The unexpected dynamics of COVID-19 in Manaus, Brazil: Was herd immunity achieved?

Daihai He[1], Yael Artzy-Randrup[2], Salihu S. Musa[1], Lewi Stone[3,4,*]

[1] Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China.
[2] Department of Theoretical and Computational Ecology, IBED, University of Amsterdam, Amsterdam, Netherlands
[3] Mathematical Sciences, School of Science, RMIT University, Melbourne, Australia
[4] Biomathematics Unit, School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel
[*] Correspondence author: LewiStone100@gmail.com

Abstract

In late March 2020, SARS-CoV-2 arrived in Manaus, Brazil, and rapidly developed into a large-scale epidemic that collapsed the local health system, and resulted in extreme death rates. Buss et al.(2021) reported ~76% of residents were infected by October 2020, implying protective herd immunity had been reached. However, in December 2020 a larger unexpected second wave of COVID-19 struck again. Novel methods are used to model the epidemic from mortality data, evaluate the impact of interventions, and to understand why the second wave appeared. The method fits a "flexible" reproductive number $R_0(t)$ that changes over the epidemic. The two-strain model takes into account the new P.1 virus lineage, provides an accurate fit to observed epidemic datasets, and finds a ~70% attack-rate by March 2021.

Keyword: COVID-19; infection attack rate; infection fatality ratio; herd immunity; mathematical modeling

Main text

In December 2020, a vicious second wave of COVID-19 erupted in Manaus, Brazil, resulting in the deaths of more than 170 residents per day, and triggering yet another collapse in the country's rundown healthcare system.(*1*) Awkwardly, the outbreak occurred just a few weeks after an important seroprevalence study(*2*) estimated that by October 2020[(2),3], ~76% of the city's population had already been infected indicating that herd immunity had been attained. Yet even though this would imply there could be no further epidemic outbreaks, a second COVID-19 wave soon followed, which proved to be even larger in terms of the number of people infected than the first wave(*3, 4*). Currently there are only speculations as to why a major second wave appeared.(*3*) One possibility is that immunity gained in the first epidemic had waned quickly enough to allow re-infections. An alternative possibility is that the different SARS-CoV-2 strains that dominated the first and second waves are antigenically different enough not to elicit effective cross-immunity from re-infection. However, assuming the original population estimates of the study are correct(*2*), this would imply that approximately 50% of the population of 2.2 million residents had been infected twice, once in the first wave and then again in the second. But unusually, to date there is no evidence to support these high reinfection numbers. On the contrary, as of March 9 2021, no more than several reinfections have been confirmed(*5, 6*) by local health authorities out of hundreds of thousands of cases. Another possibility is that the Attack Rate estimate (AR; percentage of population infected), via seroprevalence testing, was inaccurate given the potential for biases(*7, 8*) in the blood donor data assessed by Buss et al.(*2*).

These major discrepancies warrant serious further re-investigation of the epidemiological processes that took place during the first wave in order to make more reliable and intelligible assessments of the dynamics of COVID-19 now, and in future. Here we explore alternative and complementary epidemiological modelling analyses of publicly available mortality datasets to help understand why the first wave of the COVID-19 outbreak "crashed," and why a second wave suddenly appeared in Manaus in December 2020 contrary to expectations. We provide a parsimonious model that accurately reproduces the two-wave and two-strain COVID-19 dynamics in Manaus that resolves these contradictions, without assuming large reinfection rates. The modelling analysis also gives important information on the impact of non-pharmaceutical interventions (NPIs), the questionable attainment of herd immunity, and provides estimates on the transmission of the SARS-CoV-2 virus and the new Variant of Concern, the P.1 lineage.

First, it should be clarified that it is not uncommon for potentially large-scale epidemic outbreaks to die out well before predictions of conventional epidemiological models, and well before herd-immunity has developed(*9*). For example, the Ebola outbreak in Sierra Leone Guinea 2014(*10*) began with an initially susceptible population of 5,364,000 residents. While a conventional SEIR-type model predicted 5,157,000 Ebola cases,(*9*) the epidemic resulted in "only" 14,120 infected individuals. We documented a similarly perplexing account of a catastrophic epidemic outbreak in WW2.(*11*) Using a similar modelling approach to that proposed here, we showed that a dangerous outbreak of pandemic typhus crashed well before herd immunity was reached, and just before winter when the number of infected cases should accelerate. Brauer(*9*) demonstrated that the early demise of an epidemic can occur in the presence of: i) highly heterogeneous transmission and super-spreader-like events;(*10, 12*) ii) human behavioural responses which act to depress the size of epidemics by reducing the reproductive number $R_0(t)$ over time,(*9, 11-14*) providing a "protection" effect based on behaviour rather than immunity. Brauer(*9*) modelled behavioural interference by assuming that $R_0(t)$ (which is a measure of transmission) reduces exponentially over time at rate $e^{-\kappa t}$. Other studies of cholera and influenza use similar parametric approaches(*13, 15*). Here we generalise this principle by fitting a "flexible" $R_0(t)$ to the data without pre-imposing any specific functional form.(*11, 12, 16*) Our method makes

it possible to reconstruct how $R_0(t)$ changes in time, and from this gain some insights on how interventions impacted the epidemic.

A basic overview of the COVID-19 dynamics in Brazil can be gained from examining Figure 1 which plots the daily number of Severe Acute Respiratory Illness (SARI) deaths in eight arbitrarily selected Brazilian states as they changed over time (data from Ref.(*4*); Supplementary Figure S1 gives plots for all other states). The SARI deaths are considered a proxy of true COVID mortality(*2, 17*). We note that mortality data should reflect trends in disease dynamics with reasonable reliability and provides a useful reference frame, particularly when the quality of other surveillance datasets might be in question, as is often the case. The graphs indicate that the epidemic curve of the first wave was roughly synchronized over all of Brazil in the period May-July. All 27 regions peaked in this 2-month period, some earlier than others, and then declined within several months. The differences in timing between states presumably relate to spatial spread and differences in local NPI policies in terms of initiation and intensity as well as differences in local physical circumstances. The outbreaks appear to have turned around due to the impact of mitigation procedures (NPIs including social-distancing, facemasks, hand-sanitizing, work stoppages, lockdowns, travel bans, restrictions on gatherings, etc.) and not the attainment of herd immunity This has become obvious if only because of the second wave that appeared across Brazil since late last year. Even in Manaus, which was one of the worst impacted cities, the crash of the epidemic in late April was recognized to be due to the interplay of NPIs and loss of immunity, according to the findings of Buss et al (2021)., and not from having reached herd immunity. On the other hand, at the same time it was argued that the epidemic in Manaus was "largely unmitigated", a point we will return to.

In order to explore the dynamics in more depth, an SEIR-type epidemiological model (see SM1) was fitted to the SARI mortality datasets and reconstructed the epidemic trajectory over time compartmentalizing the population into mutually exclusive classes based on epidemiological states (i.e., Susceptible, Exposed, Infected, Hospitalized, Recovered and Dead). This allowed estimation of $R_0(t)$ as it changed over the epidemic, as shown in Figures 1&2. For Manaus, all analyses were repeated based on the local government's updated COVID-19 confirmed mortality (CCM) dataset (*18*), but similar outcomes were always obtained for both data sets. (The CCM appears to have some under-reporting. A comparison of the two datasets in Manaus and Amazonas in SM Figure S2.) In more formal terms, the reproductive number $R_0(t)$ is defined as the average number of secondary infections a typical infectee can generate over the period of the disease, at time *t*, in a fully susceptible population. (Note that $R_0(t)$ should not be confused with the effective reproductive number $R_e(t) = R_0(t)S(t)$ mentioned in our Supplementary Material, which is not plotted here.)

As seen in Figures 1 and S1, in all states of Brazil, $R_0(t)$ fell substantially (dashed blue line) with time over the first wave as the epidemic proceeded, and it appeared to be a regional phenomenon. In line with Brauer,(*9*) the reduction in $R_0(t)$ is thus likely to be due to mitigation practices such as state-wide NPI's, individual behavioural changes, and possibly also combined with seasonal and/or climatic factors. The former reduce the contact rate between people (e.g., decreasing under different forms of closure), and the likeliness of successful transmission given an encounter between an infected and susceptible person (e.g., decreasing when face masks are used, hands are washed and surfaces cleaned, and when practicing social distancing).

Fixed transmission rate fit. Turning specifically to Manaus, the model was first fitted to the mortality data with a fixed transmission rate $R_0(t) = constant$, as might be typical of an unmitigated epidemic. This is the usual configuration of the model when simulating single epidemics. Yet any such fit yielded outcomes that did not appear to be meaningful. For example, for Manaus, the best fit in Figure 2a with a constant transmission rate predicted a large fixed $R_0 = 3.1$ over the whole period through to January 2021, and provided a very poor fit to the data.

Dynamical fit. As such, we proceeded to model the mortality data with a flexible fitting of $R_0(t)$ so as to estimate and allow for changes in $R_0(t)$ as the epidemic evolved in time. Figures 2b&c show that with the inclusion of a flexible $R_0(t)$ an excellent fit of the Manaus mortality data is readily obtained. Figure 2b shows fits for CCM mortality data (from Ref.(*18*)) while Fig.2c fits SARI mortality data. Moreover, the SEIR-type model, despite being acclaimed for its ability to fit epidemics of all types, was unable to do so unless $R_0(t)$ decreased rapidly in April, thereby allowing a low stationary endemic state to be maintained from July to November, as was actually the case.

Demographically structured model fits. For the attack rates, the model estimated that by October 2020 in Manaus, the AR=30% (single age-class model) and AR=34% (two age-class model), as based on SARI mortality data. This is far less than that found in recent studies(*2, 19*), and far less than the herd immunity threshold estimated as 60-66% in Ref.2. It would nevertheless appear to be reasonable given the large ongoing second wave that re-emerged in December only a few months later. Note that these results were obtained before there was even awareness of the existence of a second wave, and our model had in fact predicted the occurrence of a new wave with any significant weakening of NPI practices. These estimates of AR are also in-line with the work of Mellan et al.(*20*) and Hallal et al.(*21*) who reported a far lower attack rate than Ref.2. (The seroprevalence survey of Hallal et al.(*21*) reported an IAR 14.6% with 95% confidence interval CI: (8.9%, 22.1%) by June 7, 2020, while our estimated IAR is 20% by this time, well in the CI.)

As a test of the methodology which is a Bayesian method based on iterated filtering (*22*), we generated three fictitious time-series from March 2020 to January 2021 with $R_0(t)$ arbitrarily varying in time $0.5 < R0t < 2.2$ (Fig.3a,b&c, blue dashed curve). These time-series were used to drive the model parameterised for Manaus in order to generate dummy datasets of daily mortality numbers (red trajectory with circles). We could then use our method to reconstruct $R_0(t)$ (plotted in green). The results show that the reconstructed $R_0(t)$ is a very close estimate of the original value (dashed blue) in each case. Similar to Figs 1&2, the black curve and grey regions, were based on simulations of the SEIRD model using the same reconstructed $R_0(t)$. This confirms the capability of the method to reconstruct $R_0(t)$ and that our analysis of the Manaus data is reliable. Variations of the methodology has been tested in other contexts in our own work and by several independent teams (*15, 23, 24*).

In the process of calculating the AR, our model takes into account dynamical features of the epidemic trajectory over time. The evolving shape of the trajectory itself is a source of information on the underlying dynamics that took place at the time. In this respect this approach provides a higher level of sophistication relative to gross calculations based for example on IFR estimates and seroprevalence data, particularly in cases where the latter may potentially be prone too sampling biases. According to Buss et al.(*2*), their high estimate of AR~76% based on seroprevalence data, implies that infection fatality rate (IFR) must be extremely low, in order to maintain consistency (on average IFR~0.018-0.26%). However, we were unable to fit mortality dynamics sensibly when assuming such a low values of IFR (see Figure S3). The authors attribute the high AR they found for Manaus, to its low IFR, one that was claimed to be lower by almost 40% than São Paulo. But SM Figure S5a&b shows that this seems to contradict what is seen in the ASFR (Age SARI Fatality rate which should be a proxy for the IFR) data; the ASFR of Manaus is larger than most other cities. This conflicting information makes the comparison between Manaus and São Paulo complex and insights regarding the differences in AR inconclusive.

Modeling the full two-waves of the Manaus epidemic with an explicit two strain model. Following the rapid explosion of the second wave of COVID-19 in December 2020, a new highly transmissible P.1 lineage of

SARS-CoV2 was identified. Initially it had been described at around January 11, 2021, but the same lineage was also later identified in samples that date back to early December 2020, most likely already present in November 2020(*25*). We therefore attempted to also model the full two-waves of the Manaus epidemic using an explicit two-strain model. Given the significantly low cases of re-infection found to data, we assumed the strains share full cross immunity, such that recovery from infection by one of the strains, precludes becoming infected by the second strain. Structurally, the second strain was integrated into our original model by adding compartments that explicitly represent the different epidemiological states associated with the P.1 strain (i.e., Exposed, Infectious, Hospitalized, Recovered and Dead). To facilitate the partitioning of the model fit between two strains we used data made available from Ref.(*25*) that gave the relative proportion of P.1 cases among all cases as a function of time based on genomic analyses (see Fig.4b). The goal was to check whether both simulated daily mortality data would match observed data and simulated strain proportion would match reported strain proportion as a function of time. Figure 4a shows the results of the two-strain model fit to the SARI data up until March 1 2021. The analysis shows that 30% of the population was infected by October 2020, while 67% was infected by the end of February 2021. To the best of our knowledge, this is the first model for which the two waves of the observed data falls within the 95%CI of the model simulations (cf. Refs (*25, 26*)). A good fit to the population data was only possible if the P.1 transmission rate was 2.2-fold that of the non-P.1 strain (*25, 26*) as seen in Figs 4b,d and similar to estimates in (*25, 26*). The fit of the genomic data is at least as good as or better than other attempts (*25, 26*). Finally we note that this model did not take into account that during the second wave of the Manaus epidemic, the proportion of those infected in the population of younger age (also having a lower IFR) significantly increased. This would tent to raise the Attack Rate making our estimate of AR=67% conservative.

In summary, the results reported here quantitatively confirm that the AR in Manaus was significantly over-estimated, and could not have been close to 76% by October 2020 (Ref.2). Our model estimates an attack rate of AR~30% at this time, and thus far below herd immunity. The intervention activities implemented in Manaus must have played a role in the crash of the first wave seen in early May 2020, as also argued in Refs. 2,17 & 25. It appears that the implementation of NPIs provided temporary “herd-protection”, despite the notable political and infrastructural challenges faced by the residents of Manaus, and the poor and often counter-productive response of the national authorities and their mishandling of mitigation efforts. Had the epidemic been unmitigated, herd immunity would have been reached at the peak of the epidemic late April 2020, whereas our results indicate that herd immunity could only have been achieved well after December 2020, that is, well into the second wave. Relaxation of mitigation efforts in late 2020 combined with the appearance of the highly transmissible P.1 variant are important factors that led to the triggering of the second wave, and the large numbers of deaths that rapidly followed. Through modelling, these conclusions were reached before the second wave was apparent (*27, 28*).

Our work highlights the difficulties policy makers face when attempting to predict herd immunity, as well as the importance of developing and using complementary mechanistic epidemiological modelling tools when conducting such data analyses. The methods presented here are particularly useful in cases where NPIs and human behaviour impacts the temporal development of epidemiological dynamics, for which other traditional methods can fail. The experiences in Manaus with regard to COVID-19 make clear the need for future research characterizing the nature of reinfection and protective immunity, as they are currently key factors limiting our understanding of the epidemiological dynamics at the level of the host population.

Declarations

Ethical approval and consent to participate are not applicable.

Consent for publication

Not applicable.

Availability of data and materials
All data used are publicly available.
Code availability
The codes are available upon request to the corresponding author.
Conflict of interests
DH was supported by an Alibaba (China) Co. Ltd Collaborative Research Project. Other authors declared no conflict of interest.
Funding
DH supported by Hong Kong Research Grants Council Collaborative Research Fund and Alibaba (China) Co. Ltd. Collaborative Research Project (ZG9Z). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions
LS, DH, YA designed the study, carried out the statistical analysis and wrote the manuscript. SSM participated in data analysis and critically revised the manuscript.

REFERENCES

1. O. T. Ranzani *et al.*, Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data. *Lancet Respiratory Med*.
2. L. F. Buss *et al.*, Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288-292 (2021).
3. E. C. Sabino *et al.*, Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452-455 (2021).
4. *Brazil Ministério da Saúde Opendatasus SRAG 2020* (2020 https://opendatasus.saude.gov.br/gl/dataset/bd-srag-2020/resource/d89ea107-4a2b-4bd5-8b8b-fa1caaa96550).
5. F. Naveca *et al.*, Three SARS-CoV-2 reinfection cases by the new Variant of Concern (VOC) P. 1/501Y. V3. (2021).
6. *Ministério da Saúde*. *Boletins Epidemiológicos coronavírus - N55* (2021 https://www.gov.br/saude/pt-br/media/pdf/2021/marco/25/boletim_epidemiologico_covid_55_atualizado.pdf).
7. A. Backhaus, Common Pitfalls in the Interpretation of COVID-19 Data and Statistics. *Intereconomics* **55**, 162-166 (2020).
8. I. Amigo, in *The Scientist*. (New York, 2020).
9. F. Brauer, The final size of a serious epidemic. *Bull Math Biol* **81**, 869-877 (2019).
10. C. L. Althaus, Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Currents* **6**, (2014).
11. L. Stone, D. He, S. Lehnstaedt, Y. Artzy-Randrup, Extraordinary curtailment of massive typhus epidemic in the Warsaw Ghetto. *Science Advances* **6**, eabc0927 (2020).
12. S. Zhao, L. Stone, D. Gao, D. He, Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Negl Trop Dis* **12**, e0006158 (2018).
13. G. Chowell, C. Viboud, L. Simonsen, S. M. Moghadas, Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J Royal Soceity Interface* **13**, 20160659 (2016).
14. T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846-849 (2020).

15. A. A. King, E. L. Ionides, M. Pascual, M. J. Bouma, Inapparent infections and cholera dynamics. *Nature* **454**, 877-880 (2008).
16. D. He, E. L. Ionides, A. A. King, Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J Royal Society Interface* **7**, 271-283 (2010).
17. W. M. de Souza *et al.*, Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behaviour* **4**, 856-865 (2020).
18. *Amazonas and Manaus daily COVID-19 confirmed death data* (2021 *http://www.amazonas.am.gov.br/content/uploads/2021/01/20_01_21_BOLETIM_DIARIO_DE_CASOS_COVID-19-12.pdf*).
19. T. Prowse *et al.*, Inferred resolution through herd immmunity of first COVID-19 wave in Manaus, Brazilian Amazon. *medRxiv*, (2020).
20. T. A. Mellan *et al.*, Subnational analysis of the COVID-19 epidemic in Brazil. *medRxiv*, (2020).
21. P. C. Hallal *et al.*, SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Global Health* **8**, e1390-e1398 (2020).
22. E. L. Ionides, C. Bretó, A. A. King, Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103**, 18438-18443 (2006).
23. A. Camacho *et al.*, Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proceedings of the Royal Society B: Biological Sciences* **278**, 3635-3643 (2011).
24. D. He, J. Dushoff, T. Day, J. Ma, D. J. Earn, Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology* **4**, 283-288 (2011).
25. N. R. Faria *et al.*, Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *medRxiv*, (2021).
26. R. M. Coutinho *et al.*, Model-based evaluation of transmissibility and reinfection for the P. 1 variant of the SARS-CoV-2. *medRxiv*, (2021).
27. L. Stone, Cambridge University seminar: "Epidemics and Genocide: Mathematical modelling helps understand the connection. (2021).
28. D. He, Y. Artzy-Randrup, S. S. Musa, L. Stone, The unexpected dynamics of COVID-19 in Manaus, Brazil: Herd immunity versus interventions. *medRxiv*, (2021).

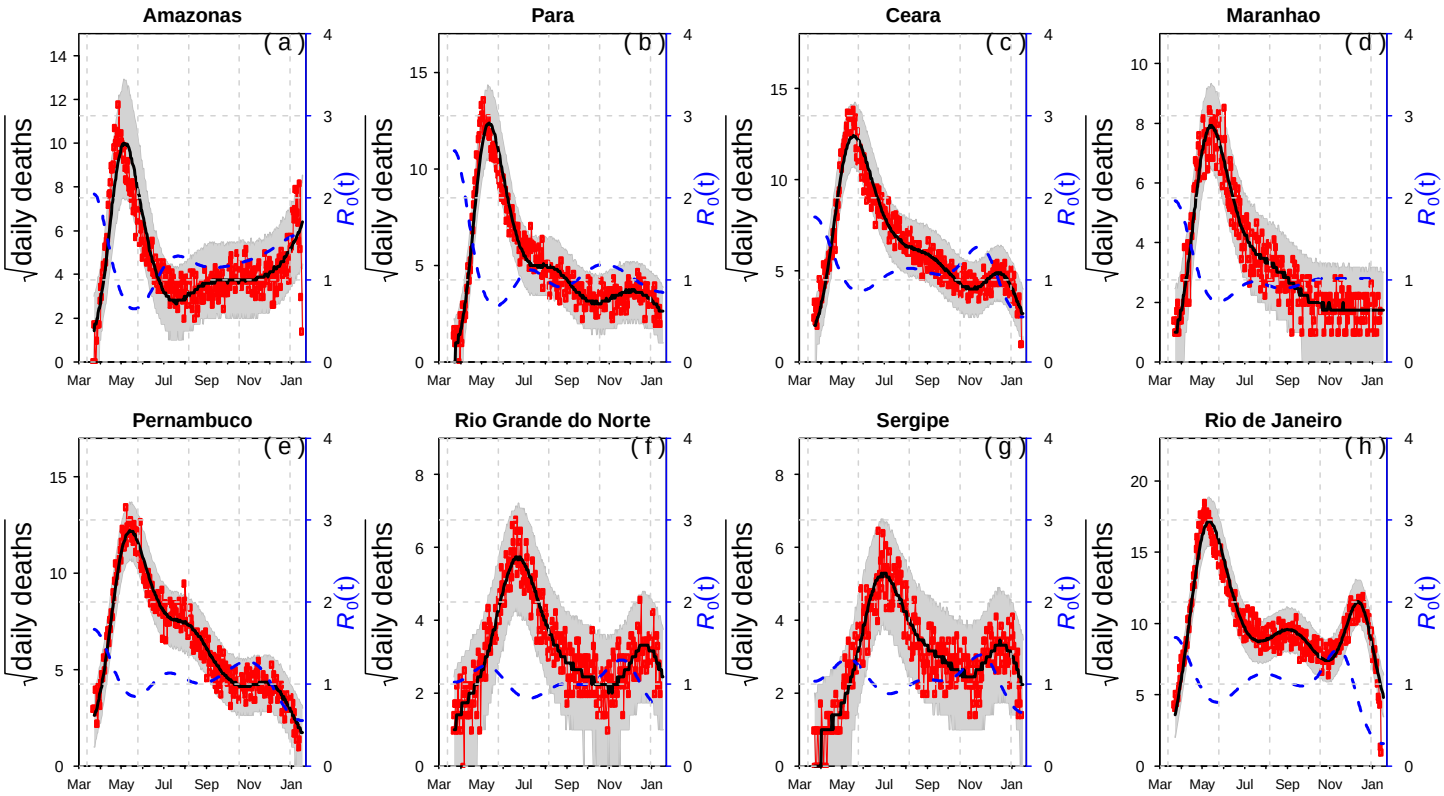


Figure 1. An SEIRD model is fitted to reported daily SARI deaths in eight Brazilian states(*4*). The number of nodes in transmission rate spline was set at n_m=7. Red points represent real data in terms of individuals dying per day, the black line is the median of model simulations of the mortality data, and the grey shading indicates the 95% range based on 1,000 simulations.

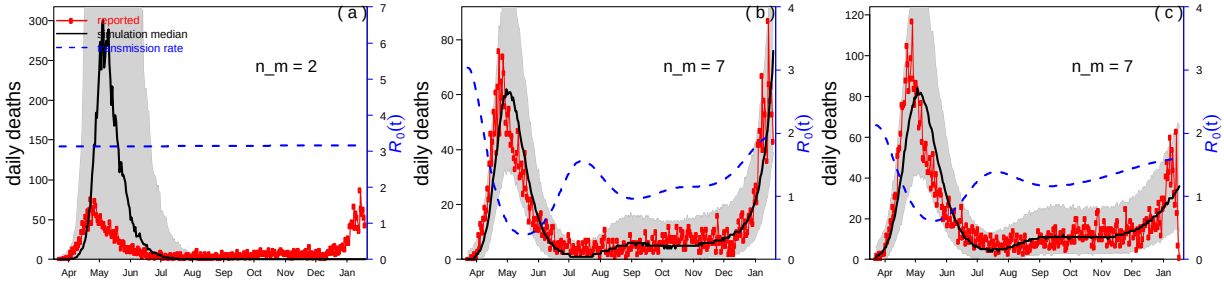


Figure 2 The SEIRD model was used to fit reported daily Covid-19 confirmed deaths in Manaus with a flexible transmission rate $R_0(t)$ (blue dashed line). The black line is the median of model simulations of the mortality data, and the grey shading indicates the 95% range based on 1,000 simulations. (a) A constant transmission rate (constant R_0) fails to fit the major wave well (panel a; number of nodes in transmission rate spline nm=2). Data from Ref.(*18*). (b) When the spline's number of nodes was increased to 7, this gave a flexible R0(t), and the fitting performance was improved. Based on the 2nd-order Akaike Information Criterion (AICc), the best performance was found for nm=7. Data from Ref. (*18*). (c) Same as panel b but based on SARI mortality data in Manaus from(*4*) (red points). In the first phase of the epidemic, our estimated R_0 for Manaus is

2.250 95% CI(1.79, 2.79) (see SM) with data from.(*18*) Again, the red points represent real data in terms of individuals dying per day, the black line is the median of model simulations of the mortality data, and grey shading provides the 95% range based on 1,000 simulations.

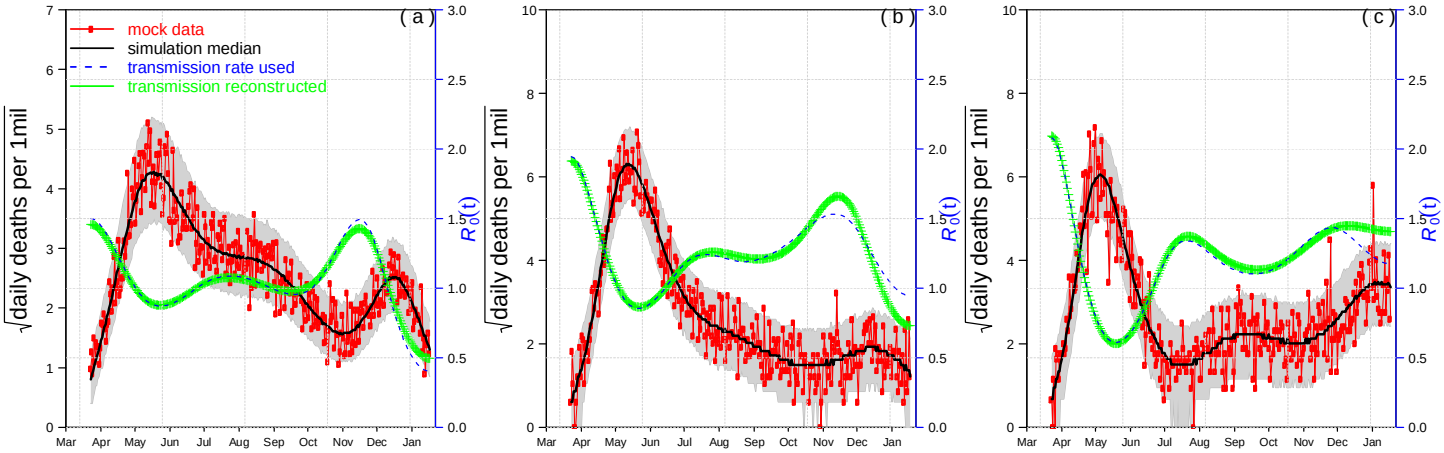


Figure 3. Fitting $R_0(t)$ from mock data. We generated three simulated time series of $R_0(t)$ changing over time in an arbitrary way between March 2020 and February 2021 (blue dashed lines). Each time series of $R_0(t)$ was used to generate a simulation of mortality data from the SEIRD model (mock data, red trajectory with circles), and basically three different patterns of death data resulted as seen in (a), (b), (c). Then we used our fitting method to reconstruct $R_0(t)$ (green trajectories) from the mock data. The solid black lines and grey regions, similar to Figs 1&2, were based on simulations using the reconstructed $R_0(t)$. The reconstructed $R_0(t)$ proves to be a good estimate of the initial $R_0(t)$.

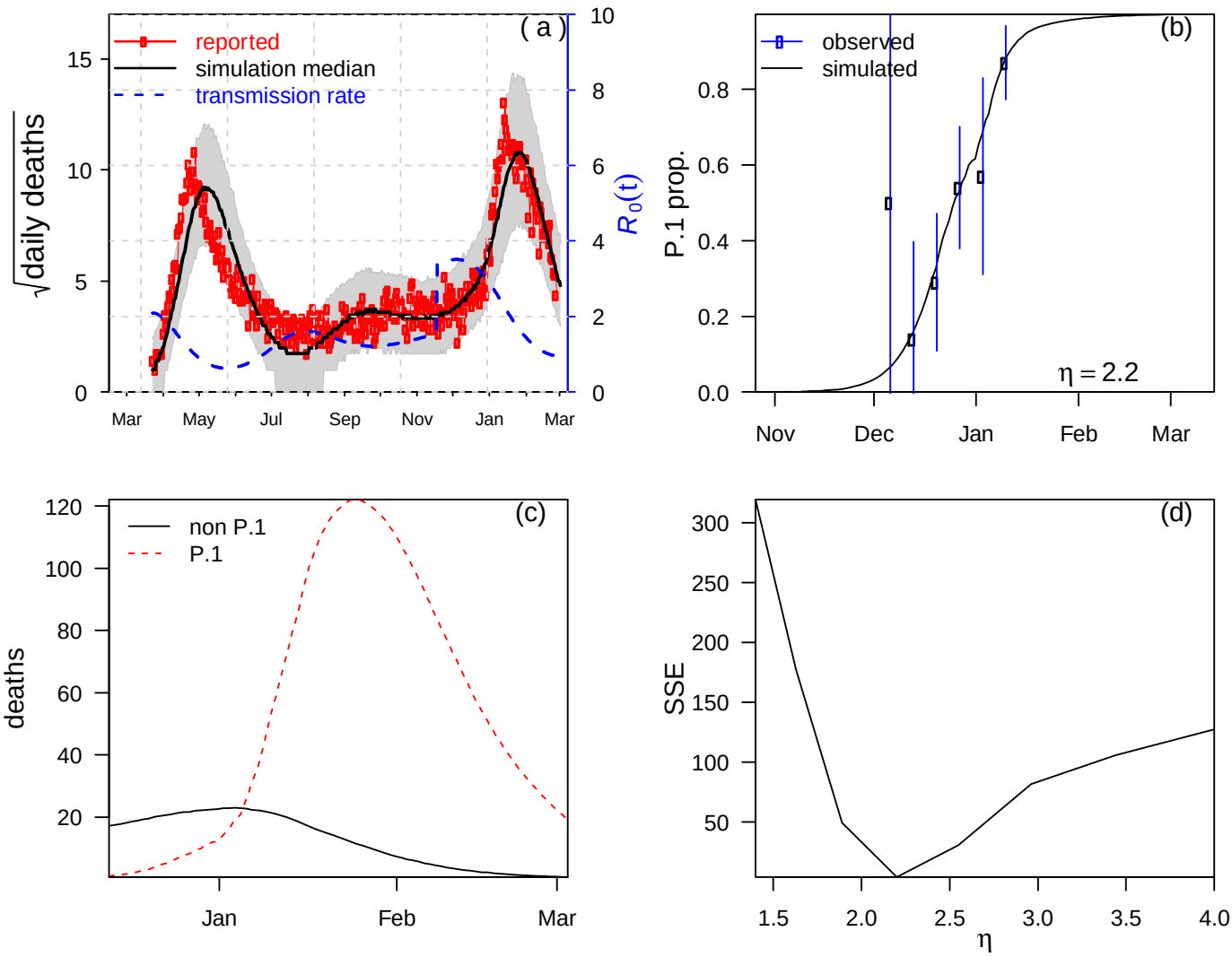


Figure 4 Fitting a two-strain model to SARI mortality data and sequencing proportion of P.1 strain simultaneously. η denotes the relative transmission of P.1 variant compared to non-P.1 strain. $\eta = 2.2$ yields the best fit for both SARI deaths (a) and sequence proportion (b). Panel (c) shows the simulated daily deaths due to P.1 and non-P.1 strains. Panel d shows the sum of squared error (SSE) based on sequence proportion data (b) as function of η. Smaller or larger η leads to mismatching between simulated and observed sequence proportion.

Supplementary Material

The unexpected dynamics of COVID-19 in Manaus, Brazil: Was herd immunity achieved?

Daihai He[1], Yael Artzy-Randrup[2], Salihu S. Musa[1], Lewi Stone[3,4,*]

[1] Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China.

[2] Department of Theoretical and Computational Ecology, IBED, University of Amsterdam, Amsterdam, Netherlands

[3] Mathematical Sciences, School of Science, RMIT University, Melbourne, Australia

[4] Biomathematics Unit, School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

[*] Correspondence author: lewistone100@gmail.com

1. The SEIHDR epidemic model
2. Fitting the model to SARI mortality data in all Brazilian states
3. Comparison of confirmed Covid-19 mortality and SARI mortality in Amazonas and Manaus
4. Fitting data with IFR<0.3% and Attack Rates greater or equal to AR=76% by Oct 31, 2020
5. Initial Reproductive number $R_0(t)$
6. IFR and AR in Manaus versus São Paulo
7. Fitting Two-age-group model to SARI mortality data including both waves
8. Reinfections
9. Blood donor data

Section 1: The SEIHDR epidemic model

We use an susceptible-exposed-infectious-recovered (SEIR) type model with a flexible time-varying transmission rate $\beta(t)$ to the reported mortality data:

$$\begin{aligned}
\dot{S} &= -\frac{\beta SI}{N}, \\
\dot{E} &= \frac{\beta SI}{N} - \sigma E, \\
\dot{I} &= \sigma E - \gamma I, \\
\dot{H} &= \theta\gamma I - \kappa H, \\
\dot{D} &= \pi\kappa H, \\
\dot{R} &= (1-\theta)\gamma I + (1-\pi)\kappa H.
\end{aligned}$$

Here the compartments *S*, *E*, *I*, and *R* denote the conventional susceptible, exposed, infectious, and recovered individuals, respectively. The constant *N* is the total population size of the city or state, *H* denotes the hospitalized cases (or severe cases), and *D* denotes the total of SARI-deaths. Parameters $\beta(t), \sigma, \gamma, \kappa$ denote the transmission rate, the infectiousness emergence rate, the infectiousness disappearance rate, and the removal rate (due to death or recovery) of hospitalized cases, respectively. Parameters θ and π denote the ratio of hospitalized cases out of all infected cases and the proportion of deaths out of hospitalized cases, respectively. Thus, the overall case fatality rate (or infection fatality rate) equals $\theta\pi$. All parameters are constant except $\beta(t)$ being time-varying. The above model was used for studying the Amazonas SARI mortality data(*4*) and a very similar model was used to study Manaus COVID confirmed mortality data.(*18*)

We follow closely our previous work for fitting a "flexible" time varying transmission rate $\beta(t)$ (*11, 12, 29*) based on fitting the mortality data. The former requires defining $\beta(t) = \exp\,(cubic_spline)$ as an exponential cubic spline with n_m nodes evenly distributed over the study period. After this was fitted, we estimated $R_0(t) = \frac{\beta(t)}{\gamma}$. For the model, we set time step size as one day and integrated $\dot{D}$ for one day to obtain the simulated daily deaths D_t. We defined the reported deaths as C_t, where:

$$C_t \sim \text{NegativeBinomial}\big(\text{mean} = D_t, \text{variance} = D_t(1+\tau D_t)\big).$$

Here, τ denotes the overdispersion, and accounts for the measurement noise due to surveillance and heterogeneity among individuals.

The parameter values of σ, γ, κ were taken as 365/2, 365/3.5, and 365/12 per year, respectively. Thus, the mean latent period and mean infectious period (the reciprocals of σ and γ) were set as 2 and 3.5 days, respectively. The generation time GI,(*30*) the sum of mean latent period and mean infectious period, equals 5.5 days, which is in line with three key studies which provided estimated GI.(*31-33*) According to,(*34*) the delay between the first symptom onset and the death is ~14 days, which justifies the choice of 12 days (the reciprocal of κ) from loss of infectiousness to death. Namely the onset of infectiousness was taken to be 2 days ahead of first symptom onset.
In addition, we assume 5% proportion of the population has pre-existing immunity. (*35, 36*) , which only changes the initially large available susceptible pool by only a small amount.

The Euler-multinomial simulation approach was used to simulate our model with a time-step size 1 day. The popular iterated filtering method(*12, 15, 16*) (note that at least 25 applications use this method, see a list at

https://en.wikipedia.org/wiki/Iterated_filtering) was implemented to fit our model to the observed data and to obtain the maximum likelihood estimates of unknown parameters, including values of nodes in $\beta(t)$ and π while holding θ=0.1. Namely, we assumed 10% of cases were hospitalized (or severe cases), as widely reported. The fitting is insensitive to the choice of θ, but sensitive to the product of π and of θ, which is equivalent to the infection fatality rate (IFR). We assumed uniform prior for all parameters. We hypothesized that the IFR lay in the interval (0.002, 0.008), and fit the daily SARI mortality data with a flexible transmission rate. The method finds the best fitting parameters including IFR. We used the 2nd-order Akaike Information Criterion (AICc) (*11, 12, 29*) to find the best number of nodes in the transmission spline n_m. The same method was used to fit a two-age-class model

Section 2: Fitting the model to SARI mortality data in all Brazilian states

In Figure 1 of the main text, we fit the above model to SARI mortality data in eight Brazilian states. Figure S1 below repeats this for the other 16 Brazilian states and the federal district.

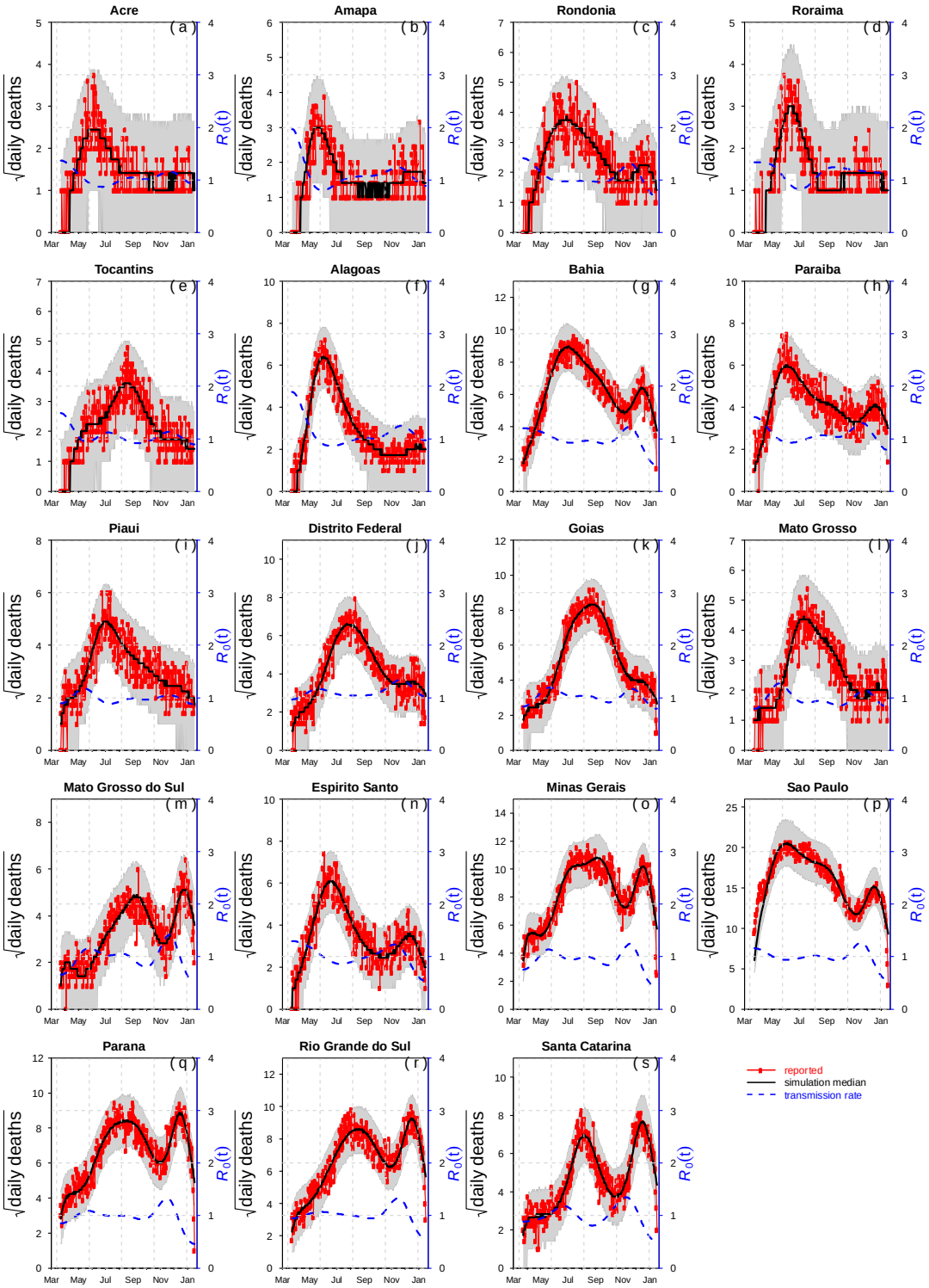


Figure S1. Fitting SEIRD model to daily SARI death in 16 Brazilian states and the federal district.

Section 3: Comparison of confirmed Covid-19 mortality and SARI mortality in Amazonas and Manaus

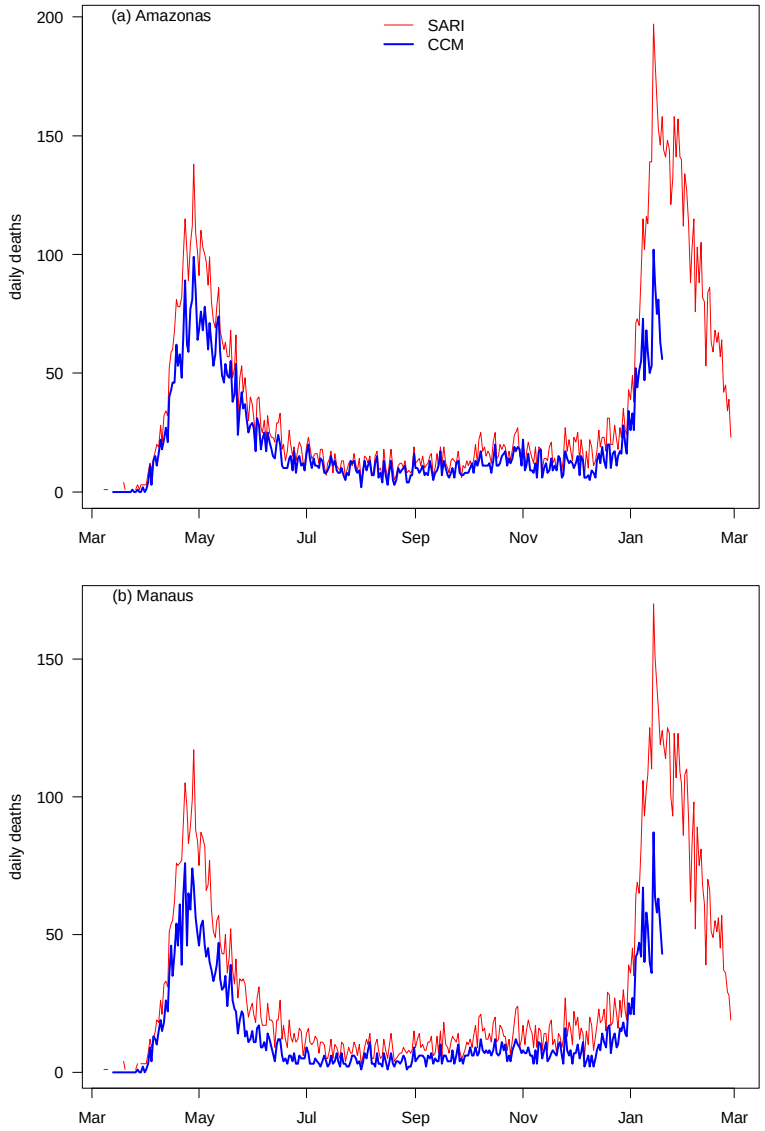


Figure S2. Comparison of confirmed Covid-19 mortality data (CCM) and SARI mortality data in Amazonas and Manaus. We stopped updating the CCM data in early January, 2021.

Section 4: Fitting data with IFR<0.3% and Attack Rates greater or equal to AR=76% by Oct 31, 2020

In Figure S3, we explored simulations with IFR<0.3 ensuring that the Attack Rate (AR) reach levels of AR >76%, based on deaths and population size in Manaus. In all cases this assumption led to $R_0(t)$ being unrealistically large (i.e., $R_0(t) > 7$) in the second wave. If $R_0(t)$was fixed to a constant, the fit to the data was poor (Figure S3a), with the best fitting $R_0(t)$ reaching values of $R_0(t) > 5$.

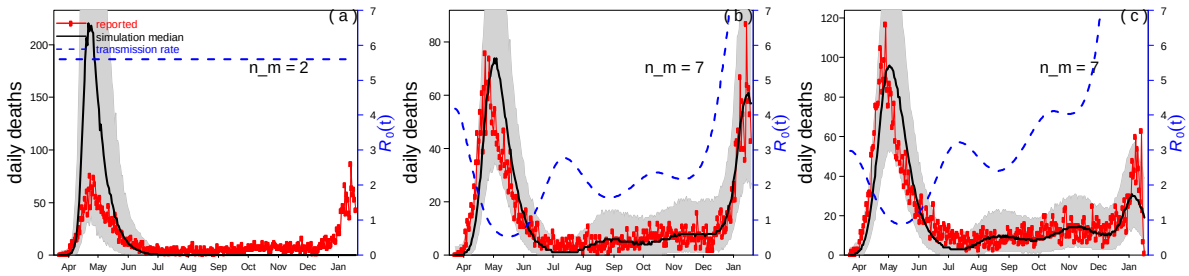


Figure S3. Reproducing the fitting in Figure 2 with IFR <0.3%. The estimated $R_0(t)$ reaches unreasonably high values during the second wave. (a,b) confirmed COVID-19 deaths. (c) SARI deaths, in Manaus.

RMIT Classification: Trusted

Section 5: Initial Reproductive number $R_0(t)$

For each dataset in Fig. S1 and in the main text we plotted $R_0(t) = \frac{\beta(t)}{\gamma}$, which is the dashed blue line seen in figures as a function of time (/date). The reproductive number $R_0(t)$ should not be confused with the effective reproductive number $R_e(t) = R_0(t)S(t)$ mentioned in the main text, which we do not plot. The plot of $R_0(t)$ allows us to visualise how the reproductive number changes e.g., as a result of behaviour, or interventions.

For Amazonas and Manaus, we also arrived at estimates for the initial reproductive number in the first phase of the epidemic, using the standard technique of Wallinga and Lipsitch.(*37*) This is shown in Figure S4 which is based on the COVID-19 mortality data from Amazonas government health department reports.(*18*)

In Figure S4, we used a standard approach, $R_0 = \frac{1}{M(-r)}$, where M is the moment generating function of the distribution of the generation time which attains a Gamma distribution with a mean of 5.5 days and s.d. 3 days.(*31, 33*) We used the interval from March 30, 2020 (when there were 2 deaths) to April 12, 2020.

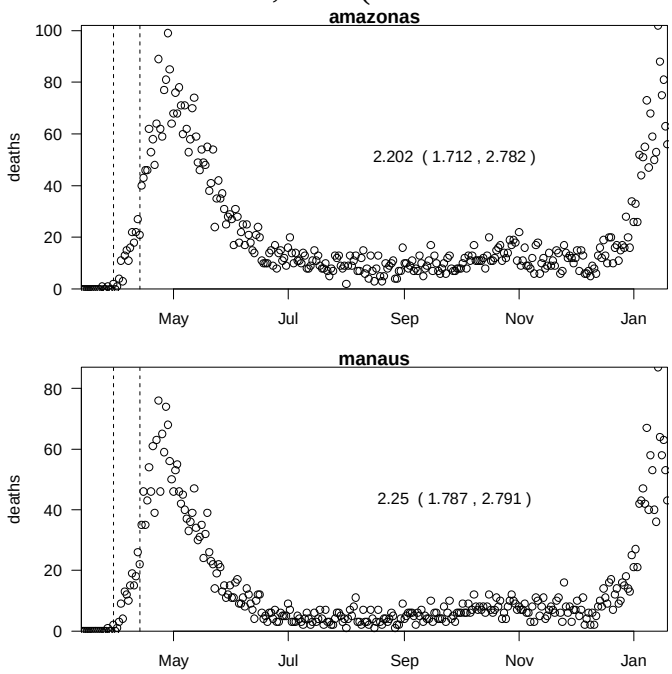


Figure S4. Estimating initial R_0 via the method of Wallinga and Lipsitch (2007)[1]. The dashed vertical lines indicate the window in which R_0 was estimated. Initial R_0 in Amazonas is 2.20 with 95%CI (1.71, 2.78). Initial $R_0(t)$ in Manaus is about 2.25 with 95% CI (1.79, 2.79), which are in line with.(*17*) Data are from [2].

RMIT Classification: Trusted

Section 6: IFR and AR in Manaus versus São Paulo

Buss et al. argued that the AR in Manaus (76%) was far larger than that of São Paulo (29%). They claimed this was due to the much lower global population infection fatality rate (IFR) in Manaus, an outcome of the age-class distribution of the IFR in a city with relatively younger population. They found the IFR in Manaus was 38.9% smaller than that of São Paulo. However, this difference between Manaus and São Paulo is questionable since it is not seen when examining the Age SARI fatality rate (ASFR), i.e., the ratio of deaths among reported SARI cases. In Figure S5 we plot the ASFR for four cities in Brazil including São Paulo, and find the ASFR is higher in Manaus in both waves compared to any of the other three cities. In the first wave it is particularly larger than the ASFR of São Paulo for all ages. This was also evident in plots of Buss et al.(*2*) in their SI.

Given the "inverse nature" of the calculations used in working with the IFR, if Manaus actually had a higher IFR than São Paulo as Figure S5 indicates, this should translate into a lower number of infected residents [for similar mortality numbers] and lower AR in Manaus, indicating another inconsistency with the suggested AR of 76%.

Also Figure S5 shows that although the population in Manaus is relatively young, we see that those of younger ages that were diagnosed as SARI, were more likely to perish in Manaus relative to São Paulo. In Manaus, the young population presumably suffered more due to the heavy burden of the epidemic conditions and the collapse of the hospital system.

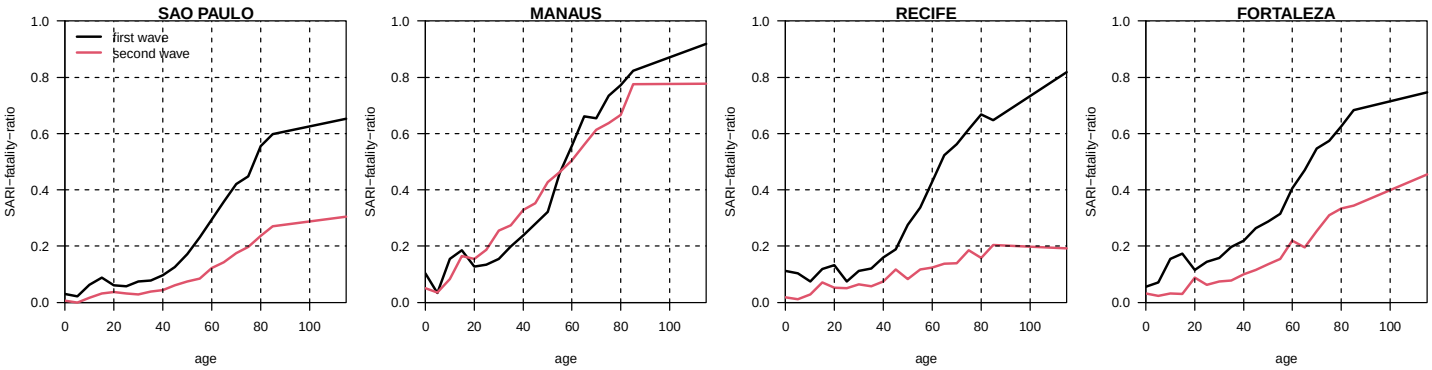


Figure S5. Age SARI fatality ratio (ASFR), i.e., the ratio of deaths among reported SARI cases. Black curve shows the ASFR for the first wave (March 13 - June 1, 2020). Red curve shows the ASFR for the second wave (Nov 23, 2020 – Feb 8, 2021). Data from Ref. (*4*).

RMIT Classification: Trusted

Section 7: Fitting Two-age-group model to SARI deaths including both waves

Age-class model rationale Our initial results were obtained from analysing an epidemic model with only a single age-class. While this ignores important considerations of the age distribution, it nevertheless has relevance for two reasons: a) The important *Science* paper of Earn et al.(*38*) demonstrated that significant changes in $R_0(t)$ have much more impact on the epidemic trajectory than age-structure and should be the first consideration, while the effects of age-structure are in comparison far smaller. b) Qualitatively, Manaus can be characterized as a two age-class system, with people aged 45 and up significantly contributing to COVID-19 related deaths (these account for roughly 20% of the population), while the rest of the population (< 45 years) display an exceptionally low IFR, contributing very little to the total deaths toll even though the account for roughly 80% of the population (e.g., according to Levin et al. (2021)(*39*), they are responsible for <7% of total COVID-19 deaths. We justified our approach based on the knowledge that in contexts similar to the one dealt with here, a single age-class system found from averaging two age-classes (themselves homogenous in transmission) should be able to characterise the epidemic dynamics of the two age-class system (see e.g., Magpantay et al.(*40*)). Nevertheless, as a check on our work, we repeated all analyses on a complete two age-class model as described later. The results were always found to be very similar to the single age-class model, as we point out in the main text and below.

In the two-age-class model, the population in each city was divided into two age groups: <50 and 50+ years of age. The population of each group was then divided into six S-E-I-H-R-D sub-groups. It was assumed that the cross-group transmission is the same as the in-group transmission. We obtained population size and age structure (the population ratio between <50 and +50) from online sources, e.g.,

https://www.citypopulation.de/en/brazil/regiaosudeste/admin/são_paulo/3550308__são_paulo/
https://www.citypopulation.de/en/brazil/regiaonorte/admin/amazonas/1302603__manaus/

The daily Severe Acute Respiratory Syndrome (SARI) deaths was aggregated into two age groups: <50 and 50+ years of age, and time series were constructed for each group. The daily deaths from our model to the two time-series were fitted simultaneously. The total log likelihood of the model was taken to be the sum of the two log likelihoods for each time series.

The infection fatality ratio (IFR) was taken to be different for the two groups. In particular, the IFR for <50 group was much smaller than that of the 50+ group. Both IFRs were estimated by our procedure. It was only assumed that the IFR for the 50+ group was between 1% to 4.5%. We assumed a flexible transmission rate (exp-cubic-spline with 7 nodes). Parameters were estimated including IFRs for both age groups, number of nodes in the transmission rates, and initial conditions. For Manaus, Figure S6 shows the fitting performance in terms of log likelihood of the two-age-class model given the data, as a function of the IFR for the 50+ group and that for the single-age-model.

RMIT Classification: Trusted

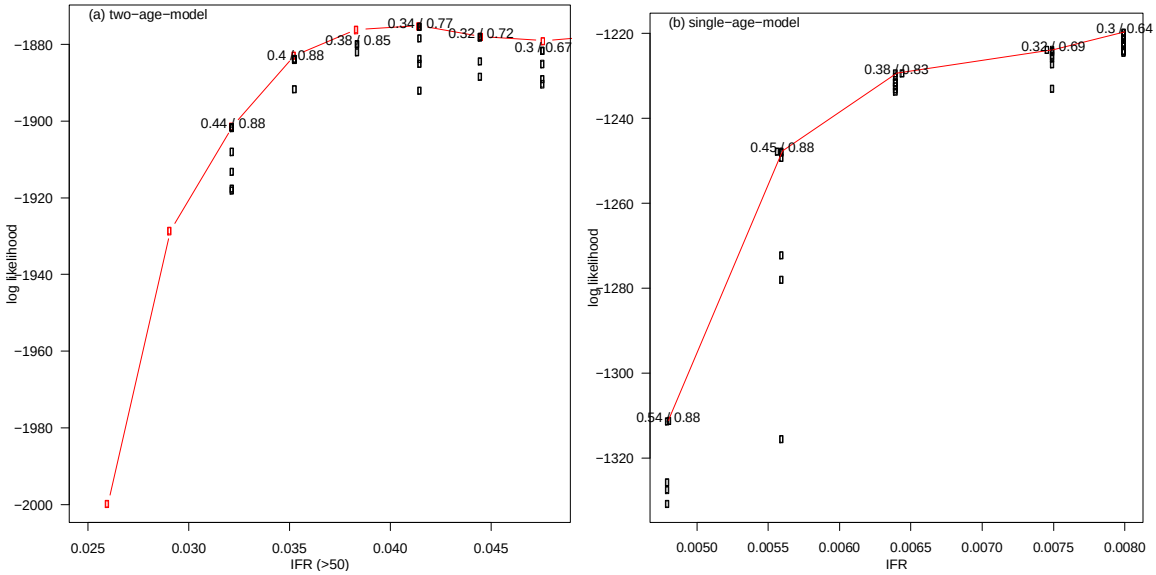
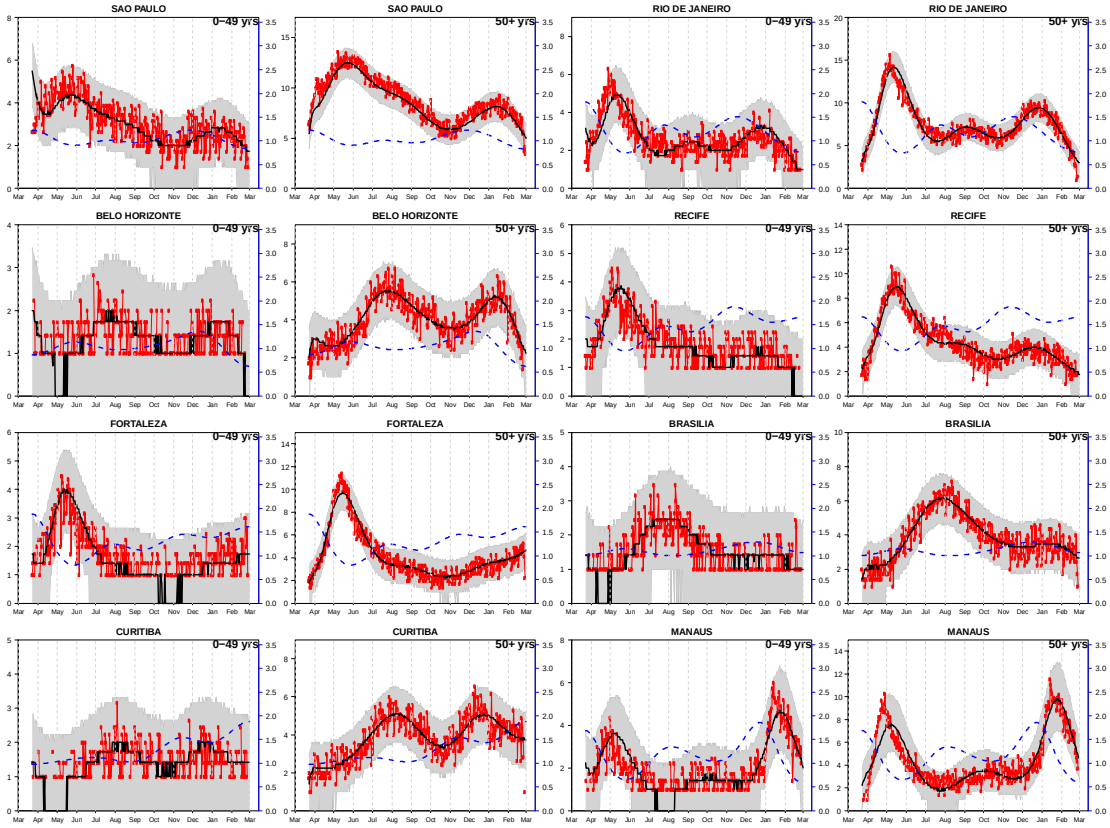


Figure S6. The fitting performance is plotted in terms of the model's log likelihood given the SARI deaths as a function of the IFR for the 50+ group in (a) and of the IFR for the single-age-model in (b) in Manaus. Larger Log likelihood implies a better fit. For each of the six IFR values plotted, we give as a pair of values the infection attack rate by Oct 31, 2020 and Feb 28, 2021. The maximum likelihood estimates of infection attack rate is thus 34% and 77% with IFR for 50+ at 4.15%. Since the peak of the Log likelihood profile is very flat, 32%/72% to 38%/85% are also plausible.

We fit the model to data from 8 Brazilian cities and achieve reasonable fitting for each of the two age-classes as indicated in Figure S7.



RMIT Classification: Trusted

Figure S7. Fitting a two-age-group model to SARI mortality from eight Brazilian cities each divided into two age-classes.

Section 9: Reinfections

If the attack rate is AR=76% up until October 2020 and let us assume that the first wave is roughly as large as the second wave. Then a simple Venn diagram would show that one needs >50% re-infection rate among infections in the second wave in Manaus. Yet currently there is no evidence or reliable reports that indicate the reinfection rates were more than 1 percent in the second wave December 2020 -February 2021. This would appear unusual if reinfections were actually occurring at high rates. A large-scale study in Israel found a rate of 0.1% (two PCR-positive at least 100 days apart) (*41*). The other large-scale study in Denmark found a rate of 0·65% two PCR-positive at least 3 months apart). (*42*) Only 3 reinfections were found in Amazonas (see main text).

Section 10: Blood-donor data

It is difficult to explain the high estimate of the attack rate of COVID-19 in Manaus, as found by Buss et al. (2021). One possibility are the sources of bias that might be introduced from working with blood-donor data. These are not discussed in their article, and thus difficult to comment on. For example, a recent COVID seroprevalence study in Kenya(*43*) made use of "family replacement donors (FRDs) who provide a unit of blood in compensation for a transfusion received by a sick relative." This may have led to major biases in the Kenyan analysis. Replacement blood donation is a common practice in middle and low-income countries. According to the Agência Nacional de Vigilância Sanitária - Anvisa (2019)(*44*) report of Brazil, 54.3% of blood donors are repeated donors, and 37.3% of blood donors are family replacement donors in the North region. The Amazonas region may have an even higher rate of replacement donations. Screening measures ("prospective donors could not have had flulike symptoms within the 30 days before donation; had close contact with suspected or confirmed covid-19 cases in the 30 days before donation")(*45*) could prevent donation from recent symptomatic cases, but not from asymptomatic cases and past symptomatic cases (who showed symptoms beyond previous 30 days). Other biased sampling of the Manaus population was also possible. For example, during the peak of the pandemic, blood donation experienced shortage, and military members were recruited(*46*) in Amazonas. Interviews in the media report other potential sources of bias such as donors being more mobile, or receiving free COVID tests or being a nonrepresentative sample of the population(*7, 8*). The study of Buss et al., could possibly be free of all such biases, but further clarification is needed.

References

1. O. T. Ranzani *et al.*, Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data. *Lancet Respiratory Med*.
2. L. F. Buss *et al.*, Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **371**, 288-292 (2021).
3. E. C. Sabino *et al.*, Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452-455 (2021).
4. *Brazil Ministério da Saúde Opendatasus SRAG 2020* (2020 https://opendatasus.saude.gov.br/gl/dataset/bd-srag-2020/resource/d89ea107-4a2b-4bd5-8b8b-fa1caaa96550).
5. F. Naveca *et al.*, Three SARS-CoV-2 reinfection cases by the new Variant of Concern (VOC) P. 1/501Y. V3. (2021).

RMIT Classification: Trusted

6. *Ministério da Saúde. Boletins Epidemiológicos coronavírus - N55* (2021 https://www.gov.br/saude/pt-br/media/pdf/2021/marco/25/boletim_epidemiologico_covid_55_atualizado.pdf).
7. A. Backhaus, Common Pitfalls in the Interpretation of COVID-19 Data and Statistics. *Intereconomics* **55**, 162-166 (2020).
8. I. Amigo, in *The Scientist*. (New York, 2020).
9. F. Brauer, The final size of a serious epidemic. *Bull Math Biol* **81**, 869-877 (2019).
10. C. L. Althaus, Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. *PLoS Currents* **6**, (2014).
11. L. Stone, D. He, S. Lehnstaedt, Y. Artzy-Randrup, Extraordinary curtailment of massive typhus epidemic in the Warsaw Ghetto. *Science Advances* **6**, eabc0927 (2020).
12. S. Zhao, L. Stone, D. Gao, D. He, Modelling the large-scale yellow fever outbreak in Luanda, Angola, and the impact of vaccination. *PLoS Negl Trop Dis* **12**, e0006158 (2018).
13. G. Chowell, C. Viboud, L. Simonsen, S. M. Moghadas, Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J Royal Soceity Interface* **13**, 20160659 (2016).
14. T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846-849 (2020).
15. A. A. King, E. L. Ionides, M. Pascual, M. J. Bouma, Inapparent infections and cholera dynamics. *Nature* **454**, 877-880 (2008).
16. D. He, E. L. Ionides, A. A. King, Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *J Royal Society Interface* **7**, 271-283 (2010).
17. W. M. de Souza *et al.*, Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behaviour* **4**, 856-865 (2020).
18. *Amazonas and Manaus daily COVID-19 confirmed death data* (2021 http://www.amazonas.am.gov.br/content/uploads/2021/01/20_01_21_BOLETIM_DIARIO_DE_CASOS_COVID-19-12.pdf).
19. T. Prowse *et al.*, Inferred resolution through herd immmunity of first COVID-19 wave in Manaus, Brazilian Amazon. *medRxiv*, (2020).
20. T. A. Mellan *et al.*, Subnational analysis of the COVID-19 epidemic in Brazil. *medRxiv*, (2020).
21. P. C. Hallal *et al.*, SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Global Health* **8**, e1390-e1398 (2020).
22. E. L. Ionides, C. Bretó, A. A. King, Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **103**, 18438-18443 (2006).
23. A. Camacho *et al.*, Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proceedings of the Royal Society B: Biological Sciences* **278**, 3635-3643 (2011).
24. D. He, J. Dushoff, T. Day, J. Ma, D. J. Earn, Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology* **4**, 283-288 (2011).
25. N. R. Faria *et al.*, Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *medRxiv*, (2021).
26. R. M. Coutinho *et al.*, Model-based evaluation of transmissibility and reinfection for the P. 1 variant of the SARS-CoV-2. *medRxiv*, (2021).
27. L. Stone, Cambridge University seminar: "Epidemics and Genocide: Mathematical modelling helps understand the connection. (2021).
28. D. He, Y. Artzy-Randrup, S. S. Musa, L. Stone, The unexpected dynamics of COVID-19 in Manaus, Brazil: Herd immunity versus interventions. *medRxiv*, (2021).

29. D. He, S. Zhao, Q. Lin, S. S. Musa, L. Stone, New estimates of the Zika virus epidemic attack rate in Northeastern Brazil from 2015 to 2016: A modelling analysis based on Guillain-Barré Syndrome (GBS) surveillance data. *PLoS Negl Trop Dis* **14**, e0007502 (2020).
30. Å. Svensson, A note on generation times in epidemic models. *Math Biosci* **208**, 300-311 (2007).
31. J. Griffin *et al.*, Rapid review of available evidence on the serial interval and generation time of COVID-19. *BMJ Open* **10**, e040263 (2020).
32. C. A. Prete Jr *et al.*, Serial interval distribution of SARS-CoV-2 infection in Brazil. *J Travel Med* **28**, taaa115 (2021).
33. G. D. Zhao S, Zhuang Z,, C. Y. Chong MKC, Ran J, Cao P,, L. Y. Wang K, Wang W, Yang L,, H. D. a. W. MH, Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. *Frontiers Physics* **8**, (2020).
34. I. Hawryluk *et al.*, Inference of COVID-19 epidemiological distributions from Brazilian hospital data. *J Royal Society Interface* **17**, 20200596 (2020).
35. K. W. Ng *et al.*, Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339-1343 (2020).
36. Q. Lin *et al.*, A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *International journal of infectious diseases* **93**, 211-216 (2020).
37. J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences* **274**, 599-604 (2007).
38. D. J. Earn, P. Rohani, B. M. Bolker, B. T. Grenfell, A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667-670 (2000).
39. A. T. Levin *et al.*, Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European J Epidemiolgy*, 1-16 (2020).
40. F. Magpantay, A. King, P. Rohani, Age-structure and transient dynamics in epidemiological systems. *J Royal Society Interface* **16**, 20190151 (2019).
41. G. Perez *et al.*, A 1 to 1000 SARS-CoV-2 reinfection proportion in members of a large healthcare provider in Israel: a preliminary report. *medRxiv*, (2021).
42. C. H. Hansen, D. Michlmayr, S. M. Gubbels, K. Mølbak, S. 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*, (2021).
43. S. Uyoga *et al.*, Seroprevalence of anti–SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science* **371**, 79-82 (2021).
44. *Produção Hemoterápica (Hemoprod) de 2019* (2021 https://app.powerbi.com/view?r=eyJrIjoiOTVhYzk3MzctYzI4MS00YWY0LWJiM2ItMzRmZTY0ZTM4NGEzIiwidCI6ImI2N2FmMjNmLWMzZjMtNGQzNS04MGM3LWI3MDg1ZjVlZGQ4MSJ9).
45. L. Amorim Filho *et al.*, Seroprevalence of anti-SARS-CoV-2 among blood donors in Rio de Janeiro, Brazil. *Revista de saude publica* **54**, 69 (2020).
46. *Militares do Comando do 9º Distrito Naval participam de campanha de doação de sangue em Manaus* (2020 https://www.marinha.mil.br/com9dn/campanha_doacao_sangue).