



# Resolving the enigma of Iquitos and Manaus: A modeling analysis of multiple COVID-19 epidemic waves in two Amazonian cities

Daihai He<sup>ab,1</sup> , Lixin Lin<sup>c</sup> , Yael Artzy-Randrup<sup>d</sup>, Haydar Demirhan<sup>c</sup> , Benjamin J. Cowling<sup>e</sup> , and Lewi Stone<sup>cf,1,2</sup> 

Edited by Simon Levin, Princeton University, Princeton, NJ; received July 6, 2022; accepted January 17, 2023

The two nearby Amazonian cities of Iquitos and Manaus endured explosive COVID-19 epidemics and may well have suffered the world's highest infection and death rates over 2020, the first year of the pandemic. State-of-the-art epidemiological and modeling studies estimated that the populations of both cities came close to attaining herd immunity (>70% infected) at the termination of the first wave and were thus protected. This makes it difficult to explain the more deadly second wave of COVID-19 that struck again in Manaus just months later, simultaneous with the appearance of a new P.1 variant of concern, creating a catastrophe for the unprepared population. It was suggested that the second wave was driven by reinfections, but the episode has become controversial and an enigma in the history of the pandemic. We present a data-driven model of epidemic dynamics in Iquitos, which we also use to explain and model events in Manaus. By reverse engineering the multiple epidemic waves over 2 y in these two cities, the partially observed Markov process model inferred that the first wave left Manaus with a highly susceptible and vulnerable population ( $\approx 40\%$  infected) open to invasion by P.1, in contrast to Iquitos ( $\approx 72\%$  infected). The model reconstructed the full epidemic outbreak dynamics from mortality data by fitting a flexible time-varying reproductive number  $R_0(t)$  while estimating reinfection and impulsive immune evasion. The approach is currently highly relevant given the lack of tools available to assess these factors as new SARS-CoV-2 virus variants appear with different degrees of immune evasion.

epidemic | pandemic | model | fitting data | COVID

To what extent do epidemics spread through populations? This innocuous question never seemed more important than in the first months of 2020, when the World Health Organization (WHO) officially declared Coronavirus disease 2019 (COVID-19) a global pandemic, and signs of the first COVID-19 outbreaks began to appear in cities across the world. Despite decades of mathematical investigation and modeling devoted to predicting the growth and size of an epidemic, in practice, this was difficult to achieve with reliability and accuracy as events played out during the pandemic. In this paper, we are particularly interested in the large-scale epidemics that ravaged the two nearby Amazonian cities of Manaus (Brazil) and Iquitos (Peru), which serve as paradigmatic examples. They were often regarded as the two cities in the world in which COVID-19 hit hardest over the first wave, because of their extremely high infection and mortality rates (1–4). It may be recalled that beginning in March 2020, the unprepared population of Manaus faced an explosively spreading infectious disease which left large numbers of residents severely ill each day, with a high rate of mortality, over a period of 2 to 3 mo (Fig. 1). The large number of cases rendered local healthcare systems and hospitals dysfunctional, while the number of deaths became so unmanageable that mass grave sites became the only practical burial solution. This extreme situation placed Manaus at the center of worldwide attention and concern. A widely cited seroprevalence surveillance study estimated that by October 2020, some 76% of the Manaus population had become infected by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (i.e., an attack rate of  $AR_1 = 76\%$ ) (2). The large first wave attack rate led to claims that herd immunity had been achieved, and thus, according to fundamentals of the epidemiological theory, a follow-up second epidemic wave was no longer possible, and the city was now “safe” or “protected.” Scientific and public interest in the events of Manaus increased further because it was one of the few cities where the mathematical theory of herd immunity could be monitored in real time.

Against all expectations, in late November, a new highly transmissible Gamma P.1 variant of the SARS-CoV-2 virus appeared in Manaus, and a major second wave of COVID-19 rapidly spread through the city and then radiated outward to all Brazilian states (7). The second wave resulted in a mortality rate even larger than the first (Fig. 1),

## Significance

The nearby cities of Iquitos (Peru) and Manaus (Brazil) experienced the world's highest infection and mortality rates during the first COVID-19 wave in 2020. Key studies suggested that >70% of the city populations were infected in this wave and thus close to herd immunity and protected. It remains an enigma as to why a deadly second wave followed in Manaus worse than the first. To resolve this, we present a data-driven model of epidemic dynamics in Iquitos which we use to help explain and model events in Manaus. The partially observed Markov process model simultaneously fits a flexible “variable  $R_0$ ”, estimates long-term immunity waning and impulsive immune evasion, and thus provides a comprehensive framework for characterizing and modeling new variants of concern.

Author contributions: D.H. and L.S. designed research; D.H., L.L., and L.S. performed research; D.H. and L.S. contributed analytic tools; D.H., L.L., Y.A.-R., H.D., and L.S. analyzed data and critically reviewed the paper; and L.S. wrote the paper.

Competing interest statement: B.J.C. has consulted for AstraZeneca, GlaxoSmithKline, Moderna, Pfizer, Roche, and Sanofi Pasteur. Other authors report no other potential conflicts of interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>D.H. and L.S. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [lewi.stone@rmit.edu.au](mailto:lewi.stone@rmit.edu.au).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2211422120/-/DCSupplemental>.

Published February 27, 2023.

with more than 170 residents dying per day at the peak, and rapidly triggering yet another collapse of the city's healthcare system. Obviously, this was not the expected herd immunity that had been hoped for. Multiple explanations have been hypothesized (8) to account for the appearance of this large second wave, but many questions still remain puzzling. Did the first wave of COVID-19 truly reach an attack rate ( $AR_1$ ) as large as 76% in Manaus in only several months [especially given the issues with blood donor surveillance data (3, 9–12); *Discussion*]? Was herd immunity attained? If so, how could a second wave appear in Manaus soon after that was larger than the first wave in terms of both deaths and cases? Going further, did reinfection with the new P.1 variant play a major role in generating the second wave? Did the similar large-scale COVID-19 epidemic in neighboring Iquitos ( $AR_1 > 70\%$ ) give further support for the massive attack rate estimate in Manaus, or as we argue shortly, did it in fact support the exact opposite? Methodologically, how can attack rates and changing transmissibility be estimated when there are multiple waves?

Here, extending principles from our earlier work (13–17), we make use of a mathematical modeling approach that has the capability of tracking reinfections and immunity loss over multiple follow-up waves, collected over several years, to help answer questions of this type.

## Background Details on Manaus and Iquitos

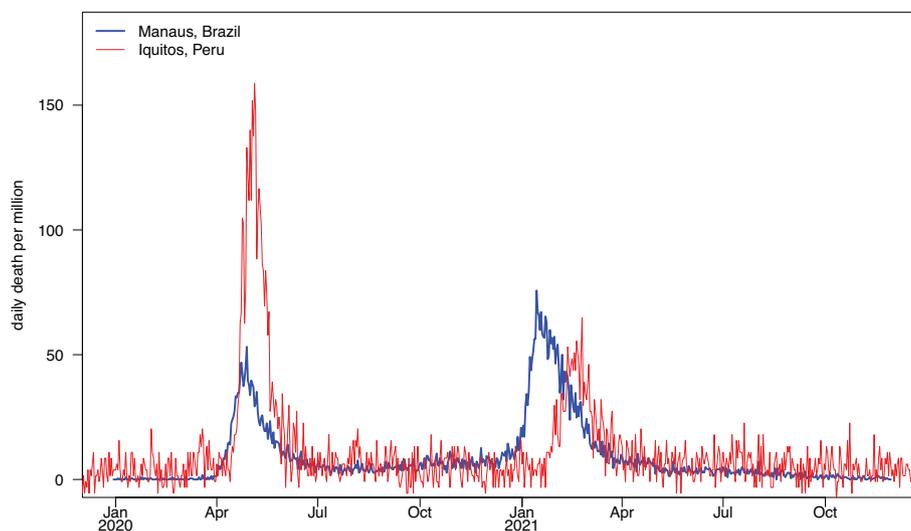
In May 2021, Álvarez-Antonio et al. (1) published their detailed analysis of COVID-19 impact in Iquitos, Peru, a neighboring city of Manaus, and the capital city of Peru's Loreto Region. The small city is extremely isolated and famously considered the largest city in the world unreachable by road. The majority of the population resides in a small strip of  $\sim 30 \text{ km}^2$ , where the population density should be on average  $>15,000$  people per  $\text{km}^2$ , i.e., roughly two times larger than typical in highly heterogeneous and spread out Manaus. In a large, randomized population seroprevalence survey (726 residents sampled from 90,354 households), Álvarez-Antonio et al. (1) recorded 70% seroprevalence at the end of the first wave in April to May 2020. Other studies in Loreto found unusually high seroprevalence levels among health workers in Iquitos hospitals (18)

and also in some rural areas (19). The large-scale city-wide survey presumably provides a more rigorous spatial sampling scheme than the blood donor sampling scheme used for Manaus (20).

A key motivation for the Iquitos seroprevalence study was to determine whether herd immunity was in fact reached or whether a second wave might appear (1). The regional director of health in Loreto, Dr. Carlos Campala, emphasized, “that the studies were carried out due to the possibility of having a second wave.... If the study is confirmed [which it was], it would show that we are in one of the safest cities in the country and could open its doors to tourism” (11). To help explain the large attack rate in Iquitos, the regional director commented, “They cannot compare us, for example, with Spain where two or three people live in a house, while in our country two or three families live, that is, between 10 and 15 people” (10). Such reports suggest that the concept of herd immunity was of considerable importance to health officials in the Amazon region at the time.

Despite the large attack rate in Iquitos and the spike in mortality numbers during the first wave, Iquitos did not manage to evade a smaller second wave at the beginning of 2021 (Fig. 1). Similar to Manaus, the second wave coincided with the appearance of new variants of concern (VoCs). These included the Lambda variant which was prevalent in other parts of Peru, but it was the Gamma variant P.1 that dominated during 2021 in Iquitos [*SI Appendix, section 6*; (21)], as also in Manaus. In what follows, we model and fit the Iquitos dataset and explore how the different qualitative patterns of their first two waves (in terms of case numbers and mortality) help resolve the enigma of Manaus and thus shed light on key events in the history of the pandemic. Apart from our work here, no modeling study has been conducted to analyze the situation in Iquitos or compare the two cities.

Broadening our earlier work (13–15), a scheme is implemented that fits a flexible time-varying reproductive number  $R_0(t)$  to COVID-19 mortality data. We explain shortly how some other related schemes adopted to achieve this goal may be unreliable, or at best have been used for single-epidemic outbreaks only, and are still under development (22). The method used here is intended to reverse engineer the two waves of the Manaus and Iquitos COVID-19 mortality data to learn about the underlying processes generating these waves. In the process, the method tracks



**Fig. 1.** Mortality data. Daily severe acute respiratory illness (SARI) deaths (*per million inhabitants*) in Manaus, Brazil (blue) (5), compared with excess daily deaths (*per million inhabitants*) in Iquitos (red). The Iquitos data were constructed from all-cause deaths in ref. 6 with a background death rate of mean 3.36 per day subtracted as based on pre-pandemic (2017 to 19) daily all-cause deaths (*SI Appendix, Fig. S4*). Note that the total population of Manaus is 2,219,580 while that in Iquitos is 426,000 (1).

reinfections and impulsive immunity loss over multiple follow-up waves. The follow-up waves assist in calibrating calculations of attack rate of the first wave ( $AR_1$ ), population immunity loss, reinfection, and other key indexes that might otherwise be difficult to determine without performing such a comparison between waves and between cities. The same principles are fundamental for understanding the dynamics of multiple waves in general and will assist in the current difficulties faced in predicting how the buildup and waning of immunity at the population level will affect the invasion of future VoCs. Relevance for the case of Omicron is briefly discussed in *SI Appendix, section 7*.

## Results

**Preliminary Data Overview.** Mortality data is considered one of the more reliable indicators of COVID-19 burden (23). In Manaus (total population of 2,219,580), there were 4,971 severe acute respiratory illness (SARI) deaths over the first wave period from March 2020 to October 2020. Over the second wave period, from November 2020 to November 2021, there were 8,949 SARI deaths reported, which is almost double in number. In the smaller city of Iquitos (total population of 426,000) (see ref. 1), there were 1,812 excess deaths in the first wave versus 1,459 in the smaller second wave. As a comparison, in Fig. 1, we have plotted the per capita (per million inhabitants) mortality rates for Manaus (blue) and Iquitos (red), which should give a reasonable representation of the true intensity of COVID-19 deaths (*Materials and Methods*). Data sources for this study are given in the *Materials and Methods* section.

What stands out in Fig. 1 is that the per capita death rate at the peak of the epidemic in Iquitos is some 3 to 4 times larger than that in Manaus. Arguably, this may possibly be due to higher typical population densities in Iquitos, but other complex factors may be involved that include the level of medical services and support from the State, the capability of the completely overstressed hospital system with limited intensive care beds, levels of poverty in the city and the effectiveness of poorly implemented mitigation schemes.

From an overview of the available SARI mortality data (*Materials and Methods*), we begin by roughly approximating the first wave attack rate in Manaus in an attempt to independently verify that  $AR_1 \approx 76\%$ . It will be assumed that at the population level, the rate of mortality (deaths per day) is approximately proportional to the incidence (infections per day) of SARS-CoV-2 in a city. Thus, when comparing epidemic waves of a single city, if the second wave has double the number of deaths, we would approximate the number of infections to also be double. Such an approximation assumes that changes in the age distribution of cases are minor (as confirmed in ref. 24), as well as deaths (which can be adjusted for), and that health system capacity changes little over the two epidemics. But allowance can be made for errors that might arise through some of these assumptions, as discussed shortly.

According to the Manaus mortality data, the second wave of deaths was thus 1.8 times larger than the earlier first wave. Using mortality as an approximate proxy for infective cases, a simple calculation shows that if reinfections were negligible, the first wave could only have a maximum of  $AR_1 \approx 36\%$  [since  $AR_1 + 1.8 AR_1 = 100\%$ ], which is far smaller than the  $AR_1 = 76\%$  estimate in refs. 2 and 25. [Note that Buss et al., also consider reinfection negligible in similar calculations (24).] Even if there were an intermediate level of reinfection, it would have minor impact on the mortality data because reinfections (pre-Omicron) rarely led to severe cases and even more rarely led to deaths (*SI Appendix, section 3*).

But even if 20% of those infected in the first wave became reinfected after recovery, and led to severe cases or deaths, this would result in  $AR_1 \approx 39\%$  [ $AR_1 + 1.8 AR_1 - 0.2 * AR_1 = 100\%$  implies  $AR_1 = 39\%$ ]. A simplifying assumption here is that P.1 and non-P.1 strains are similarly fatal. Should the P.1 strain be 1.5 or 1.8 times more fatal than the ancestral strain, this would still mean the first wave should not be larger than  $AR_1 = 45.5\%$  or  $50\%$ , respectively. Examining other realistic variations to these assumptions changes the outcome minimally, as shown by a simple model (*Materials and Methods*—Eq. 1).

The results are in agreement with the large-scale seroprevalence study of Manaus by Lalwani et al. (26), who found that the first wave had “maximum possible” disease prevalence of 45% by October 2020, and a randomized serological study by Hallal et al. found 22% by June 2020 (27). They are also similar to the nearby Mâncio Lima cohort in the Brazilian Amazon where adjusted seroprevalence was 38.9% (95% HDI 33.2 to 44.8%) at the end of the first wave (24).

While the frequency of COVID-19 reinfection cases is difficult to estimate, they were few in number in the pre-Omicron period 2020 to 2021, generally less severe, and rarely led to death (28–30). Specifically, a recent study on the P.1 (Gamma) variant in Nicaragua found that previous infection provided “some protection against infection during the second wave approximately 1 year after the first wave. Protection was higher against more severe outcomes, with 78.9% protection...against moderate or severe infection...68.1% protection...against symptomatic infection... and 63.9% protection against any detectable infection” (31). The first two waves in Manaus were separated by approximately 8 mo, indicating additional protection than given by these figures. A study on the island of Parintins, Amazonas, reports (32), “We determined that the Gamma variant was responsible for a high proportion of infections in this cluster; we did not find evidence of Gamma variant infection in persons with previous COVID-19.” Further discussion and review on the extent of reinfections may be found in *SI Appendix, section 3*. As such, it is unlikely that reinfections played a major role in generating the large second wave of deaths experienced in Manaus as graphed in Fig. 1. If the second wave of P.1 infections in Manaus was hypothetically an outcome of reinfections, then second wave mortality (due to severe cases) should not be larger than the first wave as seen in Fig. 1, especially if  $AR_1 \approx 76\%$  as claimed by Buss et al. (2). Given these statistics, it would not be unreasonable to suggest the possibility that the attack rate of the first wave in Manaus might be far smaller, with  $AR_1 \ll 76\%$ .

**Multiwave Model.** We now make use of a modeling approach to estimate epidemiological parameters of the processes driving the waves of COVID-19 mortality [see recent works (13, 15)]. Since for both cities, there was only one dominant strain in any given time period, a one-strain model is sufficient to capture the population dynamics over multiple waves (*SI Appendix, section 6*). The invasion and successful domination of the new P.1 variant in the second wave are accommodated by making changes to relevant parameters before and after the time P.1 entered the city. Following refs. 25 and 33, the model only includes a single age-class which helps to keep the number of parameters within reason for fitting purposes, and this is discussed further in the *Materials and Methods* section.

In the model formulation, at any time point, each individual in the population belongs to one of the following compartmental subpopulations: susceptible (S), exposed (E), infected (I), recovered (R), severely ill (T), or dead (D). We denote the total population size as N. The goal of the approach is to use the model to

help reconstruct the time series of  $S(t)$ ,  $E(t)$ ,  $I(t)$ ,  $R(t)$ ,  $T(t)$ , and  $D(t)$  based only on a knowledge of the mortality data  $D(t)$ . This is a technique that goes back to earlier work on epidemic time series reconstruction initiated in refs. 34 and 35 and makes use of the partially observed Markov process modeling [Materials and Methods; (36)].

The model features two loops. First, the usual SEIRS loop where susceptible individuals flow through the compartments in the usual circular  $S \rightarrow E \rightarrow I \rightarrow R \rightarrow S$  fashion (see LHS of the schematic in Fig. 2). That is, a susceptible individual who becomes exposed upon meeting an infected can potentially become ill with COVID-19 and then infectious. In this loop, a proportion  $(1 - \theta)$  of infected people eventually enter the recovered class where they gain temporary immunity to the disease. After some time delay, the immunity wanes, and recovered individuals become susceptible once more, closing the SEIRS loop, and thereby allowing reinfections to occur. As is known, previously infected individuals who have recovered have considerable immunity to reinfection (SI Appendix, section 3). Rather than tracking reinfected individuals separately (e.g., creating two susceptible classes), making the complex model considerably more complicated, we suppose that the remaining differences can be accommodated by the assignment of different infection fatality ratio (IFR) parameter values for the second wave. Note that when reinfections are relatively infrequent, especially those that lead to severe cases, the approximation is reasonable.

The second loop characterizes the proportion ( $\theta$ ) of ill people who become severely ill (T). We suppose that after an additional time delay having mean  $1/\kappa$ , a proportion  $\theta$  of these will die from the disease, while a proportion can still recover and move to the recovered class. In this setting, the infection fatality ratio becomes  $IFR = \theta^2$ . For realism, the time delay is incorporated by breaking up the T compartment into three staggered stages so that the delay  $1/\kappa$  also has a gamma distribution with mean  $\frac{1}{\kappa} = 12$  d. Details of the full-model parameter settings and fitting procedures are presented in the Materials and Methods section.

**Immunity Loss  $\phi$  and Immune Evasion Ability  $\psi$ .** Similar to ref. 37, our model incorporates immunity loss in two different ways. First, it is assumed that infected individuals, upon recovery, gain full immunity which proceeds to wane with the passage of time. To incorporate a realistic waning period (of mean  $1/\phi$  years), the recovered class is also broken down into a series of three subcompartments, which recovered individuals must pass through.

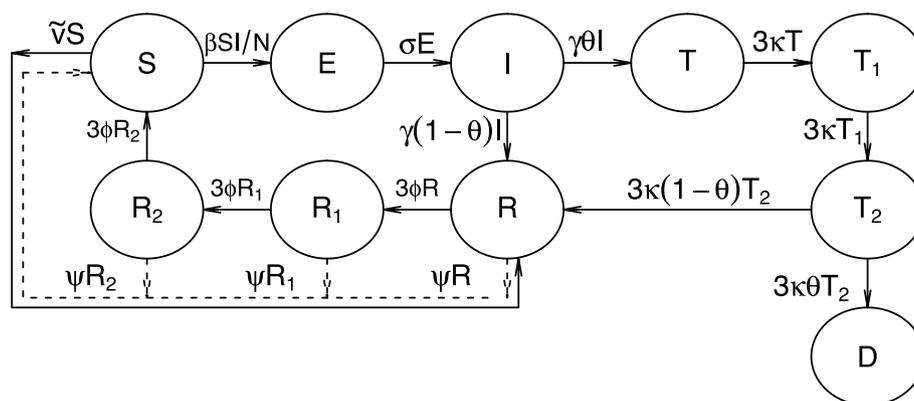
The loss time of the waning immunity ( $1/\phi$ ) is characterized by a gamma distribution with mean set to  $1/\phi = 2$  y for full immunity loss (see e.g., ref. 37).

In the case of SARS-CoV-2, the appearance of new VoCs has sometimes been characterized by their increased immune escape, which is also the case for P.1 (25). This may be achieved by including an impulsive drop in population immunity levels by a proportion  $\psi$  which is inferred from fitting the data. The parameter  $\psi$  is referred to as the immune evasion ability of P.1 (37). In Manaus, the impulsive drop takes place on November 23, 2020, and in Quito on January 1, 2021.

**Parameters.** The fitting procedure is described in the Materials and Methods section. Specifically, the fitted parameters include the transmission rate  $\beta(t)$ , the immune escape proportion  $\psi$ , and the  $IFR_1$  (infection fatality ratio  $= \theta^2$ ) for the ancestral strain and  $IFR_2$  ( $= RR * IFR_1$ ) for the variant of concern P.1. And RR is multiplied at the rate from  $T_2$  to Death. The transmission  $\beta(t)$  is assumed to be time varying with some restrictions (SI Appendix, section 1 and Materials and Methods) while we have set the restrictions  $IFR_1 < 0.6\%$  and  $IFR_2$  in the range [0.5, 1.8]-fold of  $IFR_1$ . The remaining fixed parameters were set to reflect the most current knowledge of COVID-19 dynamics (Materials and Methods).

**Time-Varying Transmission Rate.** Over the course of the pandemic in these cities, the mitigation measures, public risk perception, reactive behaviors, and seasonality continuously modify the population's contact rates and thus ultimately the transmission rate  $\beta(t)$ . Additionally, the different virus variants have different levels of transmissibility. These differences are incorporated through the inclusion of a time-varying transmission rate  $\beta(t)$  modeled using a cubic spline with  $n_\beta$  nodes and obtained by fitting the model to the mortality data (Materials and Methods). The time-varying basic reproductive number is given by  $\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma}$ , where the infectious period is given by  $1/\gamma$ , and provides a direct index of the effects of mitigation measures. To avoid confusion, we do not directly discuss the effective reproductive number  $\mathcal{R}_{eff}(t) = S(t)\mathcal{R}_0(t)$  in this paper.

The method finds the best fitting parameters based on the 2nd-order Akaike information criterion (AICc) (38) to find the optimal number of nodes in the transmission spline  $n_\beta$ . Models with too many parameters are penalized (SI Appendix, section 4), thereby taking into account the trade-off between parsimony (number of free parameters) and goodness of fit. The methodology of



**Fig. 2.** A flowchart of the SEIRS-based model. Susceptible individuals flow through compartments in a circular  $S \rightarrow E \rightarrow I \rightarrow R \rightarrow S$  loop (LHS of diagram). Infected people either recover and gain temporary immunity or become severely ill (T) and those who do not recover die (D) (RHS of diagram). The compartments R and T are staggered and broken into three serial compartments to add realistic time delays. Vaccination  $v(t)$  can also be included.

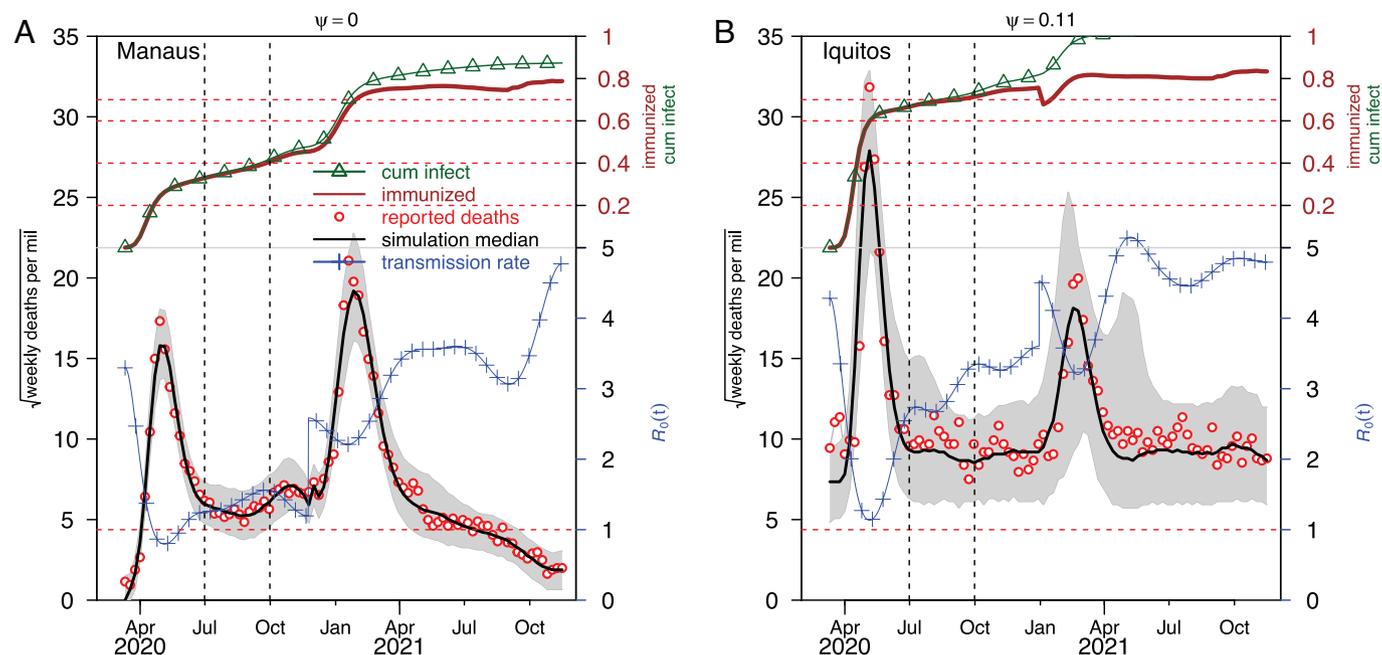
reconstructing  $\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma}$  has been thoroughly tested on mock datasets generated from complex and similar epidemiological models (4). For simple epidemics, one might argue that in principle, by varying  $\mathcal{R}_0(t)$  alone, one can fit any epidemic curve. But this is not the case here, where the complexity of the system, with its multiple waves, and considerations of immunity buildup and waning imply that there are tight restrictions on the manner in which  $\mathcal{R}_0(t)$  varies over the study period. For example, a large number of infectives in the second wave would imply that  $\mathcal{R}_0(t)$  has to be restricted in the first wave. This restriction ensures there are enough susceptibles available to support a second epidemic wave. The range of  $\mathcal{R}_0(t)$  does not need to be set by the modeler but is simply the envelope of all feasible solutions of reasonable fit.

**Model Analysis.** Fig. 3 *A* and *B* show that the best fitting models fit the Manaus and Iquitos mortality waves with good accuracy. The observed weekly deaths are represented by red circles after the data were square root-transformed to assist in observation of the troughs. The median of the simulated data (1,000 model simulations), plotted as a function of time, is given by the solid black line. The latter always falls within the gray regions that show the 95% range of the 1,000 simulations and envelope the median. Note that removal of the square root scale of the mortality numbers would give a wider envelope from a visual perspective. The blue curve shows the estimated transmission rate in units of  $\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma}$  and plotted as a function of time with its own separate scale given on the right-hand axis. In the upper panels, the brown curve indicates the proportion of the population with immunity (including exposed and infectious individuals) at any time. The upper green curve represents the total proportion of population ever infected and may be taken as the cumulative attack rate.

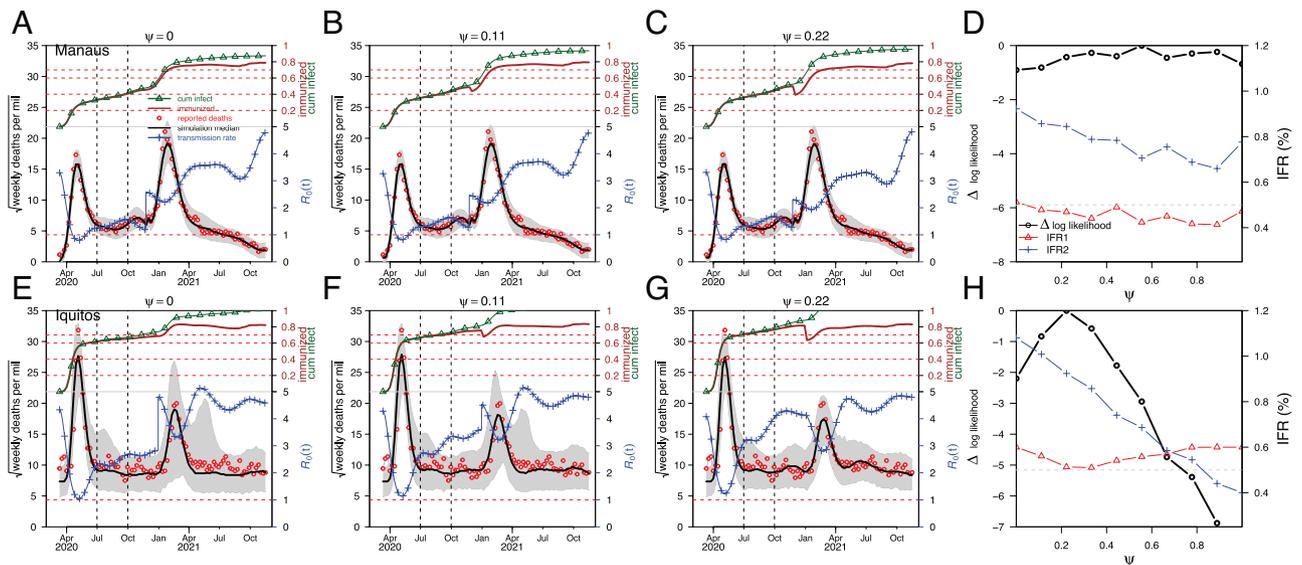
For Iquitos, the best fitting model gives estimates for the infection attack rate (green) over the first wave that sits at  $AR_1 \approx 64\%$  and  $AR_1 \approx 72\%$  by mid-July and October, respectively, for  $\psi = 11\%$  impulsive immune evasion (Fig. 3*B*). The latter is

indicated visually by the pulsed reduction in January 2021 of the proportion of the population with immunity (red line; Fig. 4). Fig. 4 *E–H* show that  $AR_1$  changes little for  $\psi = 0\%$  to  $\psi = 22\%$  immune evasion. These results for  $AR_1$  are very similar to those found by Álvarez-Antonio et al. (1) who reported a first wave  $AR_1 = 70\%$  by mid-July 2020 in Iquitos. Fig. 4 *D* and *H* show the relative log-likelihood profile of the model as a function of the immune evasion parameter  $\psi$ . (Here, this is relative to the maximum likelihood estimates—see *Materials and Methods*.) For Iquitos, the relative log-likelihood profile peaks at  $\psi = 22\%$ , but in fact, all values of  $\psi$  in the range  $10\% < \psi < 40\%$  have similar likelihoods, and their fits are difficult to differentiate by eye, as seen in Fig. 4 *E–G*. It should be recalled that to report a significant difference [with a threshold level of  $P = 0.05$ ], a change of 1.9 log-likelihood units is required.

However, the situation for Manaus, as shown in Fig. 3*A*, is very different. The simulated model (black line; for  $\psi = 0$ ) fits the Manaus mortality data well, but the attack rate of  $AR_1 = 38$  to  $40\%$  (green line) by October 2020 is significantly lower than that of  $AR_1 = 76\%$  estimated by Buss et al. (2) and  $81\%$  estimated by Faria et al. (25). Even with larger levels of immune evasion, the attack rate is at most  $AR_1 = 47\%$  (Fig. 4 *B* and *C*). Yet, in contrast, in the city of Iquitos, the model did actually succeed in estimating  $AR_1 \approx 60$  to  $70\%$  by October 2020 similar to that estimated by the seroprevalence surveys of Álvarez-Antonio et al. (1). This result, combined with our understanding from the Preliminary Data Overview above, strongly suggests that the attack rate of Manaus was significantly less than that of  $AR_1 = 76\%$ . If the attack rate in Manaus really was  $AR_1 = 76\%$ , in principle, the model should have identified this. Also remarkable is the excellent fit to the data based on a transmission spline with the optimal  $n_\beta = 14$  nodes determined via an AIC analysis, although the square root data transformation enhances the fit visually. The estimation results were robust and changed little whether one spline node was estimated every 6.4 wk ( $n_\beta = 14$ ) or one node parameter estimated every 11 wk



**Fig. 3.** Fitting results of the model to weekly mortality data collected in (A) Manaus and (B) Iquitos (B). Red circles show the reported weekly per capita deaths (square root of weekly deaths per million of population). The solid black line shows the median of simulated weekly per capita deaths from 1,000 model simulations. The gray region showed the 95% range of 1,000 simulations. Observation noise has been included in outputs. The blue curve shows the transmission rate in units of  $\mathcal{R}_0(t) = \frac{\beta(t)}{\gamma}$ . In the upper panels, the brown curve indicates the percentage of population with immunity (including exposed and infectious individuals). The upper green curve represents the total proportion of population ever infected (attack rate).



**Fig. 4.** Comprehensive fitting results of the model to Manaus (A–D) and Iquitos (E–H) weekly mortality data. Red circles show reported weekly per capita deaths (square root of weekly deaths per million of population). The black curve shows the median of simulated weekly per capita deaths from 1,000 model simulation trajectories. The gray-enveloped region indicates the 95% range of 1,000 simulations. The blue curve shows the transmission rate in units of  $R_0(t) = \beta(t)/\gamma$ . The brown curve indicates the proportion of population with immunity (including exposed and infectious individuals). The upper green curve represents the total proportion of population ever infected (attack rate). The relative log likelihood of model (black curve with circles in D and H) is plotted as a function of the rate of (impulsive) immune evasion  $\psi$ . First wave IFR<sub>1</sub> (red curve with square) and second wave IFR<sub>2</sub> (blue curve with diamond) as functions of the rate of immune evasion  $\psi$ . The scale for IFR is shown on the right-hand axis as a percentage.

( $n_\beta = 8$ ), as we show in *SI Appendix, section 4* and *Fig. S3*, although simulation variability increased in the latter case.

The model reaches these conclusions by making calculations that are essentially similar to the approximate calculations given in the Preliminary Data Overview above but in a more sophisticated manner. As indicated in the overview, a first wave attack rate larger than  $AR_1 = 40\%$  in Manaus would make it difficult to fit the data with sensible parameters and also to fit the second wave with limited susceptibles available. We demonstrate this specific point in *SI Appendix, section 2* and *Fig. S2*. Finally, the large difference between Manaus ( $AR_1 \approx 40\%$ ) and Iquitos ( $AR_1 \approx 72\%$ ) should be expected given that the fitted initial transmission rates, or reproductive numbers seen in *Fig. 3*, indicate that  $R_0(0) = 3.2$  in Manaus and  $R_0(0) = 4.2$  in Iquitos.

**Comparison with the Bayesian Model of Faria et al. (25).** One of the most sophisticated and better known COVID-19 models is the Bayesian model of the Imperial group which Faria et al. (FEA; 2021) (25, 39) modified. The core of the FEA model has been used for studying COVID-19 spread in the United States (33), Europe, and India (40). However, the method appears to have several weaknesses that may have inflated the first wave attack rate estimate of Manaus and led to inaccurate estimates of parameters, as detailed in *SI Appendix, Part B3* and summarized here. Several problems may relate to the fitting algorithm, and the possibility of potential overfitting, which does not seem to have been checked for.

FEA treat the reproductive number  $R_0(t)$  as a constant modulated by an autoregressive AR(2) model of “weekly effects.” In theory, the AR(2) modulation alone is unsuited to modeling the major impact nonstationary mitigations including lockdowns or other practices that result in strong disruption of disease transmission (face masks, school closures, and mobility loss). We have found that the FEA code estimates  $R_0(t)$  in an unusual manner and very differently to the AR(2) characterization presented in their article (*SI Appendix, Part B*). By running the model on a variety of synthetic and real mortality datasets, we observed that the FEA fitting scheme was always unusually accurate (and possibly

overfitting), despite the simplicity of the model structure and its relatively few parameters, as presented. Examples of this are given in *SI Appendix, Part B*. However, in practice, rather than fitting a statistical AR(2) model, their code fits the (49 different) “weekly updates”  $\epsilon(t)$  in a way that ensures a highly accurate fit to the mortality data, while  $R_0(t)$  is fitted in parallel. In short, this is not standard autoregressive time series modeling as implied by the authors. Instead, there appears to be puzzling overfitting which no AR(2)-based model of the type they describe could achieve and which is not being assessed (e.g., using the AIC). This combined with carefully specified priors (*SI Appendix, Part B3*) can lead to incorrect estimates of  $R_0(t)$  and indexes such as attack rate.

Sensitivity analysis of some key parameters indicates the presence of other underlying issues. The FEA model estimates the first wave attack rate in Manaus as  $AR_1 = 81\%$ , but this proves to be very sensitive to the prior for the reproductive number, which is effectively fixed at  $R_0(t = 0) \approx 3.28$  for the first 2 wk (*SI Appendix, Part B2*). With very minor changes in the profile of  $R_0(t)$  over these first 2 wk, the attack rate drops from  $AR_1 = 81\%$  to  $AR_1 = 56\%$ . Yet, the model still fits the mortality data with high accuracy (*SI Appendix, Part B2*).

The model of FEA reports a first wave attack rate of  $AR_1 = 81\%$  which leaves very few susceptibles in the population at the end of 2020 when P.1 arrived. As confirmed by running the code of FEA, a large second wave of P.1 can only occur if the reproductive number of P.1 is unrealistically high at  $R_0 > 9$ , which is implausible (*SI Appendix, Part B1*). In view of these issues, the model results we report here should be considered as a serious alternative.

## Discussion

It has been suggested that the high attack rate of  $AR_1 = 76\%$  estimated for Manaus by Buss et al. (2) has the potential to be inflated as the analysis was based on convenience sampling of blood donor data collected in hospitals (3, 9–12, 20, 41, 42). Thus, the landmark review by Levin et al. (43) discussed in depth empirical calculations of IFR<sub>1</sub> especially in Manaus, making comparisons with the estimates obtained from blood donor data in ref. 2. However, their updated

systematic review no longer makes mention of Manaus but instead clarifies (20), “Our analysis only included studies that had a random selection of participants from a sample frame representative of the general population. . . . Consequently, studies of convenience samples—such as blood donors or residual sera from commercial laboratories—were excluded. Such samples are subject to intrinsic selection biases. Indeed, there is abundant evidence from the pandemic that convenience samples provide inaccurate estimates of seroprevalence, with assessments indicating that they are likely to overestimate the true proportion infected.”

Buss et al. (2) estimated the infection fatality ratio (IFR) to be on average  $IFR_1 \approx 0.17$  to  $0.28\%$ , which is low, as would be necessary to obtain the high attack rate found,  $AR_1 = 76\%$ . However, we were unable to fit mortality dynamics sensibly when assuming such low values of IFR, and the best fitting model estimated the infection fatality ratio over the first wave to be  $IFR_1 \approx 0.47$  to  $0.5\%$ . This helps explain why the first wave attack rate estimated here is not as large as in Buss et al. (2). For neighboring Iquitos, the model estimate  $IFR_1 \approx 0.55$  to  $0.6\%$  which is almost identical to the most recent estimates of Levin et al. (20) for Iquitos.

Of particular note is the estimate of Manaus especially large infection fatality ratio during the P.1 wave in 2021 compared to the non-P.1 wave in 2020, as measured by the relative risk  $RR = IFR_2/IFR_1$ . The model fitting required  $IFR_1 = 0.5\%$  with an RR between  $RR = 1.4$  and  $1.8$  to accommodate this large second wave in Manaus with  $IFR_2 = 0.7$  to  $0.9\%$ . This underlines the fact that a relatively large infection fatality ratio  $IFR_2$  of the P.1 variant might be the only sensible way to explain the second wave, especially under the constraint of already having a large first wave  $AR_1$ . Current views suggest that the large second wave in Manaus may be due to the pressure on healthcare and hospital services at the time rather than any major biological difference between fatality rates of P.1 and the ancestral virus (24, 44).

For Iquitos, in contrast, there is a clear peak in the likelihood profile in Fig. 4H, and the problems of identifiability are less pronounced. As discussed in detail in *SI Appendix, section 8*, with a large first wave in Iquitos, it becomes difficult to fit the second wave especially when it relies more heavily on reinfections (which might not be available), as a source of new susceptibles. Thus, parameters are more constrained, and some parameter regimes become more feasible than others. According to Fig. 4H, estimates of bounds for immune evasion  $\psi$  over the second wave for Iquitos can be approximated as  $0.1 < \psi < 0.4$ . This would correspond to the fraction of the infected people in the first wave who were reinfected in the second wave. However, the panels in Fig. 4E–G suggest that the larger values of  $\psi$  are associated with unrealistically large estimates of  $\mathcal{R}_0(t)$  for the ancestral strain in the first wave. It is unlikely that the reproductive number exceeded  $\mathcal{R}_0(t) = 3$  after the crash of the first wave, indicating that it is unlikely that  $\psi > 0.11$ . Moreover, given what is known about P.1 reinfection,  $\psi = 40\%$  does seem extreme, although from reports in Nicaragua,  $\psi = 10$  to  $20\%$  after 1 y from the first infection, might be possible [(31); *SI Appendix, section 3*].

The related modeling work of Yang and Shaman (37) also encounters difficulties in estimating immune evasion as seen by the large CIs reported (37). These issues indicate the difficulties in estimating immune evasion and thus reinfection rates. The problem is further complicated by the fact that reinfections are relatively rare, making estimation from data difficult.

## Conclusion

In Iquitos, the model estimated that the first wave infected 72% of the population by July 2020, while by the time of the peak of the second P.1 wave in March 2021, almost 100% of the population was

infected (Fig. 4E–H). From this date onward, there was almost complete reliance on immune evasion and reinfection to generate new infections. The results indicate that if 70% of the Iquitos population were infected in the first wave, the remaining 30% would have been infected in the second wave. Local reports indicate that among those hospitalized in the second wave, more than 95% were new or without previous infection. The model estimated that in August 2021, at the end of the second wave, 100% of the population was infected, and 7.7% was reinfected (based on setting  $\psi = 0.11$ ).

Although Buss et al. and Faria et al.’s (2, 25) serological and modeling analysis concluded that the first wave in Manaus left some 76% of the population infected (the modeling in fact suggested 81%), this was contradicted by two independent serological studies of local teams on general population data (26, 27, 45) and was considered controversial (3, 4, 46). The best fitting model reported here for the Manaus data indicated a first wave attack rate that was large but still modest with  $AR_1 \approx 40\%$  matching our earlier first wave analysis (4). The relatively large number of susceptibles remaining in the population after the first wave makes it possible for the model to both fit and simulate a second wave without significant reliance on reinfections. The fit would not be achievable if it were really true that  $AR_1 \approx 76\%$  in Manaus, in which case a second wave would be implausible.

Clearly, the modeling showed that for both cities, herd immunity had not been reached, and even Iquitos which may have been close to herd immunity was able to generate a reasonably sized second wave. In the presence of immune evasion and reinfection, we see that large first wave attack rates as in Iquitos do not necessarily prevent smaller secondary waves, indicating limitations of the herd immunity concept under these conditions.

The goal of the present study was to design a framework for estimating attack rates and immune evasion characteristics based on available epidemic data. There are few other modeling tools available for this work, yet it is particularly important given the developing immune evasion abilities of new SARS-CoV-2 variants, as we discuss in *SI Appendix, section 7*. We demonstrated the problems that may arise in achieving these goals and how modeling via reconstruction techniques can be of service (34, 35). These techniques have a long history in the modeling of infectious diseases. Some of the more difficult problems were alleviated when information based on data from multiple epidemic waves became available, and the power of reconstruction techniques could be better exploited. An attempt was made to resolve an important debate concerning the extreme epidemics in Manaus and Iquitos. Although it can be argued that the large attack rate in Iquitos ( $AR_1 > 70\%$ ) suggests that the nearby city of Manaus might also reasonably have large  $AR_1 > 70\%$ , our analysis shows the opposite and provides a useful alternative view. The patterns of the two waves in Manaus were qualitatively and quantitatively different from those in Iquitos, and the modeling suggests that Manaus had sufficient susceptibles to drive a large second wave. This was not true for Iquitos.

## Materials and Methods

**Data.** Mortality data for Manaus were obtained from Brazil’s Ministry of Health (5) which documented severe acute respiratory illness (SARI) daily deaths taken from hospitalized cases (including COVID confirmed). The same SARI mortality data have been used in a number of recent key studies where they were considered a proxy of true COVID mortality (2, 47). We also compared these datasets with COVID-confirmed mortality (CCM) datasets (48) collected and compiled by the local government and in collaboration with the FVS (Fundação de Vigilância em Saúde do Amazonas) and Brazil’s Ministry of Health. The latter datasets were continually updated with retrospective corrections. As we show elsewhere (4), the CCM dataset of daily deaths appears to suffer from underreporting, but only to a limited extent, and it appears that the CCM dataset is a proper subset of the SARI deaths.

We obtained all-cause mortality data in Iquitos from 2017 to 2021 as given in ref. 6 and plotted in *SI Appendix, Fig. S4*. There is minor seasonality (*SI Appendix, Fig. S5*). When fitting the all-cause deaths in Iquitos, we allowed for a daily baseline of deaths due to other causes with Poisson (rate = 3.36), a rate that was calculated based on pre-pandemic daily all-cause death. The resulting excess death mortality time series is very similar to the excess mortality dataset presented by Álvarez-Antonio et al. (1). By excess deaths is meant “all-cause deaths–baseline deaths”.

**Model.** The model is based on a susceptible–exposed–infectious–recovered–dead (SEIRD) formulation with a flexible time-varying transmission rate  $\beta(t)$  to the reported mortality data. The model equations are as follows:

$$\begin{aligned} \dot{S} &= 3\phi R_2 - \beta SI / N - \eta \bar{\nu} S, \\ \dot{E} &= \beta SI / N - \sigma E, \\ \dot{I} &= \sigma E - \gamma I, \\ \dot{T}_1 &= \theta \gamma I - 3\kappa T_1, \\ \dot{T}_2 &= 3\kappa T_1 - 3\kappa T_2, \\ \dot{D} &= 3\kappa \theta T_2, \\ \dot{R} &= \eta \bar{\nu} S + (1 - \theta) \gamma I + 3\kappa(1 - \theta) T_2 - 3\phi R, \\ \dot{R}_1 &= 3\phi R - 3\phi R_1, \quad \dot{R}_2 = 3\phi R_1 - 3\phi R_2. \end{aligned}$$

On day  $t^*$ , an impulsive immune evasion occurs, whereby a proportion  $\psi$  of those with immunity ( $R$ ,  $R_1$ , and  $R_2$ ) are shunted to the susceptible class as follows:

$$\begin{aligned} S(t^*_{+}) &= S(t^*) + \psi (R(t^*) + R_1(t^*) + R_2(t^*)), \quad R(t^*_{+}) = (1 - \psi) R(t^*), \quad R_1(t^*_{+}) \\ &= (1 - \psi) R_1(t^*), \quad R_2(t^*_{+}) = (1 - \psi) R_2(t^*). \end{aligned}$$

**Parameters and Fitting.** The method uses a plug-and-play likelihood-based inference framework [(49); *SI Appendix, section 1*]. We use iterated filtering [implemented in the R package POMP (partially-observed Markov processes) (50, 51)] which was built based on the sequential Monte Carlo (particle filtering) for partially observed Markov processes. It is difficult and often impossible to estimate all the parameters that typically arise in complex epidemiological models due to statistical identifiability issues. Following common practice, we fixed some of the parameters to known and well-accepted estimates when they were available [(e.g., Yang and Shaman (37)), although careful consideration is needed [Elder et al. (52)]. The parameters were set to reflect the most current knowledge of COVID-19 dynamics, with details given in the main text. In particular, exposed individuals move to the infectious class after a mean latent time of  $\frac{1}{\sigma} = 2$  d. Infected individuals move to the recovered or severe class after a mean infectious period of  $\frac{1}{\gamma} = 3$  d. The generation time (GT) equals the sum of mean latent period and mean infectious period, which here equals 5 d, in line with previous studies (53, 54). According to ref. 55, the delay between the first symptom onset and the death is  $\sim 15$  d, which justifies the choice of  $\frac{1}{\kappa} = 12$  d from loss of infectiousness to death. That is, the onset of infectiousness was taken to be 2 d ahead of the first symptom onset. The infection fatality ratio (IFR) is given by  $IFR = \theta^2$  (Fig. 2) and also fitted. This assumes that the proportion of severe cases among all infections equals the proportion of mortality among severe cases. The rationale is that the two proportions are not jointly identifiable in this situation. Since we do not fit data of the severe cases (an intermediate class between infection and mortality), the assumption will not affect other results.

We follow closely our previous work for fitting a “flexible” time-varying transmission rate  $\beta(t)$  (13) based on fitting the mortality data. The former requires defining  $\beta(t) = \exp(\text{cubic\_spline})$  as an exponential cubic spline with  $n_\beta$  nodes

largely distributed evenly over the 89-wk study period (we divided  $n_\beta$  nodes into two halves; one half distributed evenly in the pre-P.1 time interval and the other half evenly in the P.1 time interval). To be specific,  $n_\beta = 14$  nodes, with seven nodes evenly over the pre-P.1 period and seven nodes evenly over the P.1 period (*SI Appendix, section 1*: Reproductive Number). This corresponds to fitting one parameter (spline node) every 6.4 wk to estimate  $R_0(t) = \frac{\beta(t)}{\gamma}$ . This setting was the conclusion of an AIC analysis to determine the best fitting model after penalizing for the number of parameters (*SI Appendix, section 4*). We also found that the results were remarkably robust and change little whether there is one parameter estimated per 6.4 wk ( $n_\beta = 14$ ) or one parameter estimated every 11 wk ( $n_\beta = 8$ ) (cf. *SI Appendix, section 4 and Fig. S3*).

The time step size for the model was set as 1 d. We integrated  $\dot{D}$  (with baseline deaths in Iquitos) for a week to obtain the simulated weekly deaths  $D_t$ . The reported deaths were defined as  $RD_t$ , where

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

Here,  $\tau$  denotes the overdispersion and accounts for the observation noise due to surveillance and heterogeneity among individuals. A weekly log-conditional likelihood can be defined with  $RD_t$  and  $D_t$ . The overall log likelihood is the sum of all weekly log-conditional likelihoods. The relative log likelihood was defined and plotted as  $\Delta \log \text{likelihood} = \log \text{likelihood}(\text{parameter}) - \log \text{likelihood}(\text{parameters at Maximum log-likelihood MLL})$ .

To add realism, it was also assumed that  $\sim 1\%$  proportion of the population has preexisting immunity, which only changes the initially large available susceptible pool by a small amount.

We tested a closely related model and provided details in online notes (4) that demonstrate the ability of this method to fit synthetic mortality data generated from a model using arbitrary functions for  $\mathcal{R}_0(t)$ . In all cases, the methodology succeeds in reconstructing  $\mathcal{R}_0(t)$  from simulated mortality data.

**Single Age-Class Assumption.** Following refs. 25 and 33, the model only includes a single age-class which helps to keep the number of parameters within reason for fitting purposes. In an earlier study of COVID-19 in Manaus, a model with two age-classes was implemented, and similar results were achieved in terms of estimating the attack rates (4). We recognize that a single age-class model will be unable to capture possible shifts in the age structure of infections and deaths between waves that might arise because later variants often spread more readily among children (56), as was hypothesized in the Amazonian region (24). However, “[s]uccessive SARS-CoV-2 IgG serosurveys in the Brazilian Amazon showed that age-specific attack rates and proportions of symptomatic SARS-CoV-2 infections were similar before and after Gamma variant emergence” (24). “Of note, symptomatic infections did not affect young children disproportionately more during the second wave” (24). Freitas et al. (44) found an age shift over the two waves in terms of hospitalized severe cases and death rates. Even if there were some difference in age-class dynamics, such as found by Freitas et al. (44), they could as a first approximation be accommodated by suitable adjustment in the process of fitting IFR1 and IFR2 of the two waves.

**Simplifying Assumptions in Preliminary Data Overview.** The calculations in the Preliminary Data Overview make the simplifying assumption that at the end of the second wave almost all Manaus residents were infected. If this were not the case, the  $AR_1$  should be even less than the calculation indicates. A more complete calculation is as follows.

We suppose the infection fatality ratio in the first wave is  $IFR_1$  and in the second wave is  $IFR_2$  and for reinfected individuals is  $IFR_3$ . Let the relative rate be  $RR = IFR_2 / IFR_1$ . Suppose  $Z\%$  of the population was infected at least once at the end of the two waves. Let  $r$  be the proportion of individuals in the first wave who were reinfected. Let  $\alpha$  be the ratio of the total deaths in the second wave compared to the first wave. For Manaus,  $\alpha = 1.8$ . Then, a calculation reveals the following:

$$AR_1 = \frac{RR \cdot Z}{[RR + \alpha - r^*] \%} \approx \frac{RR \cdot Z}{[RR + \alpha] \%} \quad [1]$$

where  $r^* = r \cdot IFR_3 / IFR_1$ . Here, the RHS approximation holds when the proportion of reinfections  $r$  is small and their contributions to the death rates are small. For

$Z = 100\%$  and  $RR = 1, 1.5, \text{ or } 1.8$ , one obtains,  $AR_1 = 36\%, 45.5\%, \text{ and } 50\%$ , respectively. If 10% of the population remain uninfected after the two waves ( $Z = 90\%$ ), one obtains  $AR_1 = 33\%, 41\%, \text{ and } 45\%$ .

**Data, Materials, and Software Availability.** Data and code are publicly available online at <https://github.com/linlixina/model-Manaus-IQUITOS> or [https://www.zotero.org/groups/4832328/linlixin\\_lin](https://www.zotero.org/groups/4832328/linlixin_lin). Data may be found in the file `sari_city2.csv` and code operation described in `README.md`. Previously published data were used for this work (6, 48).

**ACKNOWLEDGMENTS.** We thank Prof. Ben Bolker, Dr. Celis Salinas, Dr. Devon Graham, Dr. Contia Cornelius, Prof. Antonio Alvarez, M. Quispe, Prof. Aaron King, and Prof. Edward Ionides. D.H. and B.J.C. were supported by the Collaborative

Research Fund (grant number HKU C7123-20G) of the Research Grants Council (RGC) of Hong Kong, China. D.H. was also supported by a project of RiFood (grant number 1-CD52).

Author affiliations: <sup>a</sup>Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China; <sup>b</sup>Research Institute for Future Food, The Hong Kong Polytechnic University, Hong Kong, China; <sup>c</sup>Mathematical Sciences, School of Science, Royal Melbourne Institute of Technology (RMIT) University, Melbourne, Victoria 3000, Australia; <sup>d</sup>Department of Theoretical and Computational Ecology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1090 GE, Amsterdam, Netherlands; <sup>e</sup>World Health Organization (WHO) Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; and <sup>f</sup>Biomathematics Unit, School of Zoology, Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

1. C. Álvarez-António *et al.*, Seroprevalence of anti-SARS-CoV-2 antibodies in Iquitos, Peru in July and August, 2020: A population-based study. *Lancet Glob. Health* **9**, e925–e931 (2021).
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. L. Ferrante *et al.*, Brazil's COVID-19 epicenter in Manaus: How much of the population has already been exposed and are vulnerable to SARS-CoV-2? *J. Racial Ethn. Health Disparities* **9**, 2098–2104 (2021).
4. D. He *et al.*, The unexpected dynamics of COVID-19 in Manaus, Brazil: Was herd immunity achieved? medRxiv [Preprint] (2021). <https://doi.org/10.1101/2021.02.18.21251809>.
5. Anonymous, Brazil Ministério da Saúde Opendatas SRAG 2020 in Opendatas (Brazil) (2020).
6. Anonymous, "Plataforma Nacional de Datos Abiertos" 1 February 2022. <https://www.datosabiertos.gob.pe/group/ministerio-de-salud>.
7. F. G. Naveca *et al.*, COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence. *Nat. Med.* **27**, 1230–1238 (2021).
8. E. C. Sabino *et al.*, Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).
9. T. Rocha Filho *et al.*, Attack rate and the price of SARS-CoV-2 herd immunity in Brazil. *Res. Sq.*, 10.21203/rs.3.rs-659187/v2 (2021).
10. Anonymous, Covid-19: Seroprevalencia en la ciudad de Iquitos es de 75 por ciento. *Andina*, 6 April 2022. <https://andina.pe/agencia/noticia-covid19-seroprevalencia-la-ciudad-iquitos-es-75-ciento-815197.aspx>.
11. Anonymous, Director Regional de Salud Loreto: No podemos decir que Iquitos tiene inmunidad de rebaño. *CanalN*, 6 April 2022. <https://canaln.pe/actualidad/director-regional-salud-loreto-no-podemos-decir-que-iquitos-tiene-inmunidad-rebano-n420928>.
12. Anonymous, A Brazilian city thought it had herd immunity. It was wrong. *The Economist*, 4 April 2022. <https://www.economist.com/the-americas/2021/01/23/a-brazilian-city-thought-it-had-herd-immunity-it-was-wrong>.
13. L. Stone, D. He, S. Lehnstaedt, Y. Artzy-Randrup, Extraordinary curtailment of massive typhus epidemic in the Warsaw Ghetto. *Sci. Adv.* **6**, eabc0927 (2020).
14. L. Stone, R. Olinsky, A. Huppert, Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533–536 (2007).
15. D. He *et al.*, Evaluation of effectiveness of global COVID-19 vaccination Campaign. *Emerg. Infect. Dis.* **28**, 1873–1876 (2022).
16. A. Feng, U. Obolski, L. Stone, D. He, Modelling COVID-19 vaccine breakthrough infections in highly vaccinated Israel: the effects of waning immunity and third vaccination dose. *PLOS Glob. Public Health* **2**, e0001211 (2022).
17. J. B. Axelsen, R. Yaari, B. T. Grenfell, L. Stone, Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9538–9542 (2014).
18. R. A. Chafloque-Vásquez, L. Pampa-Espinoza, J. C. Celis Salinas, Seroprevalencia de COVID-19 en trabajadores de un hospital de la Amazonia peruana. *Acta Médica Peruana* **37**, 390–392 (2020).
19. A. Moreira-Soto *et al.*, High SARS-CoV-2 seroprevalence in rural Peru, 2021: A cross-sectional population-based study. *mSphere* **6**, e0068521 (2021).
20. A. T. Levin *et al.*, Assessing the burden of COVID-19 in developing countries: Systematic review, meta-analysis and public policy implications. *BMJ Glob. Health* **7**, e008477 (2022).
21. N. Vargas-Herrera *et al.*, SARS-CoV-2 Lambda and Gamma variants competition in Peru, a country with high seroprevalence. *Lancet Reg. Health Am.* **6**, 100112 (2022).
22. O. Kounchev, G. Simeonov, Z. Kuncheva, Scenarios for the spread of COVID-19 analyzed by the TVBG-SEIR spline model. *Biomath* **10**, 2103087 (2021).
23. N. Brazeau *et al.*, "Report 34: COVID-19 infection fatality ratio: estimates from seroprevalence" (Report 34, Imperial College London, London, UK, Jan 11, 2022). <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-34-ifr/>.
24. V. C. Nicolette *et al.*, Epidemiology of COVID-19 after emergence of SARS-CoV-2 gamma variant, Brazilian Amazon, 2020–2021. *Emerg. Infect. Dis.* **28**, 709 (2022).
25. N. R. Faria *et al.*, Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021). [10.1126/science.abb2644](https://doi.org/10.1126/science.abb2644).
26. P. Lalwani *et al.*, SARS-CoV-2 seroprevalence and associated factors in Manaus, Brazil: Baseline results from the DETECTCoV-19 cohort study. *Int. J. Infect. Dis.* **110**, 141–150 (2021).
27. P. C. Hallal *et al.*, SARS-CoV-2 antibody prevalence in Brazil: Results from two successive nationwide serological household surveys. *Lancet Glob. Health* **8**, e1390–e1398 (2020).
28. L. J. Abu-Raddad, H. Chemaitelly, R. Bertolini, Severity of SARS-CoV-2 reinfections as compared with primary infections. *N. Engl. J. Med.* **385**, 2487–2489 (2021).
29. C. F. Schuler IV *et al.*, Mild SARS-CoV-2 illness is not associated with reinfections and provides persistent spike, nucleocapsid, and virus-neutralizing antibodies. *Microbiol. Spectr.* **9**, e0008721 (2021).
30. M. E. Flacco *et al.*, Risk of SARS-CoV-2 reinfection 18 months after primary infection: Population-level observational study. *medRxiv [Preprint]* (2022). <https://doi.org/10.1101/2022.02.19.22271221>.
31. H. E. Maier *et al.*, Protection associated with previous SARS-CoV-2 infection in Nicaragua. *N. Engl. J. Med.* **387**, 568–570 (2022).
32. J. F. da Silva *et al.*, Cluster of SARS-CoV-2 gamma variant infections, Parintins, Brazil, March 2021. *Emerg. Infect. Dis.* **28**, 262 (2022).
33. H. J. T. Unwin *et al.*, State-level tracking of COVID-19 in the United States. *Nat. Commun.* **11**, 1–9 (2020).
34. G. V. Bobashev, S. P. Ellner, D. W. Nychka, B. T. Grenfell, Reconstructing susceptible and recruitment dynamics from measles epidemic data. *Math. Popul. Stud.* **8**, 1–29 (2000).
35. B. F. Finkenstädt, B. T. Grenfell, Time series modelling of childhood diseases: A dynamical systems approach. *J. R. Stat. Soc. Series C Appl. Statist.* **49**, 187–205 (2000).
36. E. L. Ionides, C. Bretó, A. A. King, Inference for nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18438–18443 (2006).
37. W. Yang, J. Shaman, Development of a model-inference system for estimating epidemiological characteristics of SARS-CoV-2 variants of concern. *Nat. Commun.* **12**, 1–9 (2021).
38. K. P. Burnham, D. R. Anderson, Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
39. S. Bhatt *et al.*, Semi-mechanistic Bayesian modeling of COVID-19 with renewal processes. arXiv [Preprint] (2020). <https://doi.org/10.48550/arXiv.2012.00394>.
40. M. S. Dhar *et al.*, Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science* **374**, 995–999 (2021).
41. A. Backhaus, Common pitfalls in the interpretation of COVID-19 data and statistics. *Intereconomics* **55**, 162–166 (2020).
42. I. Amigo, "Study estimates 76 percent of Brazilian city exposed to SARS-CoV-2" The Scientist. <https://www.the-scientist.com/news-opinion/study-estimates-76-percent-of-brazilian-city-exposed-to-sars-cov-2-68272>. 14 December 2021.
43. A. T. Levin *et al.*, Assessing the age specificity of infection fatality rates for COVID-19: Systematic review, meta-analysis, and public policy implications. *Eur. J. Epidemiol.* **35**, 1123–1138 (2020).
44. A. R. R. Freitas *et al.*, The emergence of novel SARS-CoV-2 variant P.1 in Amazonas (Brazil) was temporally associated with a change in the age and sex profile of COVID-19 mortality: A population based ecological study. *Lancet Reg. Health Am.* **1**, 100021 (2021).
45. P. Lalwani *et al.*, High anti-SARS-CoV-2 antibody seroconversion rates before the second wave in Manaus, Brazil, and the protective effect of social behaviour measures: Results from the prospective DETECTCoV-19 cohort. *Lancet Glob. Health* **9**, e1508–e1516 (2021).
46. S. Kadelka *et al.*, Science 2021: An alternative, empirically-supported adjustment for sero-reversion yields a 10 percentage point lower estimate of the cumulative incidence of SARS-CoV-2 in Manaus by October 2020. arXiv [Preprint] (2021). <https://doi.org/10.48550/arXiv.2103.13951>.
47. W. M. de Souza *et al.*, Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nat. Hum. Behav.* **4**, 856–865 (2020).
48. Anonymous, "Amazonas and Manaus daily COVID-19 confirmed death data" 1 February 2022. [http://www.amazonas.am.gov.br/content/uploads/2021/01/20\\_01\\_21\\_BOLETIM\\_DIARIO\\_DE\\_CASOS\\_COVID-19-12.pdf](http://www.amazonas.am.gov.br/content/uploads/2021/01/20_01_21_BOLETIM_DIARIO_DE_CASOS_COVID-19-12.pdf).
49. 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. R. Soc. Interface* **7**, 271–283 (2010).
50. A. A. King, "Working with ordinary differential equations in pomp" 1 February 2022. <https://kingaa.github.io/clim-dis/pares/odes.html>.
51. X. Du, A. A. King, R. J. Woods, M. Pascual, Evolution-informed forecasting of seasonal influenza A (H3N2). *Sci. Transl. Med.* **9**, eaan5325 (2017).
52. B. D. Elder, V. M. Dukic, G. Dwyer, Uncertainty in predictions of disease spread and public health responses to bioterrorism and emerging diseases. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15693–15697 (2006).
53. J. Griffin *et al.*, Rapid review of available evidence on the serial interval and generation time of COVID-19. *BMJ Open* **10**, e040263 (2020).
54. C. A. Prete Jr., *et al.*, Serial interval distribution of SARS-CoV-2 infection in Brazil. *J. Travel Med.* **28**, taaa115 (2021).
55. I. Hawryluk *et al.*, Inference of COVID-19 epidemiological distributions from Brazilian hospital data. *J. R. Soc. Interface* **17**, 20200596 (2020).
56. N. G. Davies *et al.*, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020).