In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness

Ruian Ke[a,b] (iD), Carolin Zitzmann[a], David D. Ho[c] (iD), Ruy M. Ribeiro[a] (iD), and Alan S. Perelson[a,b,1] (iD)

[a]Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545; [b]New Mexico Consortium, Los Alamos, NM 87544; and [c]Aaron Diamond AIDS Research Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032

The within-host viral kinetics of SARS-CoV-2 infection and how they relate to a person's infectiousness are not well understood. This limits our ability to quantify the impact of interventions on viral transmission. Here, we develop viral dynamic models of SARS-CoV-2 infection and fit them to data to estimate key within-host parameters such as the infected cell half-life and the within-host reproductive number. We then develop a model linking viral load (VL) to infectiousness and show a person's infectiousness increases sublinearly with VL and that the logarithm of the VL in the upper respiratory tract is a better surrogate of infectiousness than the VL itself. Using data on VL and the predicted infectiousness, we further incorporated data on antigen and RT-PCR tests and compared their usefulness in detecting infection and preventing transmission. We found that RT-PCR tests perform better than antigen tests assuming equal testing frequency; however, more frequent antigen testing may perform equally well with RT-PCR tests at a lower cost but with many more false-negative tests. Overall, our models provide a quantitative framework for inferring the impact of therapeutics and vaccines that lower VL on the infectiousness of individuals and for evaluating rapid testing strategies.

SARS-CoV-2 | viral kinetics | SARS-CoV-2 infectiousness

SARS-CoV-2 is a new human pathogen that causes COVID-19 (1). It is highly contagious, spread rapidly across the globe and has caused 5 million deaths worldwide as of the end of October 2021. At the molecular level, SARS-CoV-2 enters host cells via the angiotensin converting enzyme 2 (ACE-2) receptor. It infects cells in the upper respiratory tract (URT), can rapidly reach a high viral load (VL) and be effectively transmitted (2–4). However, it is not clear how VL, symptom onset, and infectiousness are quantitatively related.

Previously, both VL and $\log_{10}$ VL have been used as surrogates for infectiousness of influenza (5) and SARS-CoV-2 (6, 7). A quantitative understanding of the relationship is critical for both nonpharmaceutical and pharmaceutical interventions. First, it would allow for more precise prediction of the infectiousness of infected individuals, including children and pre- or asymptomatic individuals, based on their VL measurements (8, 9). This could in turn lead to quantification of their contribution to the overall transmission in a community and help to better inform public health policy decisions. Second, as administration of vaccines may lead to lowered VLs in breakthrough infections (10–12), a quantitative understanding will inform how these reductions in VL impact infectiousness and thus allow better predictions of how much transmission vaccinated individuals with breakthrough infection cause. Third, it would provide better insight into a person's infectiousness throughout the course of infection and thus inform testing strategies for work/school reopening, travel, etc. The effectiveness of test, trace, and quarantine as control strategies heavily depends on the sensitivity and specificity of the tests and rate of testing being implemented (13). It was recently proposed that antigen tests with low sensitivity are preferred over highly sensitive RT-PCR tests because of their potential for wide coverage and short turnaround time (6). However, the effectiveness of

this strategy has not been evaluated based on VL and infectiousness dynamics inferred from data.

Here, we construct viral dynamic models of SARS-CoV-2 URT infection and a model linking VL to infectiousness. Mathematical modeling has been applied, by us and others, to understand SARS-CoV-2 infection and the potential impact of therapy (14–18). However, there were large uncertainties in model parameter estimates because in almost all studies, viral dynamic models were fit to data that were taken after symptom onset without knowledge of the patients' infection dates and early VL dynamics. We resolve this issue by using two unique datasets and by using clinical and epidemiological data to inform the quantitative relationship between VL and infectiousness. Using this relationship, we further evaluate the effectiveness of testing strategies using either antigen or RT-PCR tests at different testing frequencies.

Results

Datasets. We use two unique sets of URT VL data for model inference. The first, the "German dataset," contains VL measurements from nine individuals in the first cluster of infections in Germany (3). All individuals had mild symptoms. VLs were measured longitudinally starting several days after symptom onset. We excluded one individual (Patient 16 in ref. 19) because their first VL measurement was long after infection. A unique feature of this dataset is that the detailed transmission history, including the infection dates and dates of symptom

onset, were reported (19). However, this dataset does not have good sampling during the initial VL expansion before the viral peak. Thus, we include a second data set, the NBA (National Basketball Association) dataset, which was taken from a study where individuals (staff and players) were regularly tested during an NBA tournament in 2020 (20). We selected nine individuals sampled frequently, including during the virus expansion phase. In *Dynamics of Early Infection*, we show that these unique features of the two datasets allow us to jointly infer the within-host SARS-CoV-2 dynamics in these individuals including the time of infection.

Dynamics of Early Infection. The SARS-CoV-2 dynamics in the URT are typical of an acute respiratory infection [i.e., VLs increase to a viral peak and decline afterward (Fig. 1)]. Thus, we constructed a target cell limited (TCL) model and an innate immune response model using frameworks previously developed for influenza (21, 22) and SARS-CoV-2 infection (15, 18, 23) (*Methods* and *SI Appendix*). In the innate immune response model, we assumed that innate immune mediators, such as interferons, put target cells into an antiviral state (24) that is refractory to viral infection (22, 25). We first fit these two models to the NBA dataset to estimate the time of infection. Because multiple measurements were taken before peak VL in the nine individuals we chose to study, the times of infection can be estimated relatively reliably. Both the TCL and the

innate response model gave similar estimates of the infection time (*SI Appendix*, Table S1).

We then fit the TCL model and the innate response model to the data from both datasets simultaneously using a nonlinear mixed effect modeling approach (*Methods*). We also tested variants of these models that assume immune mediators block infection of target cells or reduce virus production from infected cells (*SI Appendix*). According to the Akaike information criterion (AIC) scores, the best model overall is the model assuming the innate immune mediators convert target cells into refractory cells (*SI Appendix*, Table S2). This model fits both datasets well (Fig. 1), and it describes both the upslope and downslope of the viral dynamics in the NBA dataset. This gives confidence in our model predictions of the early viral dynamics for individuals in the German dataset. We then tested if there is any difference in estimated parameter values between the two datasets by including the source of the dataset (i.e., the NBA or the German dataset) as a covariate in the model fitting. We found that there was no statistical support for including the origin of the datasets as a covariate (*SI Appendix*, Table S2). Therefore, we use the innate immune response model Eq. **5** (*Methods*) without the covariate for further analysis and term this model the innate response model for short.

According to the best-fit parameter values, the infected cell death rate δ is 1.7 d^{-1} on average (Table 1). Because the model includes an eclipse phase of length $1/k$, where $k = 4$ d^{-1}, the
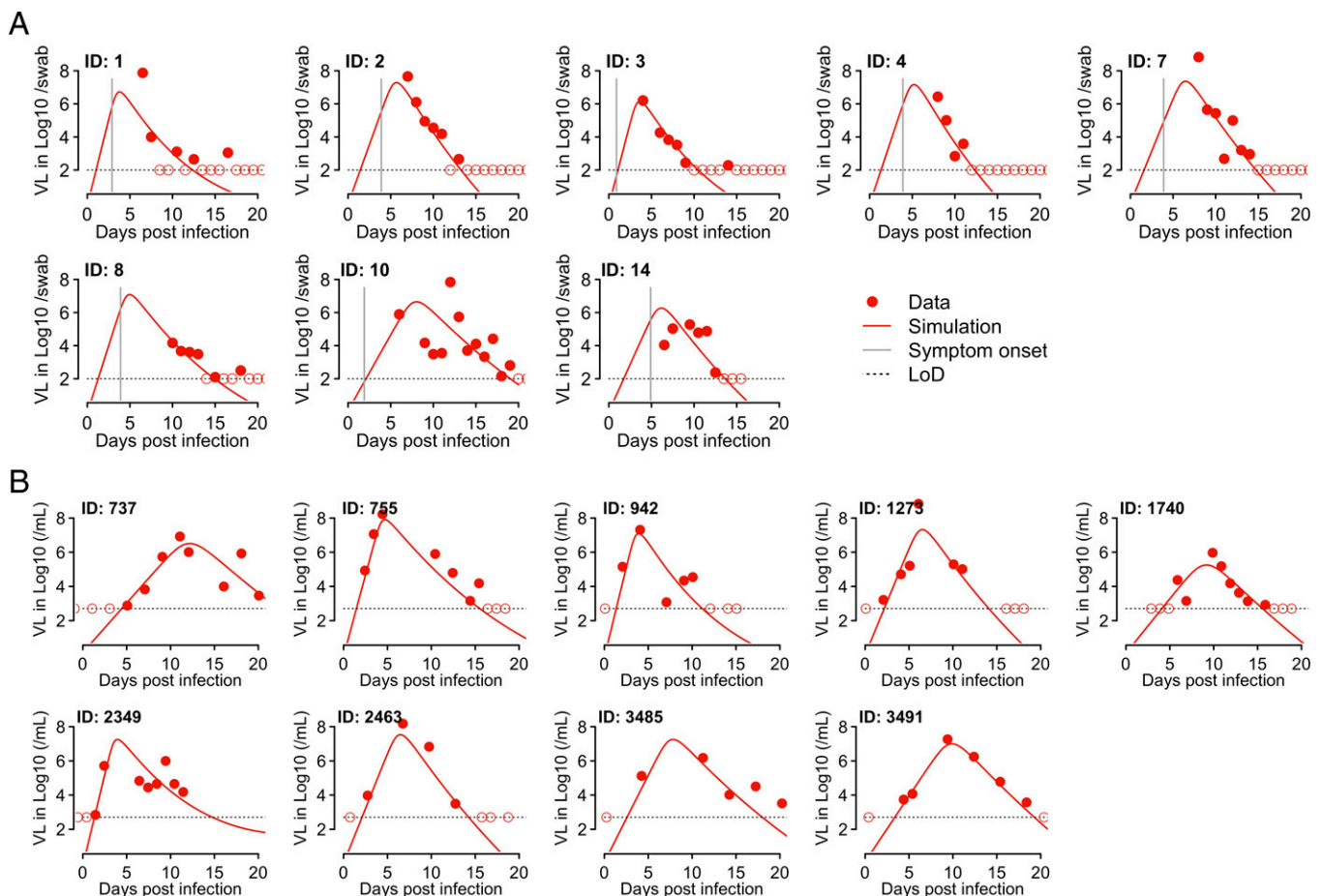
Fig. 1. Fitting results of the innate response model to the VL data from two studies. (*A*) Fitting results to data from eight individuals in the Germany study [i.e., Wolfel et al. (3)]. The model (solid lines) was simulated using the best-fit individual parameter values estimated by a nonlinear mixed effect modeling approach (Tables 1 and 2). The symbols (red dots and circles) show the data from pharyngeal swabs. The circles indicate data points below the limit of detection (LoD). Vertical gray lines denote the time of symptom onset as reported in ref. 19. Horizontal dashed black lines show the LoD. (*B*) Fitting results to data from nine individuals in the NBA study as reported in Kissler et al. (20) with symbols and colors as in *A*.

Table 1. Estimated population parameter values from the best model (i.e., the innate response model)

Parameter	Description	Mean (population estimate)	SD
β	Infection rate constant	3.2×10^{-8} ml/RNA copy/d	0.50
δ	Death rate of infected cells	1.7 /d	0.23
π	Composite parameter for virus production and sampling	45.3 /mL/d	0.24
Φ	Rate constant for the interferon-induced conversion of target cells to refractory cells	1.3×10^{-6} /cell /d	1.95
ρ	Rate at which refractory cells become target cells again	0.0044 /d	0.20

The means and SDs are derived assuming that individual parameters follow log-normal distributions.

average lifespan of infected cells is $\frac{1}{k} + \frac{1}{\delta} = 0.84$ d. We estimated that the within-host basic reproductive number, $R_{0,within}$, varies over a range between 2.6 and 14.9, with mean 7.4 (SD: ±3.8) (*SI Appendix*, Table S3).

We further tested how robust our estimates are with respect to variations in the fixed parameter values in the model by varying each of those in the ranges shown in Table 2 and then refitting the model to the data. Across the scenarios examined, the estimates of the death rate of infected cells were very consistent between 1.6 and 1.9 d^{-1} and the mean $R_{0,within}$ ranged between 5.8 and 8.9 (*SI Appendix*, Table S4). Thus, the estimated parameters and viral dynamic characteristics were robust against variations in the fixed parameters (*SI Appendix*, Table S4).

Probability of Transmission. We next examined how VL is related to the infectiousness of a person by constructing a probabilistic model to describe the various steps in viral transmission from viral shedding to establishment of infection (see Fig. 2*A* for a schematic). We define infectiousness as the probability that an infected person (i.e., a donor) will shed one or more infectious viral particles, leading to successful infection of a recipient for a typical contact of relatively short duration, τ. The typical contact here is defined as in the epidemiological survey study by Mossong et al. (26). Note that the probability defined here only characterizes the infectiousness of a person arising from virus dynamics in the URT given a contact, and it does not assume any frequency of typical contacts. The expected number of transmissions that a person causes can be calculated if the contact pattern of the person is known.

During a contact, the donor sheds both infectious and noninfectious viruses, and a transmission event occurs when one or more infectious viruses reach the recipient and establishes an infection (Fig. 2*A*). We first consider the relationship between the number of infectious viruses, V_{inf}, and the measured VL, V, in a patient sample (e.g., a swab) using three sets of cell culture positivity data [i.e., Jaafar et al. (27), Jones et al. (28), and Kohmer et al. (29)]. In these three datasets, a total of 3,790, 631, and 75 RT-PCR positive nasopharyngeal samples, respectively, with known cycle threshold (Ct) counts or VLs were tested for the presence of infectious virus using cell culture assay.

We examined the following three models describing the relationship between V_{inf} and V: 1) the linear model: V_{inf} is proportional to V; 2) the power-law model: $V_{inf} = \omega V^h$, where ω and h are constants; and 3) the saturation model: $V_{inf} = V_m \frac{V^h}{V^h + K_m^h}$, where V_m and K_m are constants. The probability of a cell culture testing positive can be expressed as $p_{positive} = 1 - \exp\left(-V_{inf}\, \varrho\right)$,

where ϱ is the probability an infectious virus will establish infection in the cell culture (*SI Appendix*). Note that because ϱ always appears as a product with ω or V_m in the expression of $p_{positive}$, ϱ and the number of infectious particles, V_{inf}, cannot be independently estimated from the data we used here. However, the estimated values of h or K_m describes how V_{inf} changes with V.
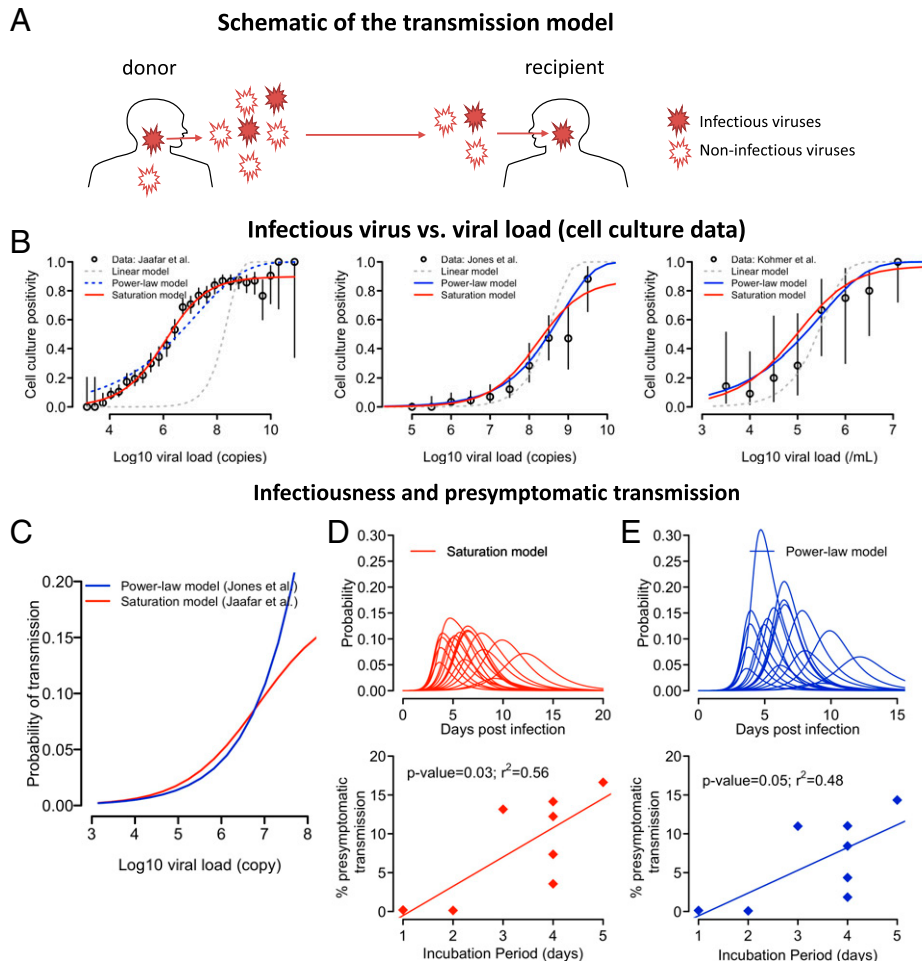
Fitting the three versions of this model to the datasets (*SI Appendix*), we found that the linear model describes all datasets poorly (Fig. 2*B*). The saturation model is the best model to describe the data from Jaafar et al. (Fig. 2*B* and *SI Appendix*, Table S5), and the best fit parameter values are $h = 0.51$ and $K_m = 8.9 \times 10^6$ RNA copies/mL (*SI Appendix*, Table S6). Both the power-law model and the saturation model describe well the data from Jones et al. and Kohmer et al. (Fig. 2*B*), which have a smaller number of samples and thus, potentially, less power to discriminate among the models. The parameter h is estimated to be 0.53 and 0.45, respectively (*SI Appendix*, Table S6), consistent with the exponent h estimated from fitting the saturation model to the Jaafar et al. data. This strongly suggests that the level of infectious viruses increases sublinearly with increases in VL (with the exponent h likely being between 0.4 and 0.6). Because the saturation model describes all datasets well, we will mainly use this model for the analyses that follow. However, we caution that the evidence is not strong enough to rule out the power-law model because the saturating behavior observed in Jaafar et al. may arise from other factors that are not part of the transmission process, such as assay limitations. In addition, another study estimating transmissibility from VL and contact tracing data did not find a saturation effect on VL (30).

We next consider viral shedding from a donor and the establishment of infection in a recipient. We used the saturation function and assumed that the mean number of infectious virions shed is proportional to the number in a sample, V_{inf}, and that the exact number is Poisson distributed. Because it is unlikely that a droplet contains more than one SARS-CoV-2 virion (31), we assume that each infectious virion acts independently, and the number of infectious virions that successfully infect the recipient follows a binomial distribution with its probability parameter defined by the probability an infectious virion establishes an infection, and the number of "trials" parameter is the Poisson-distributed number of infectious virus. Then, the probability of one or more virions generating a successful transmission event for a typical contact at time t is given by (*Methods*):

$$p(t) = 1 - e^{-\theta \frac{V(t)^h}{V(t)^h + K_m^h}}, \qquad [1]$$

Table 2. The fixed parameters in the viral dynamic models and their values

Parameter	Description	Values	Values tested in sensitivity analyses
T_0	Total number of (infection free) target cells	8×10^7 cells	NA
E_0	Initial number of infected cells	1 cell	5, 10 cells
c	Virus clearance rate	10/d	5 and 20/d
k	1/the eclipse phase duration	4/d	3 and 6/d

Ke et al.
In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness

PNAS | 3 of 9
https://doi.org/10.1073/pnas.2111477118

A Schematic of the transmission model

B Infectious virus vs. viral load (cell culture data)

Infectiousness and presymptomatic transmission

Fig. 2. The relationship between VL and host infectiousness. (*A*) A schematic of the probabilistic model describing the steps in a transmission event. A donor sheds both infectious and noninfectious viruses, of which some infectious viruses may reach a recipient during a close contact and establish an infection. (*B*) Best-fit of the three models [i.e., the linear model (gray), the power-law model (blue), and the saturation model (red)] to the data from Jaafar et al. (27), Jones et al. (28), and Kohmer et al. (29). The open circles denote the percentage of cell culture positivity reported, and vertical lines denote the 95% CIs calculated assuming a binomial distribution for the number of positive cultures. For the datasets from Jones et al. (28) and Kohmer et al. (29), VLs are binned into half-$\log_{10}$ intervals. Solid lines are used for models that describe the data well. (*C*) The predicted probability of transmission for a typical contact as a function of $\log_{10}$ VL given by the saturation model in Eq. **1** with $\theta = 0.20, h = 0.51,$ and $K_m = 8.9 \times 10^6$ RNA copies (red) or by the power model in Eq. **2** with $\phi = 2.4 \times 10^{-5}$ and $h = 0.53$ (blue). (*D* and *E*) The infectiousness profile for all individuals studied (lines in *Upper*) predicted by the infectious model assuming a saturation function (Eq. **1**) or a power-law function (Eq. **2**), respectively. (*Lower*) The relationship between the duration of the incubation period (*x* axis) and estimated presymptomatic area under the infectiousness curve. Irrespective of the model used, expected presymptomatic transmission is more likely in individuals with a longer incubation period.

where θ is a composite parameter incorporating the fraction of infectious viruses reaching the recipient and the probability of each establishing an infection (*SI Appendix*). Note that when θ is small, $p(t)$ can be approximated by the Hill function $\theta \frac{V(t)^h}{V(t)^h + K_m^h}$. A Hill function was used previously to model the transmission probability for HIV (32), influenza (5, 33), and more recently SARS-CoV-2 (16).

The values of h and K_m are determined using cell culture data. θ is a constant such that the maximum transmission probability (i.e., the maximum infectiousness) is $1 - e^{-\theta}$, which is approximately θ for $\theta \ll 1$. Multiple epidemiological studies indicate that the secondary attack rate per typical contact is low (i.e., less than 20%) (34–36). We thus set $\theta = 0.20$ in the analysis that follows so that the maximum transmission probability is ~20% for a typical contact.

Setting $\theta = 0.20$, $h = 0.51$, and $K_m = 8.9 \times 10^6$ RNA copies/mL, we calculated how infectiousness depends on VL (Fig. 2*C*) and how infectiousness varies over time, $p(t)$ (i.e., the

infectiousness profile) for each individual (Fig. 2*D* and *SI Appendix*, Fig. S1). If we define the infectious period as the period when the infectiousness, $p(t)$, is above 0.02 (i.e., 10% of the maximum probability), the infectious period ranges between 1.9 and 7.9 d with a mean of 5.5 d across the 17 individuals (*SI Appendix*, Fig. S1). For the individuals in the German dataset where the date of symptom onset is known, we calculated the presymptomatic fraction of infectiousness by dividing the area under the infectiousness curve $p(t)$ before symptom onset by the total area under the infectiousness curve. This fraction represents the expected fraction of presymptomatic transmissions (if a person is not rapidly isolated after symptom onset). We found that the fraction ranges between 0 and 17% (Fig. 2*D* and *SI Appendix*, Fig. S1). Interestingly, there is a statistically significant association between the duration of the incubation period (i.e., the time between infection and symptom onset) and the predicted probability of presymptomatic transmission (Fig. 2*D*; $P = 0.03$). This suggests that the longer the incubation period, the more likely presymptomatic

transmission occurs, and presymptomatic transmission is mostly driven by individuals who have an incubation period greater than 5 d.

To further cross validate this choice of parameters in the infectiousness model, we compared our model predictions with epidemiological data not used to derive our model. First, from the infectiousness profiles predicted by our model, we calculated using Eq. **6** (*Methods*) the expected serial interval for each individual (assuming random contacts) and found the mean serial interval across all 17 individuals studied to be 7.1 d. This is consistent with a mean serial interval of 6.5 to 8 d in the absence of active tracing and isolation efforts as estimated in ref. 37. Second, from the infectiousness profile, we calculated using Eq. **7** (*Methods*) the number of potential transmissions for each individual assuming that there are on average 13.4 typical contacts per day according to the estimates from several European countries reported in Mossong et al. (26). We then estimated the expected reproductive number of SARS-CoV-2 at the epidemiological level, $R_{0,epi}$, by taking the mean of the numbers of potential transmissions. We estimated that $R_{0,epi}$ is 5.2 for the 17 individuals (*Methods*), within the range of $R_{0,epi}$ values estimated previously for European countries (38). Therefore, these independent validations support our infectiousness model in Eq. **1**.

Similarly, we derive the probability of transmission using the power-law function as

$$p(t) = 1 - e^{-\phi V(t)^h}, \qquad [2]$$

where ϕ is a constant. We estimated $h = 0.53$ and $\phi = 2.4 \times 10^{-5}$ from the data by Jones et al. (28) such that this version of the model predicts a mean serial interval and $R_{0,epi}$ (6.9 and 5.1 d, respectively) that are consistent with epidemiological studies (37, 38). In general, the model predictions of infectiousness are similar to the predictions using the saturation model when the VL is lower than 10^7 copies/mL; however, the predictions of the two models diverge when the VL is higher (Fig. 2*C*). The power-law model estimates similar levels of infectiousness to the estimates of the saturation model, except for one individual with a high infectiousness (Fig. 2 *D* and *E* and *SI Appendix,* Figs. S2 and S3). It estimates a similar fraction of presymptomatic infections as the saturation model (Fig. 2*E*). Again, the model predicts that the fraction of expected presymptomatic transmission increases with the length of the incubation period.

Lastly, we tested whether the linear model is consistent with epidemiological data by assuming that V_{inf} is a constant fraction of V (*SI Appendix,* Fig. S4). The model predicts that the fraction of presymptomatic infections is extremely small (i.e., less than 8% in each of the patients in the German dataset [*SI Appendix,* Fig. S4*B*]) inconsistent with epidemiological data (2, 4, 39). Therefore, datasets from cell culture experiments as well as epidemiological studies suggest that the fraction of virus particles that are infectious is not constant over the course of infection.

Log VL Is a Better Surrogate Measure of Infectiousness than VL. There are two commonly used surrogate measures of infectiousness (5): the VL or the logarithm of VL. The total infectiousness of a person is then approximated by the area under the VL curve (AUC) or the area under the $\log_{10}$ of the VL curve (AUClog), respectively.

To identify the appropriate surrogate measure for SARS-CoV-2 infection, we first compared the predictions of these two measures with the epidemiological evidence that a large fraction (>30%) of transmissions occur during the presymptomatic stage of SARS-CoV-2 infection (2, 4, 39). Because the dates of infection and symptom onset are only available in the German dataset (3), we focused our analysis on this dataset. When AUC is used as a surrogate for infectiousness, this is very

similar to using the linear model for infectiousness. Therefore, AUC predict very small fractions of presymptomatic transmission (i.e., less than 8% in each of the patients in the German dataset), inconsistent with epidemiological data (2, 4, 39). This suggests the VL and its AUC are not good surrogates for infectiousness.
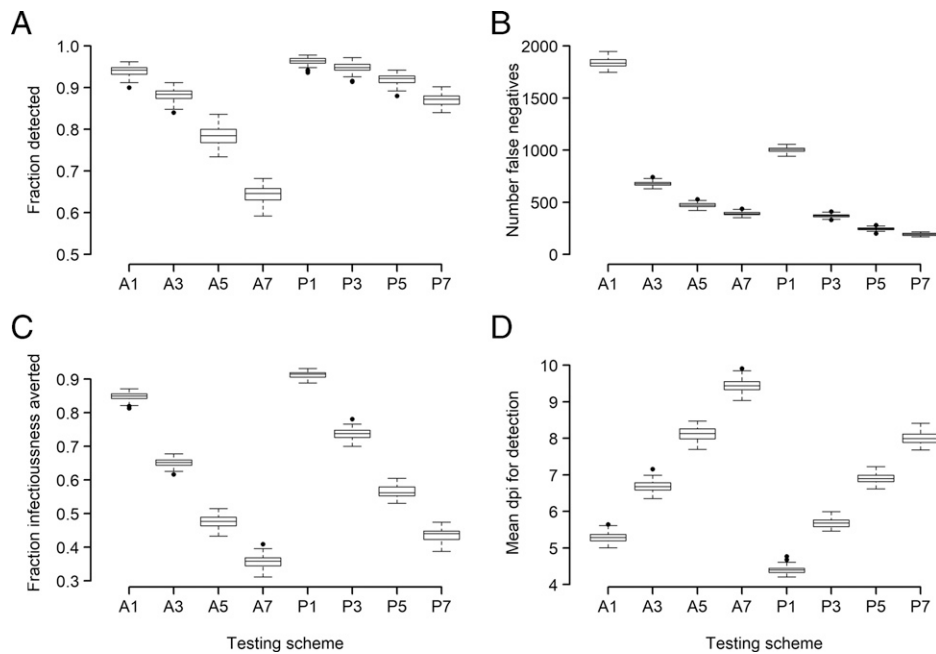
In contrast, when AUClog is used as a surrogate, we predict a sizable fraction of presymptomatic transmissions, between 2 and 27%, which is near the lower bound estimate in ref. 2. We then correlated AUClog with the cumulative infectiousness curve calculated from the probability model based on the saturation function (i.e., Eq. **1**) and found that there exists a strong correlation between the two (*SI Appendix,* Fig. S5*A*, $P = 0.002$). In addition, the fractions of presymptomatic infections predicted by AUClog are very close to those predicted using the area under the curve of infectiousness from the probability infectiousness model (*SI Appendix,* Fig. S5*B*). Therefore, the logarithm of VL, and its corresponding AUClog, serve as a better surrogate for infectiousness than the VL and its corresponding AUC.

Implications for Testing Strategies. Using our best-fit model of how VL (Fig. 1) and infectiousness (Fig. 2*D*) vary with time since infection, we analyzed the impact of possible testing strategies used to reduce the potential for SARS-CoV-2 transmission. We considered two different types of tests: 1) RT-PCR, generally considered the gold standard because of its very high sensitivity and specificity, although its performance depends on the VL and on the quality of the sample collected (40); and 2) antigen tests, which although less sensitive, generally have faster turnaround time (minutes instead of hours to days) and can be self-administered (see *Methods* and *SI Appendix,* Fig. S6 for details).

We studied a hypothetical medium-sized college setting [as described in Paltiel et al. (41)]. In this scenario, during a 12-wk semester in a cohort of 5,000 students/staff, we assume that there were 500 people infected at random times. We implemented four testing frequencies (every person every day, or every 3, 5, or 7 d) using RT-PCR or antigen testing. We assumed the sensitivity for each test varied with time since infection as in *SI Appendix,* Fig. S6 (based on data from refs. 29 and 40), and that the turnaround time was 1 d for RT-PCR and minutes for the antigen test. Given that whether infection is detected or not, as well as the time of detection, is probabilistic, for each scenario we ran 100 simulations using the best-fit model parameter values for each of the 17 individuals. In Fig. 3, we summarized the fraction of the 500 infections detected, the number of false negatives (some people may be false negatives multiple times), the average time of infection until detection, as well as the fraction of total infectiousness averted by detecting someone (assuming that person is then isolated). The fraction of total infectiousness averted was defined as the area under the infectiousness curve from time of detection until resolution of infection in detected individuals divided by their total infectiousness (AUC) averaged over the 500 people infected.

We found that with a RT-PCT test, a large fraction (>80%) of infected individuals can be detected even with a testing frequency of every 7 d (Fig. 3*A*); whereas with an antigen test, testing at least once every 3 d is needed to achieve >80% of detection. Frequent tests (every 3 d for RT-PCR tests and every day for antigen tests) are needed to identify and isolate infected individuals early and thus avert a large fraction of infectiousness (Fig. 3 *C* and *D*).

Overall, the results of these simulations show that although RT-PCR tests perform better than antigen tests in detecting infected individuals and preventing transmission, more-frequent antigen testing (e.g., every day or every 3 d) is comparable to less-frequent RT-PCR tests, at the expense of many more false-negative tests (Fig. 3*B*). This indicates that frequent

Ke et al.
In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness

PNAS | 5 of 9
https://doi.org/10.1073/pnas.2111477118

www.manaraa.com

Fig. 3. Comparison of eight testing protocols using RT-PCR or antigen tests. For each of these, we considered that every person is tested every 1, 3, 5, or 7 d, as indicated in the *x* axis by the number after the test type (e.g., A3 and P3 correspond to antigen and RT-PCR testing, respectively, every 3 d). We plot the number of people detected (*A*), the number of false-negative tests (*B*; note that some people may be false negatives multiple times), the fraction of total infectiousness averted (*C*), and the average time postinfection to detection (*D*).

antigen tests, potentially self-administered at home, could be an important tool in combating spread of infection.

Discussion

In this study, we constructed mathematical models to describe the VL kinetics of SARS-CoV-2 in the URT and their relationship with the infectiousness of an individual. Fitting a viral dynamic model that included an innate immune response to data from refs. 3 and 20, we estimated several key parameter values. The death rate of productively infected cells was estimated to be around 1.7 d^{-1}. Thus, once infected cells start producing virus, they live on average 0.6 d. We estimated the mean within-host reproductive number, $R_{0,within}$, in the URT to be 7.4 with variation among individuals examined, ranging between 2.6 and 14.9. For individuals with known dates of infection and symptom onset, we found that longer incubation periods had higher potential for presymptomatic transmission. A similar finding was reported in a recent study estimating the fraction of presymptomatic transmissions by the duration of the incubation period from transmission pair data (42).

To model viral transmission, we estimated the relationship between the number of infectious viruses in a sample and the sample VL by fitting models to three datasets on infectious virus cell culture positivity (27–29). This led to several interesting findings. First, a consistent finding across the three datasets was that the number of infectious viruses does not increase linearly with increases in VL, suggesting VL itself or the AUC is not a good surrogate for infectiousness. Instead, we found that the number of infectious viruses increases sublinearly with increases in VL. This makes log VL or the AUClog good surrogates for infectiousness. Further experiments are needed to understand this sublinear relationship. Second, a saturation effect on the infectious viruses when VL is very high (e.g., $>10^9$ copies/mL) is needed to explain data from Jaafar et al. (27); however, saturation is not needed to explain the data from Jones et al. (28) and Kohmer et al. (29). The saturation effect, if present, could be due to assay inaccuracies at very high VLs

or could arise from processes in vitro or in vivo that inactivate the virus in high-VL samples. This inconsistency in results vis-à-vis saturation leads to uncertainties in predicting infectiousness when VL is very high. Further experiments measuring the infectious virus concentration especially from samples with high VLs is needed to address this issue. In our study, irrespective of the model used, we found that the risk of transmission for a typical contact of relative short duration becomes high when the VL exceeds between 10^6 and 10^7 RNA copies/mL. This is consistent with the results from Wolfel et al. (3), where infectious viruses were recovered only when VL exceeded 2×10^5 RNA copies/swab and the results from ref. 43, where infectious virus was mainly isolated from specimens with $\geq 10^6$ virus N gene copies/mL. The results are also consistent with the findings in van Kampen et al. (44) where in hospitalized patients with COVID-19, VLs $> 10^7$ copies/mL were associated with isolation of infectious virus.

Using the predicted infectiousness over time for each individual, we evaluated the effectiveness of two testing platforms: RT-PCR and antigen tests. RT-PCR tests are highly sensitive; however, they are costly and may take days to obtain the result. On the other hand, antigen tests are less sensitive but are easy to administer and provide results in less than an hour. Our modeling suggests that RT-PCR tests are better than antigen tests at both detecting infected individuals and effectively reducing total infectiousness when testing is used as a tool for safe reopening of schools and workplaces. However, when frequent RT-PCR testing, say every 7 d, is not feasible due to its high cost and complexity in properly administering these tests, more-frequent antigen tests (i.e., every 1 to 3 d) could be used instead; however, this will lead to higher number of false-negative results due to the large number of antigen tests performed.

Administration of vaccines or effective therapeutics may lead to reduced VLs in the URT (10, 12). Our modeling approach is well suited to quantify the impact of vaccination on the infectiousness of a person. It is beyond the scope of this study to

www.manaraa.com

formally estimate infectiousness of vaccinated individuals who had breakthrough infections. However, as an illustrative example, we use VL data from participant 737 (Fig. 1*B*) to demonstrate how our model can be used to make such predictions. We considered two scenarios of how vaccination impacts viral dynamics. In the first scenario (*SI Appendix*, Fig. S7), we assumed for simplicity, that in breakthrough infections, full vaccination reduces VLs uniformly across time by 10-, 100-, or 1,000-fold (as seen in nasal swabs of some individuals in ref. 10). Our model then predicts the infectiousness of this participant would decrease by 62, 87, or 96%, respectively. In the second scenario (*SI Appendix*, Fig. S7), we assumed that in breakthrough infections, full vaccination reduces peak VLs by 10-, 100-, or 1,000-fold (as seen in nasal swabs of other individuals in ref. 10). Our model then predicts the infectiousness of this participant would decrease by 33, 68, or 87%, respectively. These results demonstrate that the relationship between VL reduction and infectiousness reduction is highly nonlinear. Further modeling work that takes into consideration the possibility that the relationship between VL and infectivity is different in vaccinated and unvaccinated individuals is needed. For example, virus isolated from vaccinated individuals may have vaccine-induced antibodies bound to it, reducing its infectivity. This is consistent with a recent report showing that messenger RNA–vaccinated individuals have reduced infectious VLs that correlate with respiratory antiviral IgG levels (45).

There are limitations to our models. First, the data we used for model inference were from infected individuals with relatively mild or no symptoms (3, 20), who rapidly cleared the virus. The parameter values and relationships we estimated between VL and infectiousness thus may be biased toward mildly symptomatic and asymptomatic individuals. Further work is needed to extend our analysis to individuals with different levels of symptom severity (46) as well as to vaccinated individuals. However, we note that people with severe symptoms will likely often be hospitalized and/or quarantined and contribute less to the spread of the virus. Second, the relationship between VL and the number of infectious particles is inferred from data aggregated from many individuals, and thus it assumes homogeneity across individuals. Further work measuring individual level heterogeneity in the relationship between infectious viral shedding and VL (such as refs. 10 and 18) will help to characterize heterogeneity in individual infectiousness and help make more-precise predictions of the impact of testing strategies on transmission.

Overall, our model linking within-host VL dynamics to infectiousness provides a crucial tool for evaluating both nonpharmaceutical and pharmaceutical interventions and aiding public health policy decisions (47).

Methods

TCL Model. We first study a within-host model based on target cell limitation. The model, which has been used for other viruses (21, 48), keeps track of the total numbers of target cells (*T*), cells in the eclipse phase of infection (*E*) (i.e., infected cells not yet producing virus), productively infected cells (*I*), and viruses measured in swab samples (*V*). The ordinary differential equations (ODEs) describing the model are

$$\frac{dT}{dt} = -\beta VT$$
$$\frac{dE}{dt} = \beta VT - kE$$
$$\frac{dI}{dt} = kE - \delta I \qquad [3]$$
$$\frac{dV}{dt} = \pi I - cV.$$

In this model, target cells are infected by virus with rate constant β. Cells leave the eclipse phase and become productively infected at per capita rate *k*. Productively infected cells die at per capita rate δ. We use *V* to describe viruses measured in pharyngeal swabs, which we assume are a constant proportion

of the total virus in the URT. Therefore, the rate, π, is the product of the viral production rate per infected cell and the proportion of virus that is sampled in a swab. Viruses are cleared at per capita rate *c*. See *SI Appendix* for further details.

From this model, we calculate the within-host reproductive number for SARS-CoV-2, $R_{0,within}$, as

$$R_{0,within} = \frac{\beta\pi}{c\delta}T_0, \qquad [4]$$

where T_0 is the initial number of target cells.

Innate Response Model. We extend the TCL model by including a prototypical innate response (e.g., type-I interferon) following the framework presented in previous models for influenza infection dynamics (21, 22, 25). Immune mediators are produced from infected cells and bind to receptors on target cells stimulating an antiviral response that makes cells refractory to viral infection (*R*). Such cells are said to be refractory cells or cells in an antiviral state (24, 49). In addition to the compartments in the TCL model, the innate response model keeps track of cells refractory to infection (*R*). For simplicity and due to a lack of data, we do not explicitly consider the specific immune mediators (e.g., cytokines) or their concentration. Instead, we make the quasi–steady-state assumption that the dynamics of these mediators are fast and thus their concentration is proportional to the number of infected cells (see *SI Appendix* for details).

The ODEs for the innate response model are

$$\frac{dT}{dt} = -\beta VT - \Phi IT + \rho R$$
$$\frac{dR}{dt} = \Phi IT - \rho R$$
$$\frac{dE}{dt} = \beta VT - kE \qquad [5]$$
$$\frac{dI}{dt} = kE - \delta I$$
$$\frac{dV}{dt} = \pi I - cV,$$

where Φ is a constant describing the rate that innate signaling makes target cells refractory, and ρ is the rate that refractory cells transition back into target cells.

Data, Estimating Time of Infection, Parameter Fitting, and Analysis. For the German dataset, we digitalized longitudinal VL data from throat swabs of the nine infected individuals reported in Wolfel et al. (3). The infected individuals are young to middle-aged professionals, without underlying disease, who were identified because of known close contact with an index case. All patients were hospitalized but had a comparatively mild clinical course of disease. For the NBA dataset, we used data reported in Kissler et al. (20). We included nine individuals for whom multiple detectible VL measurements were available before the viral peak. Note that VLs were reported in copies/swab by Wolfel et al. (3) and in copies/mL in Kissler et al. (20). Since we did not find significant differences in parameter estimates between the two datasets (*Results*), the unit of choice/reporting may not strongly impact our results. For consistency, we use copies/mL as the reporting unit.

We use a population approach, based on nonlinear mixed effect modeling (unless specified otherwise), to fit the model simultaneously to VL data from the two datasets, using the software Monolix (Lixoft SAS, Antony, France). We calculated correlations between the incubation periods and the fractions of predicted presymptomatic transmission using Pearson correlation.

The Model for Infectiousness. To calculate the probability of transmission given a typical contact of duration τ, we assume that τ is small enough (on the order of minutes or hours) that the total VL in the URT of the donor and thus the level of infectious viruses, V_{inf}, is approximately constant during the contact between time t and $t + \tau$. We then assume that the number of infectious viruses shed per unit time is μV_{inf}, where μ is a constant. Of these, a fraction, φ, reaches the URT of the recipient. Then on average, the total number of infectious viruses reaching the recipient for a contact of duration, τ, is $n = \varphi\mu\tau V_{inf}$. Airborne pathogens tend to be randomly distributed in the air (50). Thus, we assume the number of infectious viruses reaching the recipient during a contact is a random variable *X* that is Poisson distributed with parameter *n*. We further assume that each infectious virus that reaches the recipient has a probability ν to successfully establish infection and that if *X* viruses reach the recipient the probability to establish an infection is given by the binomial distribution Bin(*X*, ν). However since *X* follows a Poisson distribution, one can show the distribution of the number of viruses that successfully establish an infection follows a Poisson distribution with parameter $\lambda = n\nu = \varphi\mu\tau V_{inf}\nu = \theta\frac{V_T(t)^h}{V_T(t)^h + K_m^h}$, for the saturation model where $\theta = \varphi\mu\tau V_m\nu$.

Ke et al.
In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness

PNAS | 7 of 9
https://doi.org/10.1073/pnas.2111477118

POPULATION BIOLOGY

APPLIED MATHEMATICS

www.manaraa.com

Then, the probability of one or more virions generating a successful transmission event for a typical contact at time t is given by Eq. 1.

Estimating the Expected Serial Intervals and $R_{0,epi}$ from Infectiousness Profiles. To calculate the expected serial interval (or the generation interval), we assume that contacts are randomly distributed over time. Then, the expected serial interval for the ith individual, SI_i, can be calculated as

$$SI_i = \frac{\int_0^\infty t\, p_i(t)dt}{\int_0^\infty p_i(t)dt}, \qquad [6]$$

where $p_i(t)$ is the probability of transmission (Eq. 1) given a typical contact for individual i. The mean serial interval across all individuals in our study is calculated as the mean of the SI_i values calculated for all the individuals in the two datasets.

To calculate the expected epidemiological reproductive number, we assume that there are on average 13.4 contacts of a relatively short duration per day according to the estimates in Mossong et al. (26). Then, the expected epidemiological reproductive number for individual i is calculated as

$$R_{0,epi,i} = 13.4 \int_0^\infty p_i(t)dt. \qquad [7]$$

The mean epidemiological reproductive number across all individuals in the two datasets, $R_{0,epi}$, is calculated by taking the average of $R_{0,epi,i}$ across all individuals.

Note that the calculation of $R_{0,epi}$ above is a rough approximation because it implicitly makes the simplifying assumption that contacts are randomly distributed over time and every individual has the same number of contacts per day. This is used in our study to show that the choice of parameter values (for θ, h and K_m) are broadly consistent with estimates of epidemiological parameters such as the mean serial interval and $R_{0,epi}$. However, it should not be treated as an exact expression. See ref. 51 for discussion of formally calculating $R_{0,epi}$ in the context of SARS-CoV-2 transmission.

Model and Assumptions for Evaluating Testing Strategies. Several studies have remarked that testing sensitivity in clinical practice can be much lower than the theoretical detection limit would indicate. For example, Kucirka et al.

(40) suggested that the sensitivity of a RT-PCR test depends on the time since infection (a reflection of the VL) and that it is never more than 80%. Although there are many RT-PCR test platforms and protocols in use, the general sensitivity over the infection duration is likely not substantially different. To examine testing protocols under the best of circumstances, we assume much better performance for RT-PCR tests than suggested by Kucirka et al. (40), with no detection if the VL is below 10^3 copies/mL but 90% sensitivity for any VL above that (*SI Appendix*, Fig. S6B). We compare this test with an antigen test with characteristics as presented in Kohmer et al. (29), who compared the performance of several antigen tests with the results of RT-PCR. Based on their data for the SARS-CoV-2 Rapid Antigen Test (Roche Diagnostics) versus the VL in the sample, we fit the performance of the test to a logistic type relation between VL and positivity detection yielding the curve shown in *SI Appendix*, Fig. S6C (see *SI Appendix* for further details). An infected person's probability of being detected is a Bernoulli trial based on the sensitivity of the test (as in *SI Appendix*, Fig. S4).

1. N. Zhu et al., China Novel Coronavirus Investigating and Research Team, A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
2. X. He et al., Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
3. R. Wölfel et al., Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).
4. M. M. Arons et al., Public Health–Seattle and King County and CDC COVID-19 Investigation Team, Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* **382**, 2081–2090 (2020).
5. A. Handel, P. Rohani, Crossing the scale from within-host infection dynamics to between-host transmission fitness: A discussion of current assumptions and knowledge. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 1675 (2015).
6. D. B. Larremore et al., Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.* **7**, eabd5393 (2021).
7. N. K. Shrestha et al., Distribution of transmission potential during nonsevere COVID-19 Illness. *Clin. Infect. Dis.* **71**, 2927–2932 (2020).
8. A. Sakurai et al., Natural history of asymptomatic SARS-CoV-2 infection. *N. Engl. J. Med.* **383**, 885–886 (2020).
9. L. M. Yonker et al., Pediatric severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Clinical presentation, infectivity, and immune responses. *J. Pediatr.* **227**, 45–52.e5 (2020).
10. R. Ke et al., Longitudinal analysis of SARS-CoV-2 vaccine breakthrough infections reveal limited infectious virus shedding and restricted tissue distribution. *medRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.08.30.21262701 (Accessed 9 September 2021).
11. S. M. Kissler et al., Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated individuals. *medRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.02.16.21251535 (Accessed 22 September 2021).
12. M. Levine-Tiefenbrun et al., Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nat. Med.* **27**, 790–792 (2021).
13. M. J. Mina, R. Parker, D. B. Larremore, Rethinking Covid-19 test sensitivity—A strategy for containment. *N. Engl. J. Med.* **383**, e120 (2020).
14. P. Czuppon et al., Success of prophylactic antiviral therapy for SARS-CoV-2: Predicted critical efficacies and impact of different drug-specific mechanisms of action. *PLOS Comput. Biol.* **17**, e1008752 (2021).
15. A. Gonçalves et al., Timing of antiviral treatment initiation is critical to reduce SARS-CoV-2 viral load. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 509–514 (2020).
16. A. Goyal, D. B. Reeves, E. F. Cardozo-Ojeda, J. T. Schiffer, B. T. Mayer, Viral load and contact heterogeneity predict SARS-CoV-2 transmission and super-spreading events. *eLife* **10**, e63537 (2021).
17. N. Néant et al., French COVID Cohort Investigators and French Cohort Study groups, Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017962118 (2021).
18. R. Ke et al., Daily sampling of early SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *medRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.07.12.21260208 (Accessed 29 August 2021).
19. M. M. Böhmer et al., Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series. *Lancet Infect. Dis.* **20**, 920–928 (2020).
20. S. M. Kissler et al., Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLoS Biol.* **19**, e3001333 (2021).
21. P. Baccam, C. Beauchemin, C. A. Macken, F. G. Hayden, A. S. Perelson, Kinetics of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599 (2006).
22. K. A. Pawelek et al., Modeling within-host dynamics of influenza virus infection including immune responses. *PLoS Comput. Biol.* **8**, e1002588 (2012).
23. A. Goyal, E. F. Cardozo-Ojeda, J. T. Schiffer, Potency and timing of antiviral therapy as determinants of duration of SARS-CoV-2 shedding and intensity of inflammatory response. *Sci. Adv.* **6**, eabc7112 (2020).
24. C. E. Samuel, Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809 (2001).
25. R. A. Saenz et al., Dynamics of influenza virus infection and pathology. *J. Virol.* **84**, 3974–3983 (2010).
26. J. Mossong et al., Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74 (2008).
27. R. Jaafar et al., Correlation between 3790 quantitative polymerase chain reaction-positives samples and positive cell cultures, including 1941 severe acute respiratory syndrome coronavirus 2 isolates. *Clin. Infect. Dis.* **72**, e921 (2021).
28. T. C. Jones et al., Estimating infectiousness throughout SARS-CoV-2 infection course. *Science* **373**, eabi5273 (2021).
29. N. Kohmer et al., The comparative clinical performance of four SARS-CoV-2 rapid antigen tests and their correlation to infectivity in vitro. *J. Clin. Med.* **10**, 328 (2021).
30. A. Marc et al., Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *eLife* **10**, e69302 (2021).
31. V. Stadnytskyi, C. E. Bax, A. Bax, P. Anfinrud, The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11875–11877 (2020).
32. C. Fraser, T. D. Hollingsworth, R. Chapman, F. de Wolf, W. P. Hanage, Variation in HIV-1 set-point viral load: Epidemiological analysis and an evolutionary hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17441–17446 (2007).
33. A. Handel, I. M. Longini, Jr, R. Antia, Neuraminidase inhibitor resistance in influenza: Assessing the danger of its generation and spread. *PLOS Comput. Biol.* **3**, e240 (2007).

8 of 9 | **PNAS**
https://doi.org/10.1073/pnas.2111477118

Ke et al.
In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness

www.manaraa.com

34. H. Y. Cheng *et al.*, Taiwan COVID-19 Outbreak Investigation Team, Contact tracing assessment of COVID-19 transmission dynamics in taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern. Med.* **180**, 1156–1163 (2020).

35. L. Luo *et al.*, Contact settings and risk for transmission in 3410 close contacts of patients with COVID-19 in Guangzhou, China: A prospective cohort study. *Ann. Intern. Med.* **173**, 879–887 (2020).

36. W. Zhang *et al.*, Secondary transmission of coronavirus disease from presymptomatic persons, China. *Emerg. Infect. Dis.* **26**, 1924–1926 (2020).

37. S. T. Ali *et al.*, Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109 (2020).

38. R. Ke, E. Romero-Severson, S. Sanche, N. Hengartner, Estimating the reproductive number R_0 of SARS-CoV-2 in the United States and eight European countries and implications for vaccination. *J. Theor. Biol.* **517**, 110621 (2021).

39. M. A. Johansson *et al.*, SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Netw. Open* **4**, e2035057 (2021).

40. L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, J. Lessler, Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure. *Ann. Intern. Med.* **173**, 262–267 (2020).

41. A. D. Paltiel, A. Zheng, R. P. Walensky, Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States, Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. *JAMA Netw. Open* **3**, e2016818 (2020).

42. L. Ferretti *et al.*, The timing of COVID-19 transmission. *medRxiv* [Preprint] (2020). https://doi.org/10.1101/2020.09.04.20188516. Accessed 9 September 2021.

43. R. A. P. M. Perera *et al.*, SARS-CoV-2 virus culture and subgenomic RNA for respiratory specimens from patients with mild coronavirus disease. *Emerg. Infect. Dis.* **26**, 2701–2704 (2020).

44. J. J. A. van Kampen *et al.*, Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19). *Nat. Commun.* **12**, 267 (2021).

45. H. H. Mostafa *et al.*, SARS-CoV-2 infections in mRNA vaccinated individuals are biased for viruses encoding spike E484K and associated with reduced infectious viral loads that correlate with respiratory antiviral IgG levels. *medRxiv* [Preprint] (2021). 10.1101/2021.07.05.21259105 (Accessed 3 October 2021).

46. L. Zou *et al.*, SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **382**, 1177–1179 (2020).

47. M. Gandhi, D. S. Yokoe, D. V. Havlir, Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19. *N. Engl. J. Med.* **382**, 2158–2160 (2020).

48. K. Best, A. S. Perelson, Mathematical modeling of within-host Zika virus dynamics. *Immunol. Rev.* **285**, 81–96 (2018).

49. A. García-Sastre, C. A. Biron, Type 1 interferons and the virus-host relationship: A lesson in détente. *Science* **312**, 879–882 (2006).

50. G. N. Sze To, C. Y. Chao, Review and comparison between the Wells-Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor Air* **20**, 2–16 (2010).

51. S. W. Park *et al.*, Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: Framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. *J. R. Soc. Interface* **17**, 20200144 (2020).

POPULATION BIOLOGY

APPLIED MATHEMATICS

www.manaraa.com