR with $R_0 = 6$.

3.2 Modeling examples with pre-immunity

We exemplify by applying the model to the situations in Manaus, India (first wave) and Stockholm (second wave), and we satisfy with modeling pre-immunity using the simpler approach with a "binary" parameter Θ for the percentage of pre-immune (using the age-activity stratified model does not lead to drastically different conclusions). In SM Section 6.1 we show that the age-activity stratified model by Britton et al. is not realistic[1], especially for modeling poorer places like Manaus and India.

We will first focus on Manaus, since this is an example of a relatively unmitigated spread that is believed to have reached saturation without seasonal interference. In Figure 6 we display the result using pre-existing immunity level $\Theta = 42\%$. The output overlaps surprisingly well with observed data, the shape and T_{wave} are reasonable and $C19_{tot}$ becomes 44%, which is what was measured in [13]. For this to occur we chose $R_0^{NPI} = 3.45$, which corresponds to an "observed" R_0-value of $R_e(0) = 2$, via formula (1). The age-distribution has been adjusted for Brazil, as explained in Section 7.1.[2]

We then ran the same model for the country of India, again setting $\Theta = 42\%$ but now choosing a lower value for $R_0^{NPI} = 2.21$ in order to get $C19_{tot} = 21.4$, which was the national seroprevalence as estimated in the end of February by ICMR (Indian Council of Medical Research). That R_0^{NPI} vary between countries could have a number of explanations, but it is very interesting that the same level of pre-immunity Θ as Manaus gives a perfect data-fit. India is currently seeing a substantial second wave, which recent data indicates is due to a mutant virus variation.

Before proceeding, we would like to stress that this research is in no way meant to support mitigation strategies to control the pandemic, since the examples set by e.g. Sweden and Brazil clearly shows that this has led to substantial amount of human suffering and unnecessary loss of life [9]. Moreover the tragic second wave hitting Manaus shows with all clarity that this is not a path to be considered; even if Manaus had "herd-immunity" for some time, the recent surge is probably related to both mutations and loss of immunity [46], and both issues speak firmly for relying on suppression/vaccination strategies [4]. On the other hand, if some degree of pre-existing immunity is present in the world's population, we believe that this is an important fact that can be used for better prediction of the pandemic and understanding of the virus.

[1] at least not with activity level parameter $\eta = 2$, and for reasonable values of η the difference with only age-stratification becomes negligible

[2] From now on we use the measured data values and not age-sex reweighted. Whether we pick $C19_{tot} = 44\%$ or 52% has no bearing on the major conclusions, using 52% gives $R_0^{NPI} = 3.0$ and $\Theta = 33\%$

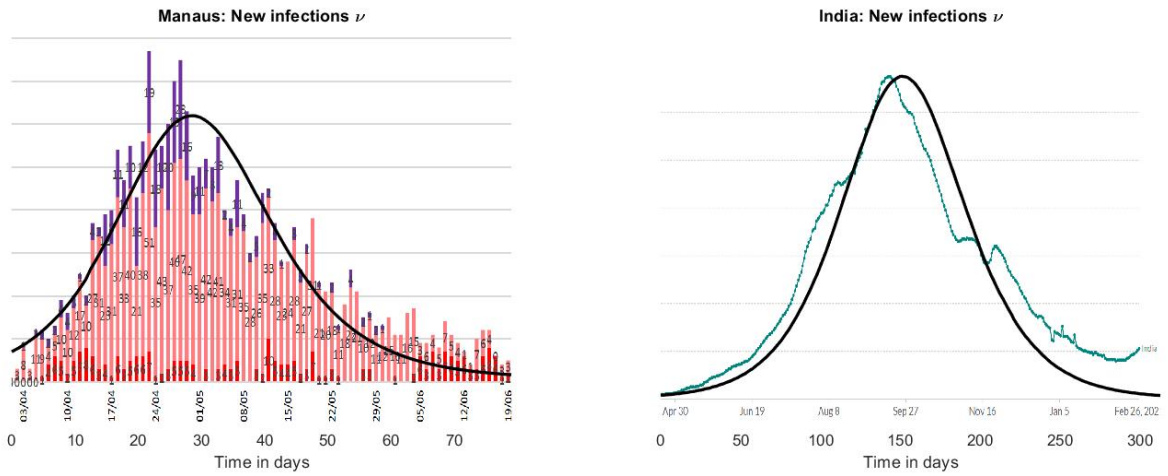

Figure 6: Left: Age-stratified SEIR with $R_0^{NPI} = 3.45$ and pre-immunity 42%. For this choice $C19_{tot}$ equals 44% and "the apparent R_0-value" $R_e(0)$ is 2.0. y-axis have been scaled to fit curve of deaths in Manaus.
Right: Same model run for India with pre-immunity 42% (again). We set $R_0^{NPI} = 2.21$ in order to achieve $C19_{tot} = 21.4\%$. y-axis have been re-scaled.

In fact, the new waves caused by the mutations are commonly explained by attributing to them a higher $R-$value, but the hypothesis of a pre-existing immunity offers an alternative interpretation. If there was a pre-existing immunity caused by cross-reactivity with other viruses, which according to the above modeling could have protected around 40 percent of the population, then any mutation which is able to get past this pre-immunity will seem like it has a higher R, when in fact it simply has a larger pool of people it can infect.

3.3 Stockholm, second wave

We now describe in further detail our in depth study of Stockholm from Section 2.3. In June, Stockholm was estimated to have around 12% sero-prevalence [21]. Note that the recent study [15] shows a moderate decline of antibodies to SARS-CoV-2, and also the study [27] has established that prior infection results in 80% protection against reinfection, which did not seem to wane over time. Therefore we have not included loss of immunity over time in our models, which typically model a couple of months at the time. Instead, it is important to have a measurement of the immunity level at the onset of the wave to be modeled. Our own measurement of the seroprevalence of SARS-CoV-2 among 300 blood donors in September gave 13%, which thus correlate well with the 12% measured in June [21].

The Swedish second wave is a good option for mathematical modeling, as public recommendations have remained rather constant except for the closing of high-schools on December 7. The sero-prevalence in Stockholm has now risen to 21.5%, which we base on a recent study of 450 blood donors sampled 22 Feb-7 March 2021. Due to the relatively small number of donors the figure is uncertain, so we also consider death data. Stockholm had 2400 deaths in the first wave and 1500 in the second wave. If we assume constant IFR and extrapolate from 12% seroprevalence corresponding to 2400 deaths, then the increase in sero-prevalence in the second wave should be at least $12 * 1500/2400 = 7.5\%$. Finally, an independent study [43] concluded that 17% of Stockholm blood donors were seropositive in the midst of the second wave, which is also in line with an 8.5% increase.

We tried to find parameters that would fit the shape of the time-series of cases in stockholm during the second wave (blue curve, Figure 3) while matching a final size of the epidemic of 8.5%. In short conclusion, it is impossible to match this data with any SEIR-model not incorporating levels of pre-existing immunity above 50% and $R_0^{NPI}-$values above 4. In order to get a good fit with the age-activity stratified SEIR [12], we had to choose the activity parameter $\eta = 100$, which is absurd. Or rather, in practice this means that the spread only takes place among the 25% "highly active", and hence this is the same as assuming that 75% are either pre-immune or isolating, see SM Section 7.9.

We display a few of these tests in Figure 3. The dark blue curve is the 7 day average of actual measured cases in Stockholm for the period 1 Sept. 2020 to 28 Jan. 2021 (so 1/9 displays the average of 1-7/9). Here, we do not re-scale the y-axis as in the previous modeling examples, and hence the dark blue is expected to be lower than modeled data due to underreporting. The light blue curve is the cases re-scaled so that it integrates to 8.5% infected during the given

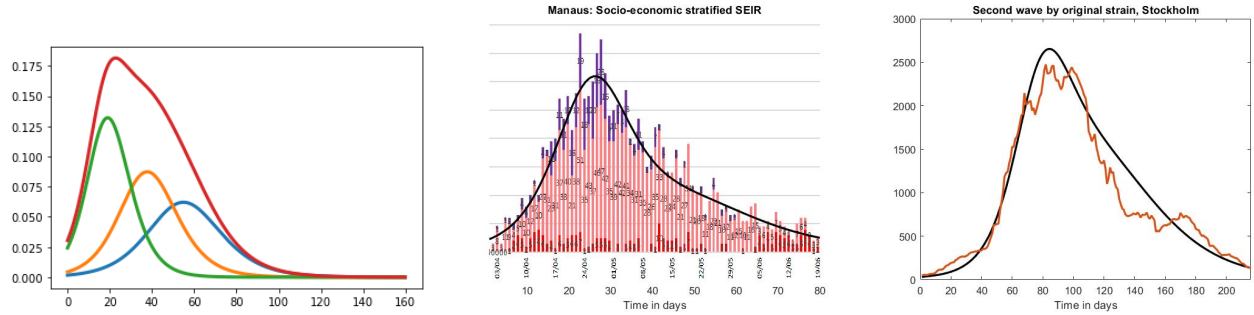

Figure 7: Left: 3 independent SEIR-curves for different R_0-values, and their sum. Middle: Manaus modeled using $R_0^{NPI} = 4.2$, $\Theta = 43\%$ giving $C19_{tot} = 44\%$. Right: Stockholm modeled using $R_0^{NPI} = 4.3$, $\Theta = 62\%$ giving $C19_{tot} = 8.5\%$.

time-period.

We first ran Britton et al's Age-Activity stratified SEIR model with pre-existing immunity to 13%, corresponding to pre-pandemic immunity of 0%. In order to get 8.5% as the "final size of the second wave", we needed to pick $R_0^{NPI} = 1.27$, the result is seen in red. In order to get a realistic value for T_{wave} we need to pick $R_0^{NPI} = 2.5$, leading to the yellow curve and a final size of the second wave of 54%. Looking at the corresponding curves, it is apparent that both alternatives are completely off.

Next we tried to find values of R_0^{NPI} and Θ that would match 8.5% as the final size of the second wave, using the simpler Age-Stratified SEIR. The choice $\Theta = 78\%$ and $R_0^{NPI} = 6.3$ gives a perfect fit. If again we incorporate the various activity levels by Britton et al. ($\eta = 2$), then these values drop further to $\Theta = 70\%$ and $R_0^{NPI} = 4.6$. The age-activity model is more realistic in the Swedish scenario than Manaus/India, due to the social well-fare state and extensive IT-infrastructure, combined with many single households and online based jobs, which makes it easier for large portions of the population to self-isolate to various degrees. For this reason we ran both models for Stockholm, and so combined this gives evidence for a pre-existing immunity in the range 70-78%. Since this is the second wave, the values of pre-existing immunity also includes the 13% acquired during the first wave, so a value of e.g. $\Theta = 78\%$ corresponds to pre(pandemic)-existing immunity of 65%. Moreover, it is believed that a good portion of the 15% that are above 70 are self-isolating (which is much easier to achieve in Sweden than India or Manaus, due to the structure of society and pension-systems), so the above 65% could in practice be 55%. Another source of potential error is that we have not corrected the value 8.5% for false negatives. However, this will not alter the magnitude of the numbers reported above. In summary, it is impossible to give firm predictions on what was the effective level of prepandemic-immunity to SARS-CoV-2 in Stockholm, but the above analysis gives support for the hypothesis that this value is somewhere between 45 and 65%, irrespective of which model/measurement one relies on.

3.4 The Gompertz curve and stability of estimates

In this section we briefly discuss if the heterogeneous model can be further tuned to more accurately fit observed data, and test if such refinements could alter the above conclusions. The first waves in a number of places had the shape of a "Gompertz"-curve (characterized by a sharp rise and a slow decline). In the supplementary material, we propose a socio-economic stratified SEIR model to explain this. The idea is simple, if you have a rich part of town and a poorer part of town, then the R_0^{NPI}-value is likely to be higher in the former and lower in the latter. Once you add up the curves, the result looks like a Gompertz-function, as shown in Figure 7 (left). We also managed to fit Manaus almost perfectly by setting $R_0^{NPI} = 4.2$ and pre-existing immunity $\Theta = 43\%$ (Figure 7, middle), again abiding by the restriction $C19_{tot} = 44\%$ as measured in [13]. This should be compared with $\Theta = 42\%$ and $R_0^{NPI} = 3.45$ used in Figure 6. It is remarkable that, despite variations in the model, key behavior is surprisingly consistent, and we have not been able to get a similar fit with any reasonable model and $\Theta = 0\%$, indicating a certain robustness to choice of model.

The same comment goes for Stockholm, to the right in Figure 7 we see the second wave in stockholm not including B.1.1.7, (compare Fig. 3). We modified the Socio-economic age-activity SEIR-model a bit further to take into account that the 70+ group is to a large extent self-isolating, and that social mixing in between age-groups certainly is reduced during the pandemic. The outcome is seen in black, using pre-immunity 62% and $R_0^{NPI} = 4.3$, not too far off from the corresponding values used for Manaus. In summary, we see that more advanced models can yield even better accuracy, but the values of R_0^{NPI} and Θ remain of the same magnitude. In particular, it is still impossible to get near observed

data with $\Theta = 0$, and this gives mathematical support to the hypothesis that a certain level of pre-existing immunity was present before the arrival of SARS-CoV-2.

4 Conclusion

We undertook a major evaluation of different SEIR-models taking factors such as age, activity level, susceptibility variation, socio-economic level and pre-existing immunity into account. We tested these models against data acquired from Manaus, India and Stockholm, and found that a simple age-stratified SEIR-model with pre-immunity in the range 40-65% gave good fit with observed data, which is impossible to attain in the absence of pre-immunity, as the other factors can not affect the model sufficiently. We also analyzed a number of potential alternative explanations and found that they did not seem backed up by data, so we argue that if SEIR-models are at all useful for modeling COVID-19, then pre-existing immunity is a plausible explanation to the unexpected development of the pandemic that should encourage further investigation in order to reveal its nature.

Acknowledgement

We thank the authors of [8] for sharing their data with us. We thank Jakob Svensson (Max Planck Inst.) for interesting and fruitful discussions on the subject. We thank Andrea B. Simons for help with Brazil data. We thank Cecilia Hellström, Peter Nilsson and Sophia Hober for their contribution to antibody pre-immunity data and Birger Sørensen and Andres Susrud at Immunor AS for collaborative work on identifying the Influenza A/SARS-CoV-2 cross protective immunity. We thank Afsar Rahbar, Nerea Martín Almazán, Mattia Russel Pantalone for data regarding pre-immunity serology prevalence.

References

[1] Dillon C Adam, Peng Wu, Jessica Y Wong, Eric HY Lau, Tim K Tsang, Simon Cauchemez, Gabriel M Leung, and Benjamin J Cowling. Clustering and superspreading potential of sars-cov-2 infections in hong kong. *Nature Medicine*, 26(11):1714–1719, 2020.

[2] The Amazonas Health Surveillance Foundation (FVS / AM). BOLETIM DIARIO COVID-19 NO AMAZONAS 1/1/2021. "`http://www.fvs.am.gov.br/media/publicacao/01_01_21_BOLETIM_DIARIO_DE_CASOS_COVID-19.pdf`".

[3] Hakan Andersson and Tom Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.

[4] Christie Aschwanden. Five reasons why covid herd immunity is probably impossible. *Nature*, 591.

[5] Norman TJ Bailey et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[6] Kristina L Bajema, Ryan E Wiegand, Kendra Cuffe, Sadhna V Patel, Ronaldo Iachan, Travis Lim, Adam Lee, Davia Moyse, Fiona P Havers, Lee Harding, et al. Estimated sars-cov-2 seroprevalence in the us as of september 2020. *JAMA internal medicine*, 2020.

[7] Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.

[8] Qifang Bi, Yongsheng Wu, Shujiang Mei, Chenfei Ye, Xuan Zou, Zhen Zhang, Xiaojian Liu, Lan Wei, Shaun A Truelove, Tong Zhang, et al. Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. *The Lancet Infectious Diseases*, 2020.

[9] Benjamin Born, Alexander M Dietrich, and Gernot J Müller. The lockdown effect: A counterfactual for sweden. *Plos one*, 16(4):e0249732, 2021.

[10] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*. Springer, 2019.

[11] Julian Braun, Lucie Loyal, Marco Frentsch, Daniel Wendisch, Philipp Georg, Florian Kurth, Stefan Hippenstiel, Manuela Dingeldey, Beate Kruse, Florent Fauchere, et al. Sars-cov-2-reactive t cells in healthy donors and patients with covid-19. *Nature*, 587(7833):270–274, 2020.

[12] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2. *Science*, 369(6505):846–849, 2020.

[13] Lewis F Buss, Carlos A Prete, Claudia MM Abrahim, Alfredo Mendrone, Tassila Salomon, Cesar de Almeida-Neto, Rafael FO França, Maria C Belotti, Maria PSS Carvalho, Allyson G Costa, et al. Three-quarters attack rate of sars-cov-2 in the brazilian amazon during a largely unmitigated epidemic. *Science*, 2020.

[14] Shi Chen, Patrick Robinson, Daniel Janies, and Michael Dulin. Four challenges associated with current mathematical modeling paradigm of infectious diseases and call for a shift. In *Open forum infectious diseases*, volume 7, page ofaa333. Oxford University Press US, 2020.

[15] Jennifer M Dan, Jose Mateus, Yu Kato, Kathryn M Hastie, Esther Dawen Yu, Caterina E Faliti, Alba Grifoni, Sydney I Ramirez, Sonya Haupt, April Frazier, et al. Immunological memory to sars-cov-2 assessed for up to 8 months after infection. *Science*, 2021.

[16] Kieran Dee, Daniel M Goldfarb, Joanne Haney, Julien AR Amat, Vanessa Herder, Meredith Stewart, Agnieszka M Szemiel, Marc Baguelin, and Pablo R Murcia. Human rhinovirus infection blocks sars-cov-2 replication within the respiratory epithelium: implications for covid-19 epidemiology. *Journal of Infectious Diseases*, 2021.

[17] Peter Doshi. Covid-19: Do many people have pre-existing immunity? *Bmj*, 370, 2020.

[18] Zhanwei Du, Xiaoke Xu, Ye Wu, Lin Wang, Benjamin J Cowling, and Lauren Ancel Meyers. Serial interval of covid-19 among publicly reported confirmed cases. *Emerging infectious diseases*, 26(6):1341, 2020.

[19] Akira Endo et al. Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. *Wellcome Open Research*, 5, 2020.

[20] Zhilan Feng, Dashun Xu, and Haiyun Zhao. Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bulletin of mathematical biology*, 69(5):1511–1536, 2007.

[21] Folkhalsmyndigheten. Påvisning av antikroppar efter genomgången covid-19 i blodprov från öppenvården (delrapport 1).

[22] Karl Friston, Anthony Costello, and Deenan Pillay. Dark matter, second waves and epidemiological modelling. *BMJ Global Health*, 5(12):e003978, 2020.

[23] Masao Fukui and Chishio Furukawa. Power laws in superspreading events: evidence from coronavirus outbreaks and implications for sir models. *MedRxiv*, 2020.

[24] Johan Giesecke. The invisible pandemic. *The Lancet*, 395(10238):e98, 2020.

[25] Alba Grifoni, Daniela Weiskopf, Sydney I Ramirez, Jose Mateus, Jennifer M Dan, Carolyn Rydyznski Moderbacher, Stephen A Rawlings, Aaron Sutherland, Lakshmanane Premkumar, Ramesh S Jadi, et al. Targets of t cell responses to sars-cov-2 coronavirus in humans with covid-19 disease and unexposed individuals. *Cell*, 181(7):1489–1501, 2020.

[26] Deepti Gurdasani and Hisham Ziauddeen. On the fallibility of simulation models in informing pandemic responses. *The Lancet Global Health*, 8(6):e776–e777, 2020.

[27] Christian Holm Hansen, Daniela Michlmayr, Sophie Madeleine Gubbels, Kåre Mølbak, and Steen Ethelberg. Assessment of protection against reinfection with sars-cov-2 among 4 million pcr-tested individuals in denmark in 2020: a population-level observational study. *The lancet*.

[28] Herbert W Hethcote and David W Tudor. Integral equation models for endemic infectious diseases. *Journal of mathematical biology*, 9(1):37–47, 1980.

[29] Michael A Johansson, Talia M Quandelacy, Sarah Kada, Pragati Venkata Prasad, Molly Steele, John T Brooks, Rachel B Slayton, Matthew Biggerstaff, and Jay C Butler. Sars-cov-2 transmission from people without covid-19 symptoms. *JAMA network open*, 4(1):e2035057–e2035057, 2021.

[30] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[31] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

[32] Sonja Lehtinen, Peter Ashcroft, and Sebastian Bonhoeffer. On the relationship between serial interval, infectiousness profile and generation time. *Journal of the Royal Society Interface*, 18(174):20200756, 2021.

[33] Michael Levitt, Andrea Scaiewicz, and Francesco Zonta. Predicting the trajectory of any covid19 epidemic from the best straight line. *medRxiv*, 2020.

[34] Natalie M Linton, Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R Akhmetzhanov, Sung-mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine*, 9(2):538, 2020.

[35] Alex Lipton and Marcos Lopez de Prado. Mitigation strategies for covid-19: Lessons from the k-seir model. *Available at SSRN 3623544*, 2020.

[36] Alun L Lloyd. Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1470):985–993, 2001.

[37] Jose Lourenco, Francesco Pinotti, Craig Thompson, and Sunetra Gupta. The impact of host resistance on cumulative mortality and the threshold of herd immunity for sars-cov-2. *MedRxiv*, 2020.

[38] Jose Mateus, Alba Grifoni, Alison Tarke, John Sidney, Sydney I Ramirez, Jennifer M Dan, Zoe C Burger, Stephen A Rawlings, Davey M Smith, Elizabeth Phillips, et al. Selective and cross-reactive sars-cov-2 t cell epitopes in unexposed humans. *Science*, 370(6512):89–94, 2020.

[39] Mélodie Monod, Alexandra Blenkinsop, Xiaoyue Xi, Daniel Hebert, Sivan Bershan, Simon Tietze, Marc Baguelin, Valerie C Bradley, Yu Chen, Helen Coupland, et al. Age groups that sustain resurging covid-19 epidemics in the united states. *Science*, 2021.

[40] Kevin W Ng, Nikhil Faulkner, Georgina H Cornish, Annachiara Rosa, Ruth Harvey, Saira Hussain, Rachel Ulferts, Christopher Earl, Antoni G Wrobel, Donald J Benton, et al. Preexisting and de novo humoral immunity to sars-cov-2 in humans. *Science*, 370(6522):1339–1343, 2020.

[41] Institute of Brazilian Geography and Statistics (IGBM). https://cidades.ibge.gov.br/, 2021. [Online; accessed 15-Jan-2021].

[42] Akira Ohnishi, Yusuke Namekawa, and Tokuro Fukui. Universality in covid-19 spread in view of the gompertz function. *Progress of Theoretical and Experimental Physics*, 2020(12):123J01, 2020.

[43] Elena Percivalle, Giuseppe Cambiè, Irene Cassaniti, Edoardo Vecchio Nepita, Roberta Maserati, Alessandro Ferrari, Raffaella Di Martino, Paola Isernia, Francesco Mojoli, Raffaele Bruno, et al. Prevalence of sars-cov-2 specific neutralising antibodies in blood donors from the lodi red zone in lombardy, italy, as at 06 april 2020. *Eurosurveillance*, 25(24):2001031, 2020.

[44] Allyson M Pollock and James Lancaster. Asymptomatic transmission of covid-19. *BMJ*, 2020.

[45] Public-Health-England. Disparities in the risk and outcomes of covid-19, 2020.

[46] Ester C Sabino, Lewis F Buss, Maria PS Carvalho, Carlos A Prete, Myuki AE Crispim, Nelson A Fraiji, Rafael HM Pereira, Kris V Parag, Pedro da Silva Peixoto, Moritz UG Kraemer, et al. Resurgence of covid-19 in manaus, brazil, despite high seroprevalence. *The Lancet*, 397(10273):452–455, 2021.

[47] Takuya Sekine, André Perez-Potti, Olga Rivera-Ballesteros, Kristoffer Strålin, Jean-Baptiste Gorin, Annika Olsson, Sian Llewellyn-Lacey, Habiba Kamal, Gordana Bogdanovic, Sandra Muschiol, et al. Robust t cell immunity in convalescent individuals with asymptomatic or mild covid-19. *Cell*, 183(1):158–168, 2020.

[48] Alessandro Sette and Shane Crotty. Pre-existing immunity to sars-cov-2: the knowns and unknowns. *Nature Reviews Immunology*, 20(8):457–458, 2020.

[49] Carlo Signorelli, Alberto Zucchi, Carlo Alberto Tersalvi, Roberta Ciampichini, Elvira Beato, Federica Balzarini, Anna Odone, and John Middleton. High seroprevalence of sars_cov-2 in bergamo: evidence for herd immunity or reason to be cautious? *International Journal of Public Health*, 65(9):1815–1817, 2020.

[50] Henrik Sjödin, Anders F Johansson, Åke Brännström, Zia Farooq, Hedi Katre Kriit, Annelies Wilder-Smith, Christofer Åström, Johan Thunberg, Mårten Söderquist, and Joacim Rocklöv. Covid-19 healthcare demand and mortality in sweden in response to non-pharmaceutical mitigation and suppression scenarios. *International journal of epidemiology*, 2020.

[51] SOU 2020:80 Statens offentliga utredningar. Äldreomsorgen under pandemin.

[52] Marco Vinceti, Tommaso Filippini, Kenneth J Rothman, Silvia Di Federico, and Nicola Orsini. Sars-cov-2 infection incidence during the first and second covid-19 waves in italy. *Environmental research*, page 111097, 2021.

[53] Patrick GT Walker, Charles Whittaker, Oliver J Watson, Marc Baguelin, Peter Winskill, Arran Hamlet, Bimandra A Djafaara, Zulma Cucunubá, Daniela Olivera Mesa, Will Green, et al. The impact of covid-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, 2020.

[54] Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10):936–944, 2006.

[55] Gizachew Tadesse Wassie, Abebaw Gedef Azene, Getasew Mulat Bantie, Getenet Dessie, and Abiba Mihret Aragaw. Incubation period of severe acute respiratory syndrome novel coronavirus 2 that causes coronavirus disease 2019: A systematic review and meta-analysis. *Current Therapeutic Research, Clinical and Experimental*, 93, 2020.

[56] Arabella Widders, Alex Broom, and Jennifer Broom. Sars-cov-2: the viral shedding vs infectivity dilemma. *Infection, Disease & Health*, 2020.

[57] Hisham Ziauddeen, Naresh Subramaniam, and Deepti Gurdasani. Modelling the impact of lockdown easing measures on cumulative covid-19 cases and deaths in england. *medRxiv*, 2020.