

RESEARCH ARTICLE

# Random time-shift approximation enables hierarchical Bayesian inference of mechanistic within-host viral dynamics models on large datasets

Dylan J. Morris<sup>1</sup>\*, Lauren Kennedy, Andrew J. Black<sup>1</sup>

School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, South Australia, Australia

\* [dylan.morris@adelaide.edu.au](mailto:dylan.morris@adelaide.edu.au)



## Abstract

Mechanistic mathematical models of within-host viral dynamics are tools for understanding how a virus' biology and its interaction with the immune system shape the infectivity of a host. The biology of the process is encoded by the structure and parameters of the model that can be inferred statistically by fitting to viral load data. The main drawback of mechanistic models is that this inference is computationally expensive because the model must be repeatedly solved. This limits the size of the datasets that can be considered or the complexity of the models fitted. In this paper we develop a much cheaper inference method for this class of models by implementing a novel approximation of the model dynamics that uses a combination of random and deterministic processes. This approximation also properly accounts for process noise early in the infection when cell and virion numbers are small, which is important for the viral dynamics but often overlooked. Our method runs on a consumer laptop and is fast enough to facilitate a full hierarchical Bayesian treatment of the problem with sharing of information to allow for individual level parameter differences. We apply our method to simulated datasets and a reanalysis of COVID-19 monitoring data in an National Basketball Association cohort of 163 individuals.

## OPEN ACCESS

**Citation:** Morris DJ, Kennedy L, Black AJ (2025) Random time-shift approximation enables hierarchical Bayesian inference of mechanistic within-host viral dynamics models on large datasets. PLoS Comput Biol 20(12): e1013775. <https://doi.org/10.1371/journal.pcbi.1013775>

**Editor:** Edward M. Hill, University of Liverpool, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

**Received:** June 19, 2025

**Accepted:** November 20, 2025

**Published:** December 4, 2025

**Copyright:** © 2025 Morris et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All relevant data are available in the following public Github repository: [https://github.com/djmorris7/random\\_time\\_shift-within\\_host\\_inference](https://github.com/djmorris7/random_time_shift-within_host_inference).

## Author summary

Understanding how viruses reproduce within an infected host is crucial for predicting disease progression and evaluating treatments. One way to study this is by using mathematical models that describe how viruses reproduce in the body, such as in the respiratory tract. These models help understand key biological processes, but they require significant computing power, making it difficult to analyse data from large groups of individuals. In this study we developed a new, faster, method for analysing viral load data. Our approach combines randomness,

**Funding:** DJM acknowledges funding from the Australian Government through a Research Training Program Scholarship. LK acknowledges partial funding from the National Institutes of Health (R01AG067149). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

to account for chance events early in infection, with a more predictable model once the infection is well established. This allows us to speed up calculations while capturing key biological processes and maintaining a high degree of accuracy. Our method is efficient enough to run on a laptop, making it possible to study large datasets using a full statistical approach that accounts for individual differences. We tested our method on 200 simulated datasets and applied it to real-world data from 163 individuals in a COVID-19 monitoring study. Our results show that the model can accurately estimate key parameters and track viral load dynamics over time.

## 1 Introduction

Within-host viral dynamics (WHVD) models describe the progression of viral levels within an individual post-infection [1–3]. Depending on their complexity, these models can capture key biological processes such as viral replication, immune response, and clearance, or focus on summarising features like peak timing and growth/decline rates [1–4]. In a typical acute infection, the amount of virus in an individual—referred to as the viral load (VL)—grows exponentially before peaking (due to the immune response and/or the carrying capacity of the respiratory tract) and then declining [5]. Modelling this progression of virus inside a host can provide insight into important features like when an individual's infectiousness (VL) peaks, differences between variants of a disease, and the vaccination status of individuals [6], all of which are drivers of between-host (or transmission) dynamics [7–9] and are key in informing health policies [10].

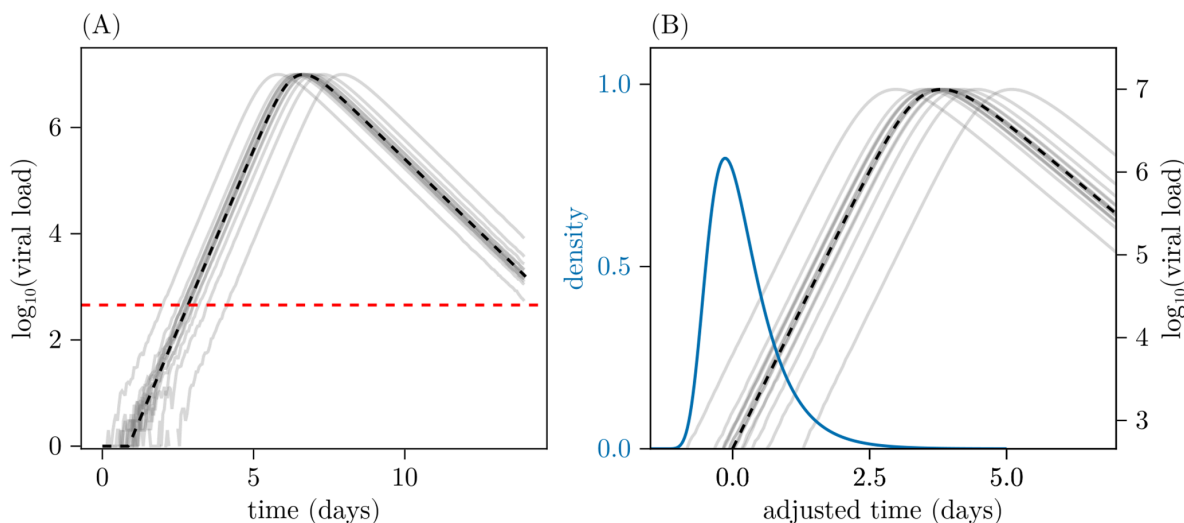
Mathematical models of WHVDs are broadly either phenomenological or mechanistic [4,5,11]. Phenomenological models are typically regression models that are able to capture key observed macroscopic (measurable) quantities like growth and decline rates as well as the timing of the peak of infection [4,11]. Such models can provide a good fit to data and predictions of the viral load trajectory but do not yield any understanding of the underlying biology of the system. Mechanistic modelling overcomes this by constructing models that explicitly track the populations of cells and virus in a region of the body [5,12–15]. Typically such a model is specified in terms of the biological interactions of the lower level entities from which solutions for the macroscopic viral load can be derived or simulated. The major drawback of such models is that they are much more expensive to fit to data because they do not have simple parametric solutions. One of the main contributions of our work here is to develop a radically faster inference approach, such that fitting complicated mechanistic models is now possible on basic consumer hardware.

Our second main contribution is the ability to fit a fully stochastic model rather than just a deterministic one. To put this in context, mechanistic models are themselves broadly either stochastic or deterministic. Deterministic models are much cheaper to solve, but stochastic models capture random variation from when the populations are initially small (i.e. early in the infection) that can still have an impact on the time for the viral load to peak [16,17] and ignoring this effect potentially biases inferences.

The additional complexity of stochastic models has led—with one exception [18]—to most inference approaches ignoring early time noise and only fitting simple deterministic models for the mean populations [5,8,11]. In this work we leverage the large scales over which the VL changes (from  $10^0$  to  $10^8$  virions per  $\mu\text{l}$  of plasma) to derive a novel approximation for the solutions of a fully stochastic WHVD model. This facilitates our first contribution by allowing the derivation of a fast and accurate approximation to the likelihood for VL time-series data. It also properly takes account of the early stochasticity in the process that is required for performing unbiased inference.

Our method is fast enough to fit a full hierarchical Bayesian model to datasets of hundreds of individuals, such as the COVID-19 testing dataset collected from National Basketball Association (NBA) players during the 2020–2021 seasons [4,6]. This dataset is among the most complete and detailed publicly available, featuring longitudinal measurements of individuals' VLs over time using common data formats found in the literature [2,5,11,19,20]. In contrast to others [4–6,18], we fit a full stochastic model to the 163 measured individuals and the inference algorithm is run on a consumer laptop in approximately one hour. The method can produce accurate posterior predictive VL trajectories and parameter estimates of the key within-host parameters (e.g. the within-host basic reproduction number,  $R_0$ ) that are consistent with previous studies of respiratory illnesses [2,5,11,19]. The hierarchical structure of the model allows us to share information from individuals with sufficient data to those whose data is limited, or noisy, during the growth and/or the decline phases of the infection. Our method can be easily extended to other datasets and other within-host models (i.e. models that include immune responses [8,21]) and other viral infections and so provides a general tool for fitting to WHVD data.

The key idea upon which our method builds is that the macroscopic dynamics of a fully stochastic WHVD model can be accurately approximated using a deterministic model with a random time-shift applied to the initial conditions [16,17]. Fig 1 illustrates this random time-shift idea. In Fig 1A, the grey lines show realisations of a stochastic model (defined fully later) and the black line shows the deterministic approximation to the stochastic model. At early times the process is strongly affected by stochasticity due to the small numbers of cells and virions, but this occurs below a detection threshold (indicated by the horizontal red line) where data is not available. Once the virion numbers grow large enough, growth



**Fig 1. Illustration of the time-shift concept.** (A) Example simulations from the fully stochastic model (grey lines) and the deterministic approximation (black). The detection limit for observations is indicated by the red dashed line. (B) The same simulations as shown in panel (A) but starting the plot from when the simulations cross the detection limit. Translating the deterministic trajectory through time approximates the stochastic solutions on the macroscale. The blue curve shows the probability density of this shift, defined by aligning threshold-crossing times and is equivalent to a shift in the initial conditions of the deterministic model. The models and parameters are described in Sects 2.1 and 2.2, with  $\theta = (8, 4, 1.7, 3, 10, 0)$  used in these simulations.

<https://doi.org/10.1371/journal.pcbi.1013775.g001>

becomes exponential, and appears deterministic. The early time stochasticity is reflected on the macroscale (the scale on which measurements can be made, i.e. above the detection threshold) in the time for the virus to peak. We see that the deterministic model (black line) accurately captures the growth and decline rates as well as the peak height, but not the randomness in the time to peak. But, by time shifting the initial conditions of the deterministic trajectory, while keeping the initial numbers fixed, the model can capture the randomness in the time to peak. The blue curve in Fig 1B shows the density of this random variable; our previous work details how this can be accurately approximated by a closed form function that can be calculated for a given model [17].

Within-host models are an ideal system for an application of this approximation method because the approximation is most accurate where we have data, i.e. when the populations have grown significantly. During the initial stages of the infection when virus and infected cell numbers are small, the approximation is less accurate, but in this region data is effectively unavailable anyway because there is a detection threshold for measurements (i.e. the data is censored) [2,5,11,19]. The important effect of stochasticity in this early time regime is to shift the time-to-peak, and this is captured accurately by the approximation.

Another way of looking at this approximation is that it allows fast forward simulation of the model without the need for costly stochastic simulation, for example using the Gillespie algorithm [22,23]. For any given set of parameters, only a *single* solution of the deterministic model and samples from the corresponding time-shift distribution are required. In many existing studies of this problem the initial viral load is treated as being unobserved and hence a parameter to be inferred [11] or fixed at some value [5]. The time-shift idea facilitates a shift in perspective from inferring the initial variable viral load at a fixed time, to instead inferring a fixed viral load at a random time. This turns out to be a key insight as it allows the derivation of a number of further accurate approximations for the likelihood function that provide large computational speed ups over standard stochastic simulation techniques, allowing inference on large datasets using a relatively complicated model.

We begin in Sect 2 by outlining our particular stochastic viral kinetics model and show mathematically how the dynamics can be approximated using the time-shift ideas. We then provide details of approximations that can be made as a result of how the time-shift is incorporated into the likelihood in order to improve the efficiency of our inference method. This provides the foundations for a Metropolis-Hastings within Gibbs sampler for sampling from the posterior distribution. In Sect 3 we apply it to two types of datasets. Firstly, multiple synthetic datasets are used to validate the method and quantify its inference capabilities in contexts where the true values of the parameters are known. We find that our approach is able to disentangle the measurement noise from the process noise and results in parameter estimates that are consistent with the simulated parameters. Secondly, we fit the model to the NBA COVID-19 dataset [4,6], where we find that our posterior predictive fits indicate that the method is able to produce a realistic fit to the data and obtain  $R_0$  estimates that are consistent with previous studies. Additionally, we show that the method is able to share information across individuals, to those with few observations and missing data, to estimate their viral dynamics.

## 2 Materials and methods

### 2.1 The viral kinetics model

In this work we consider one of the simplest viral kinetics model referred to as the Target-Cell-Limited ( $\mathcal{TCL}$ ) model with an eclipse-phase [2,24,25]. This model captures the basic cell infection cycle and viral production cycles observed during an infection while accounting for an appropriate delay (the eclipse phase) between a cell becoming infected and producing virus. We choose this simpler model and not something more complex (see [14] for an example) as some of the parameters are already unidentifiable from VL data [25,26] and so adding more complexity without additional data sources only exacerbates this. The method itself can easily be applied to more complex models if required. The  $\mathcal{TCL}$  model is commonly specified as a set of deterministic ordinary differential equations [2,24,27] for the total population counts, but here we will instead start with a stochastic version that captures the individual nature of the populations from

which we will derive the random time-shift approximation consisting of a deterministic ODE and the time-shift distribution itself.

The  $\mathcal{JCL}$  model tracks the numbers of susceptible target cells,  $S(t)$ , cells in the eclipse phase,  $E(t)$ , infected cells,  $I(t)$ , and free virus,  $V(t)$ , in an effective volume  $K$ , corresponding to the volume over which the within-host infection process occurs. The dynamics of these populations are modelled as a continuous-time Markov chain (CTMC) [28,29]  $\{\mathbf{X}(t), t \geq t_0\}$  where  $t_0$  is the time at which the infection is initiated, and is unknown, and the state of the system at time  $t \geq t_0$  is specified by the vector  $\mathbf{X}(t) = (S(t), E(t), I(t), V(t))$ . Adopting this formulation means that populations are non-negative integers. Target cells are uninfected cells that become infected when virus binds to them, a process that follows mass action kinetics [2,30]. The instantaneous rate at which target cells become infected is given by  $\alpha S(t)V(t)/K$ , where  $\alpha$  represents the per-virus infection rate of target cells. In most formulations of this model [1,2,5,27], the parameter  $\beta = \alpha/K$  is the parameter of interest. For consistency, we also take  $\beta$  as the parameter of interest in our model. Following infection, cells enter an eclipse phase, which follows an exponential distribution with mean  $1/\sigma$  days, during which infected cells are not yet actively producing virus. Infectious cells are then removed (either by immune clearance or cell death) after an exponentially distributed time with mean  $1/\delta$  days. While infected, cells produce new virions at rate  $\rho$  copies per day, which are cleared at rate  $c$  per day, due to immune response or other clearance mechanisms. The state transitions and rates of the model are summarised in Table 1.

There is a choice to be made regarding the initial condition,  $\mathbf{X}(t_0)$ . The infection is triggered by the entry of virus into the system, so ideally, the initial viral load,  $V(t_0)$ , would be set. However, this introduces complexity regarding the amount of virus required to cause an infection [12,31]. Smith et al. [31] show that the dynamics of the model are not significantly altered by taking the initial condition as a single infected cell, as any initial viral particles must initiate an infection by infecting one or more cells (at a rapid rate) before being cleared. Throughout this work we assume that  $S(t_0) = 8 \times 10^7$  based on Zitzmann et al. [5]. Hence, throughout this work, we consider viral infections triggered by a single infected (but not actively infectious) cell, i.e., the initial condition is  $\mathbf{X}(t_0) = (S(t_0) - 1, 1, 0, 0)$ . The methods presented here can be extended to handle more complex initial conditions, and extensions of this are discussed in Sect 4.

A pivotal quantity in such models is the basic reproduction number  $R_0$ , which is defined as the average number of secondary infected cells that are produced by an initially infected cell when the target cell population is fully susceptible [32]. The importance of this quantity is that it is a measure of the virus' ability to invade the population of susceptible cells and result in an acute infection [33]. If  $R_0 \leq 1$  then the virus will die out before infecting a large number of cells. If  $R_0 > 1$  then there is a chance that the virus will invade the cell population and result in an acute infection. The basic reproduction number for this model can be computed through the next generation matrix method [34] and is given by [5],

$$R_0 = \frac{\beta S(t_0) \rho}{\delta c}. \tag{1}$$

## 2.2 Random time-shift approximation

Here, we show that the macroscopic dynamics of the stochastic model described in Sect 2.1 can be approximated by a deterministic ODE model, which corresponds to standard formulations with an added random time shift to the initial

**Table 1. Rates for the CTMC viral dynamics model.** Each row gives an event type, the associated change in state  $\mathbf{X}(t)$ , and the rate at which the event occurs in the  $\mathcal{JCL}$  model.

Event	$\Delta\mathbf{X}$	Rate
Infection	$(-1, +1, 0, 0)$	$\beta S(t)V(t)$
Cell becomes infectious	$(0, -1, +1, 0)$	$kE(t)$
Infectious cell recovery	$(0, 0, -1, 0)$	$\delta I(t)$
Production of virus	$(0, 0, 0, +1)$	$\rho V(t)$
Clearance of virus	$(0, 0, 0, -1)$	$cV(t)$

<https://doi.org/10.1371/journal.pcbi.1013775.t001>

conditions. The deterministic approximation follows from a mean-field approximation of the CTMC model in the limit  $K \rightarrow \infty$  (i.e., for large volume). In the context of respiratory infections,  $K$  corresponds to the total volume over which the within-host infection process occurs and can therefore be considered “large” in the sense that the number of interacting particles (e.g., cells and virions) is sufficiently high. Throughout this work, however, the exact value of  $K$  is unimportant, as it can be absorbed into the model parameters. Taking the large-volume limit yields a system of coupled ODEs for the cell and virus populations (see Sect 1 of S1 text for the derivation). This system governs the mean-field dynamics, denoted  $S_d(t), E_d(t), I_d(t), V_d(t)$ , where the subscript  $d$  indicates deterministic solutions. To maintain consistency between the stochastic and deterministic formulations, we again make the substitution  $\beta = \alpha/K$ , giving

$$\begin{aligned} \frac{dS_d(t)}{dt} &= -\beta S_d(t)V_d(t), \\ \frac{dE_d(t)}{dt} &= \beta S_d(t)V_d(t) - kE_d(t), \\ \frac{dI_d(t)}{dt} &= kE_d(t) - \delta I_d(t), \\ \frac{dV_d(t)}{dt} &= \rho I_d(t) - cV_d(t). \end{aligned} \tag{2}$$

These equations coincide with the ODEs typically used in purely deterministic modelling approaches [3,5].

It is difficult to work directly with the transmission parameter  $\beta$  as it is strongly dependent on the experimental setup in the study of interest (e.g., how viral load samples are collected) and negatively correlated with  $\rho$  [5]. Therefore, we reparameterise the model in terms of  $R_0$  and back-transform to obtain  $\beta$  when needed. We collect all the parameters in the vector  $\theta = (R_0, k, \delta, \rho, c, t_0)$ .

Denote the solution of Eqs (2),  $\mathbf{X}_d(t) = (S_d(t), E_d(t), I_d(t), V_d(t))$ , then this provides an approximation to the mean dynamics of the full stochastic CTMC  $\{\mathbf{X}(t), t \geq t_0\}$ , but does not capture the effect of the noise during the early-time phase when the numbers of infected cells and virions is low (see Fig 1 and discussion in Sect 1). The predominant effect of the stochasticity is to shift the time at which the peak VL occurs and subsequently the time at which the infection is cleared [16,17,35]. The work of Barbour et al. [16] formalises the connection between the shifted mean field kinetics and samples of the full stochastic process. They show that  $\{\mathbf{X}(t), t \geq t_0\}$  is well approximated on the macroscale, i.e. the growth and decline rates and peak time and height, by another stochastic process  $\{\mathbf{X}_d(t + \tau), t \geq t_0\}$  where  $\tau := \tau(\theta) \in \{-\infty\} \cup \mathbb{R}$  is a scalar random variable referred to as the *random time-shift* [16,17].

The practical importance of this result is that sample paths of  $\{\mathbf{X}(t), t \geq t_0\}$  resemble those of the deterministic model,  $\mathbf{X}_d(t)$ , except for a translation in time by the time-shift  $\tau$ . The case  $\tau = -\infty$  corresponds to when the number of virions and infected cells drops to zero, i.e. extinction of the virus within the host. Our interest lies in the modelled VL once  $V(t)$  has grown to be measurable, i.e. outside of the early-time small population regime. Approximating the stochastic process  $\{\mathbf{X}(t), t \geq t_0\}$  in this way means we can efficiently generate sample paths  $V(t)$  by solving for the mean field kinetics and obtaining the distribution of the time-shift. The process  $\{\mathbf{X}_d(t + \tau), t \geq t_0\}$  does not capture the noisy early dynamics (i.e. the actual stochastic fluctuations in the populations when these counts are small) but this is of little consequence as small VLs fall below a detection threshold and are therefore not measurable [4,5,11].

The distribution of the random variable  $\tau$  is derived by analysing a branching process approximation to the early time dynamics [17]. Previous work [17] showed that the density function for  $\tau$ , conditional on non-extinction, can be accurately approximated by a member of a family of functions with the form

$$f(\tau | \theta) = \frac{\rho \lambda \mu_W^d}{a^d \Gamma\left(\frac{d}{\rho}\right)} \exp\left(-\left(\frac{\mu_W e^{\lambda \tau}}{a}\right)^\rho + \lambda \tau d\right) \tag{3}$$

with  $(a, d, p)$  determining the particular member of the family. The quantity  $\lambda$ , representing the early growth rate of the model, and the time-shift parameters  $\zeta = (a, p, d)$  are numerically calculated from the branching process approximation to the CTMC (see [17] for concrete examples). The blue curve shown in Fig 1 is this density evaluated for the parameters given in the caption. Recall that as  $\tau$  is a function of  $\theta$ , the parameters  $\zeta$  are also a function of  $\theta$  but that this has been suppressed in the notation for clarity. One final and important point is that the time-shift is independent of the infection time  $t_0$ .

Computing the parameters of a single time-shift distribution is not expensive, but when incorporated into an Markov chain Monte Carlo (MCMC) routine, where they must be repeatedly evaluated, it can become a bottleneck. To overcome this, we train a neural network (NN) to learn the function  $g : \mathbb{R}^5 \mapsto \mathbb{R}^3$ , mapping  $\theta = (R_0, k, \delta, \rho, c)$  to  $\zeta = (a, d, p)$ . This approach, known as amortised optimisation [36–38], replaces repeated optimisation with a learned mapping, significantly decreasing the cost of computation during the running of the algorithm, but at the expense of initially training the network [38]. Full details of our implementation are provided in Sect 6 of S1 text. In testing we found that this modification to our method accelerates the computation of the time-shift parameters by roughly 200 times.

### 2.3 Viral load measurement model

The data we assume available for each individual is a time-series of the  $\log_{10}$  viral load (log-VL) measurements, denoted  $\mathbf{y} := (y_1, \dots, y_J)$ , where  $y_j := y_{t_j}$  is the log-VL at time  $t_j$ . In practice the  $t_j$  could be determined in a variety of ways depending on the data-collection process. For a longitudinal study like that run on the NBA cohort [4], daily observations are taken over a long period of time and only the start point of the time-series is effectively random. In a household study, for example, the first time for each individual in a given household is intrinsically linked to both the household being first tested and could come after individuals are infectious, meaning that the first times are random for each index case [39]. Additionally, the timings of the test between individuals are likely to be non-equidistant and also prone to variability as well. The data is left-censored at a threshold  $\eta$  (i.e.  $y_j = \eta$  means that  $y_j \leq \eta$ ) which describes the limit of detection of the assay used to measure the viral load [4, 11, 12].

We assume that the data for a single individual are noisy measurements of some underlying VL trajectory. Once more we note that the time-shift,  $\tau$ , is a random variable that is a function of the VL model parameters,  $\theta$  and that  $V_d(t, \theta)$  now has explicit dependence on the VL model parameters  $\theta$ . Inclusion of the time-shift means that the modelled trajectory is effectively translated in time and so the modelled trajectory is delayed or advanced depending on the sign of the time-shift. In the case where there is a slow start to infection (i.e. a negative time-shift) the modelled trajectory is translated to the right and so is effectively delayed. To handle inconsistencies with the modelled trajectory from applying the time-shift we simply set the VL over this time to the initial condition. For the case where the time-shift is positive, the opposite is true and the modelled trajectory is effectively translated to the left. In this case we can use the modelled trajectory without changes. The only other consideration is that the time  $t$  we evaluate the modelled trajectory at must be after the infection time, i.e.  $t \geq t_0$ . This can be summarised by defining

$$z(t, \tau, \theta) = \begin{cases} \log_{10} V_d(t + \tau, \theta), & \text{if } t \geq t_0 \text{ and } t + \tau \geq t_0 \\ -\infty, & \text{otherwise.} \end{cases} \quad (4)$$

Note that predicted viral loads below one particle (i.e.  $V(t) < 1$ ) were capped at a small pseudocount ( $10^{-5}$ ) before log-transformation to avoid numerical instabilities.

Let  $\mathbf{z}(\tau) := (z_1(\tau), \dots, z_J(\tau))$  denote the modelled viral load at the observation times  $(t_1, \dots, t_J)$  for a particular time-shift  $\tau$ , that is  $z_j(\tau) := z_{t_j}(\tau) := z(t_j, \tau, \theta)$ . Throughout this work we will use the convention that normal distributions are parameterised in terms of mean and standard deviation, i.e.  $\mathcal{N}(\mu, \sigma)$ , as this parallels the formulation in the code. We assume that

the observations are noisy measurements of the modelled trajectory,  $y_j = z_j(\tau) + \epsilon_j$  and  $\epsilon_j \sim \mathcal{N}(0, \kappa)$  for  $j = 1, \dots, J$  and some standard deviation  $\kappa$ , which is assumed to be unknown.

To provide further clarity on the idea of the time-shift and how this connects to the complete data generation process we now illustrate the idea in the context of a forward simulation framework. Algorithm 1 details the key steps to simulate an approximate stochastic realisation of the WHVD model given model parameters  $\theta$ . This process assumes some discrete times that the stochastic trajectory is sampled at and a particular value of the scale parameter for the observation noise,  $\kappa$ .

### Algorithm 1 Forward simulations using time-shifts.

**Inputs:**

- The WHVD model parameters  $\theta$ .
- Observation times  $\mathbf{t} = (t_1, \dots, t_J)$ .
- Scale parameter on the observation noise,  $\kappa$ .

- 1: Solve Eqs (2) to get the deterministic trajectory  $V(t)$  through time.
- 2: Compute the time-shift parameters  $\zeta = (\mathbf{a}, \mathbf{d}, \mathbf{p})$  using methods in Morris et. al [17].
- 3: Sample  $\tau \sim f(\tau | \theta)$  (given by Eq (3)).
- 4: Shift  $V(t)$  via Eq (4) to get the approximate stochastic trajectory  $\mathbf{z}(\mathbf{t}, \tau, \theta)$ .
- 5: **Return**  $\mathbf{z}(\mathbf{t}, \tau, \theta)$ .

## 2.4 Mixed effects model

In this section we provide the general form (deferring details for priors and the calculation of the likelihood to Sects 2.5 and 2.6, respectively) of our mixed effects model. This includes detailing the data and writing down the posterior distribution.

We assume a simple experimental setup of a sample of  $N$  individuals where  $\theta_i$  denotes the VL model parameters for individual  $i$  and  $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,J_i})$  denotes the time-series of their measured VL, which consists of  $J_i$  observations. Define  $\Theta = \{\theta_1, \dots, \theta_N\}$  and  $\mathcal{D} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$  as the collections of parameters and data, respectively. We assume that the subject-specific parameters,  $\theta_i$ , are drawn from a population distribution governed by what we refer to as the shared parameters  $\phi$  (the exact relationships are defined in Sect 2.6). The joint posterior is given by,

$$f(\Theta, \phi, \kappa | \mathcal{D}) \propto \left[ \prod_{i=1}^N f(\mathbf{y}_i | \theta_i, \kappa) f(\theta_i | \phi) \right] f(\phi) f(\kappa). \tag{5}$$

The difficult part of evaluating Eq (5) is in the calculation of the marginal likelihood terms,  $f(\mathbf{y}_i | \theta_i, \kappa)$ , due to the need to marginalise over the distribution of trajectories.

## 2.5 Likelihood approximations

In this section, we outline the construction of the *marginal likelihood* for an individual, breaking it down into its components: the time-shift and the *path likelihood*, the likelihood function given a viral load (VL) sample path. This approach mirrors the general method used in the literature for constructing the likelihood in purely deterministic models [5,11,40].

The *marginal likelihood* (likelihood for individual  $i$ 's data,  $\mathbf{y}_i$ , given the parameters,  $\theta_i$ ) is given by  $f(\mathbf{y}_i | \theta_i, \kappa)$ . Since the observation process depends on an unobserved random time-shift  $\tau$ , we marginalise over  $\tau$ , considering only the non-extinction case

$$f(\mathbf{y}_i | \theta_i, \kappa) = \int f(\mathbf{y}_i | \tau, \theta_i, \kappa) f(\tau | \theta_i) d\tau. \tag{6}$$

The conditional likelihood given  $\tau$  depends on the system's trajectory  $\mathbf{z}_i(\tau)$ , which is deterministic for a given  $\tau$  and  $\theta$ , so we can rewrite the marginal likelihood as

$$f(\mathbf{y}_i | \theta_i, \kappa) = \int f(\mathbf{y}_i | \mathbf{z}_i(\tau), \theta_i, \kappa) f(\tau | \theta_i) d\tau. \quad (7)$$

Observations  $y_{i,j}$  correspond to viral load (VL) measurements at times  $t_j$ , which are assumed to be independent given the model trajectory. The deterministic trajectory  $\mathbf{z}_{i,j}(\tau) := \mathbf{z}(t_j, \tau, \theta_i)$  represents the modelled VL at time  $t_j$  for individual  $i$ . Given  $\mathbf{z}_{i,j}(\tau)$ , the conditional likelihood of the observed data  $\mathbf{y}_i$  is

$$f(\mathbf{y}_i | \mathbf{z}_i(\tau), \theta_i, \kappa) = \prod_{j=1}^{J_i} f(y_{i,j} | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa), \quad (8)$$

where the terms inside the product are given by

$$f(y_{i,j} | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa) = \begin{cases} \mathcal{N}_{\text{PDF}}(y_{i,j} | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa), & \text{for } y_{i,j} > \eta, \\ \mathcal{N}_{\text{CDF}}(\eta | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa), & \text{for } y_{i,j} \leq \eta \end{cases}$$

where  $\mathcal{N}_{\text{PDF}}(y_{i,j} | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa)$  and  $\mathcal{N}_{\text{CDF}}(\eta | \mathbf{z}_{i,j}(\tau), \theta_i, \kappa)$  are the probability density function, and cumulative distribution function (CDF), respectively, of a  $\mathcal{N}(\mathbf{z}_{i,j}(\tau), \kappa)$  random variable. The CDF terms account for observations below the detection threshold [11]. To distinguish this from other likelihood terms, we refer to Eq (8) as the *path-likelihood* for a single individual. We note here that if the VL model is replaced with the deterministic model without the time-shift, then this is equivalent to the time-shift density collapsing to a point mass at  $\tau = 0$ . Hence, the likelihood is simply  $f(\mathbf{y}_i | \mathbf{z}_i(0), \theta_i, \kappa)$  as often seen in VL kinetics modelling [5,8,11,40].

The marginal likelihood, Eq (7), can be evaluated by substituting the path-likelihood, Eq (8), and integrating. However, it involves an integral over a function that does not admit an analytical solution, so numerical methods are required. Numerical integration requires repeated computation of the modelled viral load as we integrate over  $\tau$ , which appears inside the ODE solution (i.e.,  $\mathbf{z}_i(\tau)$ ). Evaluating the ODE at each integration step, even with interpolation, is computationally expensive.

To address this, we consider a Laplace approximation [41, Chapter 27] of the interior of the integral of the marginal likelihood. The approximation is constructed by noting that for fixed  $\theta_i, \kappa$  and data  $\mathbf{y}_i$ , the marginal likelihood is simply a function of  $\tau$ . Let

$$h(\tau) = f(\mathbf{y}_i | \mathbf{z}_i(\tau), \theta_i, \kappa) f(\tau | \theta_i), \quad (9)$$

and let  $g(\tau) = \log h(\tau)$ . Then the Laplace approximation to the marginal likelihood (Eq (7)) is given by (see Sect 4 of S1 text for derivation)

$$f(\mathbf{y}_i | \theta_i, \kappa) \approx \frac{\sqrt{2\pi} h(\tau_0)}{\sqrt{-g''(\tau_0)}}, \quad (10)$$

noting that  $g(\tau)$  is assumed to be concave (implicit in applying the Laplace approximation), which implies  $g''(\tau) < 0$  around  $\tau_0$ . The quantity  $\tau_0$ , is the maximiser of Eq (9) and can be estimated efficiently using optimisation. The term,  $(-g''(\tau_0))^{-1/2}$ , is the scale term and has a more complicated form but can also be computed through either numerical differentiation or automatic differentiation of the log of Eq (9). In this work, all our code is written in the Julia programming language [42],

and the package ForwardDiff.jl [43] enables efficient computation of exact derivatives through forward mode automatic differentiation. In Sect 4 of S1 text, we plot profile likelihoods—i.e., the likelihood with all parameters fixed except for one, which is allowed to vary—for a simulated VL trajectory to show the accuracy of the approximation in Eq (10) versus the exact calculation in Eq (7). The profile likelihoods had a maximum (absolute) error of  $< 10^{-7}$  at the tested points, which indicates strong agreement between our Laplace approximation and the exact likelihood.

## 2.6 Priors

In this section, we detail our choices of priors and fixed parameters. We center our priors at values consistent with other studies of COVID-19, and use weakly informative priors with higher variance to account for uncertainty in these estimates. Prior predictive checks confirm that the effective priors on individual parameters remain weakly informative. Our primary focus in this work is methodological, demonstrating how to incorporate stochasticity into WHVD models and speed up the inference, rather than conducting a thorough analysis of this data. Moreover, because our hierarchical model is fit to a larger number of individuals than in previous studies, the population-level estimates are better informed by the data and less reliant on the prior. While hierarchical models can be sensitive to prior misspecification, the larger sample size here mitigates this concern, as the data dominate the inference.

Due to identifiability concerns, some VL parameters are fixed at values that are well established in the literature. We assume that the eclipse phase of newly infected cells,  $k_i$ , and the clearance rate of virus,  $c_i$ , are fixed at  $k_i = 4 \text{ d}^{-1}$  and  $c_i = 10 \text{ d}^{-1}$  for  $i = 1, \dots, N$ , respectively [5]. The individual-level parameters,  $R_{0,i}$ ,  $\delta_i$ , and  $\rho_i$ , are all assumed to be normally distributed (truncated to be non-negative) with unknown mean and scale parameters. Here we detail the hyper-priors on these parameters. As outlined in Sect 2.2, we parameterise the model in terms of the basic reproduction number,  $R_0$ . Table 2 gives some estimates of  $R_0$  for COVID-19 and the dataset size (number of individuals) from the literature. None of these studies implemented hierarchical models and instead fit each individual time-series. To capture the range of values in Table 2, we assume that

$$\mu_{R_0} \sim \text{Gamma}\left(\frac{10}{3}, 3\right).$$

This choice results in a mean of 10 with a 95% quantile interval of (2.3,23.3), providing a weakly informative prior centered near the commonly reported estimates.

To parameterise the priors on the mean infectious period,  $\mu_\delta$ , and the mean viral replication rate,  $\mu_\rho$ , we draw on previous studies fitting similar models to COVID-19 data (see [44,45] and the SI of [5]). We assume

$$\mu_\delta \sim \text{Gamma}(26, 0.05),$$

which corresponds to a mean removal rate of  $1.3 \text{ d}^{-1}$  and a 95% interval of (0.85,1.85), consistent with published estimates. The parameter  $\mu_\rho$  is difficult to assign a prior due to identifiability issues similar to those faced by  $\beta$  [5,18]. To address this, we assume

$$\mu_\rho \sim \text{Gamma}(10, 0.3),$$

**Table 2. Estimated  $R_0$  values from published literature.** Estimates and 95% CI where available of  $R_0$  from published models. The number of individuals in the study are also provided. The dataset column indicates what dataset was of interest for that particular study, with NBA denoting a subset of our dataset and other denoting a different dataset.

Estimate (95% CI)	Number of individuals	Reference	Dataset
8.2 (NA)	25	[5, SI]	NBA
13.51 (4.5, 45.12)	9	[44]	Other
14.2 (NA)	6	[18]	NBA

<https://doi.org/10.1371/journal.pcbi.1013775.t002>

with median  $2.9 \text{ d}^{-1}$  (roughly based on fits in the SI of [5]) and a 95% interval of (1.44, 5.13). This choice reflects prior knowledge while remaining flexible enough for the data to inform the estimate. Note that the main-text model in [5] includes an immune response, leading to a much larger magnitude ( $\sim 100$ ), so direct numerical comparison is not appropriate.

The remaining parameters are assigned weakly informative priors that were selected to guide the fitting process but are not taken from the literature. We assign Half-Normal priors on  $\sigma_{R_0}$ ,  $\sigma_\delta$ , and  $\sigma_\pi$

$$\begin{aligned}\sigma_{R_0} &\sim \mathcal{N}_+(0, 3), \\ \sigma_\delta &\sim \mathcal{N}_+(0, 1), \\ \sigma_\rho &\sim \mathcal{N}_+(0, 3).\end{aligned}$$

The hyper-prior for  $\sigma_{R_0}$  and  $\sigma_\rho$  have a larger standard deviation compared with  $\sigma_\delta$  as we found in our testing that a  $\mathcal{N}_+(0, 1)$  over-constrained the fitted estimates which caused excessive shrinkage of the individual level parameter estimates to the population mean. The standard deviation of the measurement noise,  $\kappa$ , is also assigned a weakly informative prior

$$\kappa \sim \mathcal{N}_+(0, 1).$$

The priors on the infection times are also chosen to be weakly informative, where we only make the assumption that they are most likely to occur prior to the peak VL measurement,

$$t_{0_i} \sim \text{Gumbel}(-7, 3).$$

This choice results in a median time of infection (pre-peak VL) of  $-5.9$  days. This prior assumes infection occurs most likely before peak VL but allows for post-peak infection times, which may occur in some cases due to measurement variability or delayed detection. Typical scenarios where this prior is useful is when the peak VL is a low measurement or all measurements are tightly clustered in time so it's difficult to distinguish whether the peak is early or late on in the infection. This distribution is consistent with the incubation periods of COVID-19 (pre-Omicron variant) [46] when assuming that peak infectiousness (i.e. peak VL) corresponds to symptom onset. For ease of notation we will denote the set of all hyper-parameters as

$$\phi = (\mu_{R_0}, \mu_\delta, \mu_\rho, \sigma_{R_0}, \sigma_\delta, \sigma_\rho).$$

## 2.7 Inference

The form of Eq (5) suggests a Gibbs sampler that updates the individual-level parameters ( $i = 1, \dots, N$ ) and the shared parameters ( $\phi, \kappa$ ) via their conditionals. However, these conditionals are not available in closed form and cannot be sampled from directly, so we employ Metropolis-Hastings updates within a Gibbs sampling framework [47, Chapter 10]. Full algorithmic details of the sampler, including parallelisation strategies, are provided in Sect 5 of S1 text.

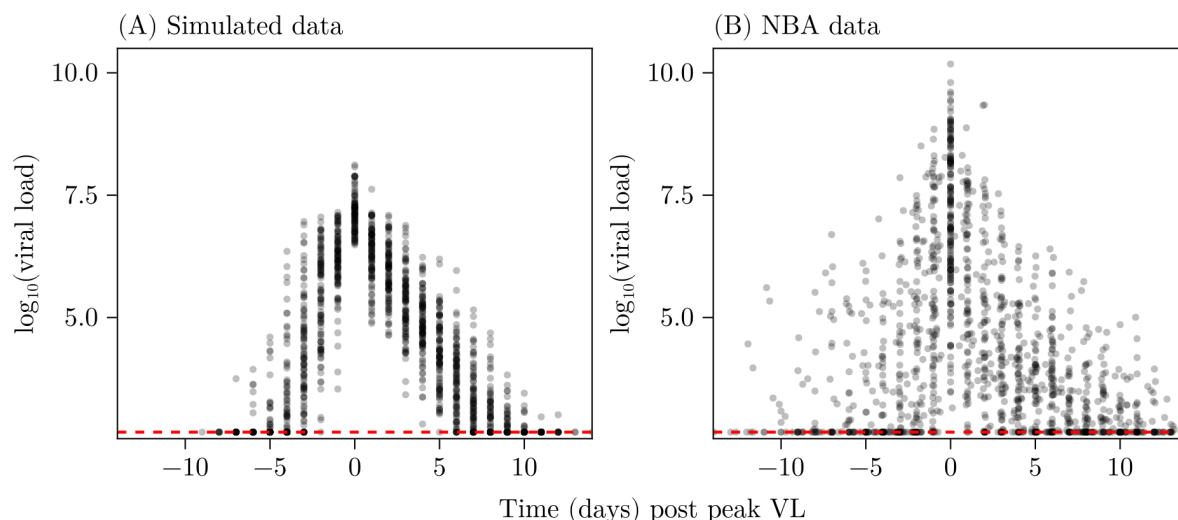
All analyses in this paper were implemented in the Julia programming language [42], which provides seamless integration of model specification, automatic differentiation, and inference within a single environment. For solving ordinary differential equations, we used the high-performance solvers provided by OrdinaryDiffEq.jl [48]. The neural network used to approximate the time-shift distributions was implemented and trained using Flux.jl [49]. Automatic differentiation was performed using ForwardDiff.jl [43].

## 2.8 Datasets

In this analysis, we use a collection of datasets: a published dataset of COVID-19 VL measurements collected during the 2020–2021 NBA season restart, and 200 synthetic datasets for method validation. Each synthetic dataset serves as a controlled setting with frequent, low-noise observations and known parameters. The purpose of the simulated datasets is to validate the method can recover parameters in an ideal setting. Each dataset consists of  $N = 100$  individual time-series and are generated using our prior model for the hyperparameters  $\phi$ . The full data generation process is detailed in Sect 2.8.1, and an example simulated cohort is shown in Fig 2A.

The NBA published COVID-19 monitoring data on players and staff during the 2019–2020 season restart [4,6]. Individuals were regularly tested for SARS-CoV-2 from mid-2020 to the season’s end, yielding log-transformed viral load measurements relative to peak VL. Other studies using mechanistic models [5,18] had to selectively include individuals with frequent pre- and post-peak observations; in our analysis we fit to a much larger set of individuals, many with much poorer data, simultaneously. There are 241 individuals in total (68 in [6] and 173 in [4]) but some of these time-series are quite noisy so we performed some pre-processing before fitting. Following the approach of Zitzmann et al. [5] we excluded individuals with viral loads detected outside a 14-day window around peak VL, as our model does not account for complex behaviours like long-term infections.

The limit of detection (LOD) in the NBA dataset is 2.658 log copies/mL and frequent testing and individuals experiencing only single infections resulted in many observations (i.e. some individuals having 80% LOD observations) at this threshold for each individual. Since it is most likely that individuals are only actually positive for some number of days pre/post the time where they have above LOD observations, we truncated the time series, retaining only the LOD measurements such that above-threshold observations remained between them. In some cases, multiple values below the LOD occur before another measurement slightly above the detection limit which complicates fitting. To address this, we further truncated time series at three consecutive threshold observations. Finally, there were some short time-series with less than 2 observations above the LOD and these were excluded. After filtering our dataset contains 163 of the original 241 timeseries. Note that we use the same post-processing procedure and LOD for the synthetic dataset. The NBA



**Fig 2. Scatter plots of an example simulated dataset and the NBA dataset.** (A) An example synthetic VL dataset generated according to the process in Sect 2.8.1. (B) The NBA VL dataset for the 2020–2021 seasons. In each panel an individual data point corresponds to a single measurement for a single individual. Points are shown as semi-transparent to approximate the “density” of points, highlighting outliers in panel (B). The red dashed line indicates the limit of detection.

<https://doi.org/10.1371/journal.pcbi.1013775.g002>

data is shown in Fig 2B alongside an example simulated cohort in panel Fig 2A. The NBA data (Fig 2B) exhibits visibly higher variance across the entire time horizon compared to the simulated dataset (Fig 2A). This is primarily because the synthetic data were deliberately generated with lower noise to enable more controlled validation of our method.

Our hierarchical model shares information from individuals with high information to improve estimates for those with sparse data, a process known as partial pooling [50, Chapter 5]. Previous analyses of this dataset [4–6,12] used deterministic models [5,12] or piecewise linear functions [4,6,12], whereas our model explicitly accounts for stochasticity in VL trajectories. Since we include individuals with both high and low information, direct parameter comparisons to the model presented in the SI of Zitzmann et al. [5] are not possible. However, we can compare our  $R_0$  estimates to their results and those from other SARS-CoV-2 studies [51] as well as other respiratory illnesses [31,51]. Additionally, we compare predicted VL trajectories to those of some of the SI fits in Zitzmann et al. [5]. Phenomenological models [4,11] estimate parameters based on piecewise linear models, which are not directly comparable to our mechanistic approach, but our model can estimate the same summary statistics as phenomenological models while also inferring biological quantities like  $R_0$ .

**2.8.1 Generating synthetic datasets.** We generated synthetic viral load (VL) datasets by sampling model parameters from the priors described in Sect 2.6, fixing the shared (hyper-)parameters and drawing individual-level parameters accordingly (e.g. for the basic reproduction numbers,  $R_{0,i} \sim \mathcal{N}(\mu_{R_0}, \sigma_{R_0})$  and similarly for  $\delta_i$  and  $\rho_i$ ). The shared parameter means ( $\mu_{R_0}, \mu_\delta, \mu_\rho$ ) were chosen to align with previous estimates [5, SI], while their corresponding scales ( $\sigma_{R_0}, \sigma_\delta, \sigma_\rho$ ) were selected heuristically based on comparisons between simulated and NBA datasets. The measurement noise parameter  $\kappa$  was set to 0.5, representing a moderate, “low-noise” scenario suitable for evaluating model performance under idealised conditions. The full set of hyper-parameter values is given in Table 3.

For each individual, infection times were drawn from the prior, and corresponding time series were simulated using Algorithm 1. Because variability in the priors can produce unrealistic trajectories (e.g., non-clearing infections or implausible peak VLs), we applied a rejection-sampling procedure to retain only biologically plausible cases. Specifically, we accepted simulations with peak log-VL between 5 and 10 (consistent with the NBA dataset), peak time between 4 and 9 days post-infection, and log-VL below 4.5 at 21 days post-infection to ensure viral clearance. As in the NBA data, a limit of detection (LOD) of 2.658 log copies/mL was imposed and processed as described in Sect 2.8. The resulting datasets thus represent an idealised setting in which individuals have well-resolved observations across both the growth and decline phases of infection.

### 3 Results

In this section, we apply our method to the datasets described in Sect 2.8. We first analyse the synthetic datasets to demonstrate the algorithm’s ability to recover accurate parameter estimates in a setting where the true values are known. We then apply the method to the NBA dataset to illustrate its utility in a real-world scenario where the true parameter values are unknown.

**Table 3. The values of the shared parameters used to generate the synthetic datasets.**

Parameter	Value
$\mu_{R_0}$	8
$\sigma_{R_0}$	0.5
$\mu_\delta$	1.3
$\sigma_\delta$	0.15
$\mu_\rho$	3
$\sigma_\rho$	0.25
$\kappa$	0.5

<https://doi.org/10.1371/journal.pcbi.1013775.t003>

All simulations and models were implemented in Julia v1.12 [42] and executed on a 2021 MacBook Pro with an M1 chip and 16 GB of RAM. Likelihood computations were parallelised across four high-performance CPU cores, as described in Sect 5 of S1 text. Pilot runs were used to tune the proposal distributions for both individual and shared parameters. Following the pilot runs, we initialised three chains from perturbed versions of the posterior means obtained in the pilot phase. For each simulated dataset, each chain was run for 100,000 iterations, with the first 20,000 discarded as burn-in. For the NBA dataset, each chain was run for 200,000 iterations, discarding the first 20,000 as burn-in. The runs for the NBA dataset (the larger of the two) required approximately 15 minutes per chain, giving a total runtime of about 45 minutes.

Trace plots for all shared parameters and a subset of individual parameters were visually inspected for divergences and showed no evidence of poor mixing across chains in either analysis. All parameters had  $\hat{R} \leq 1.01$ , indicating good within- and between-chain convergence [52], with minimum effective sample sizes exceeding 500. The acceptance rates were approximately 0.30 for both the individual (averaged across groups) and shared parameters, within the expected range for well-tuned random walk Metropolis proposals [53].

### 3.1 Simulation study

Table 4 reports numerical summaries of the posterior fits (see Sect 7 of S1 text for details on each of the scoring rules shown in Table 4). Fig 3 displays the 95% credible intervals for the population parameters across 50 of the 200 simulated datasets. In most cases, the true values are contained within the 95% credible intervals, with the lowest empirical coverage being 0.95. The credible intervals are narrow relative to the scale of the parameters (see the Interval width column in Table 4), indicating that the model recovers the population values with good precision. Relative bias was computed as the bias divided by the true parameter value, giving a signed percentage difference. For all parameters, zero lies within the 95% confidence interval for the relative bias, suggesting that the method is unbiased. For  $\sigma_{R_0}$  and  $\sigma_\rho$ , the average relative biases are  $-0.109$  and  $0.585$ , respectively. However, Fig 3 shows that the 95% credible intervals for these variance parameters are much wider (average interval widths of 1.05 and 0.826, respectively), reflecting substantial uncertainty. This explains why, despite noticeable bias in the posterior means of these two parameters, interval coverage remains adequate.

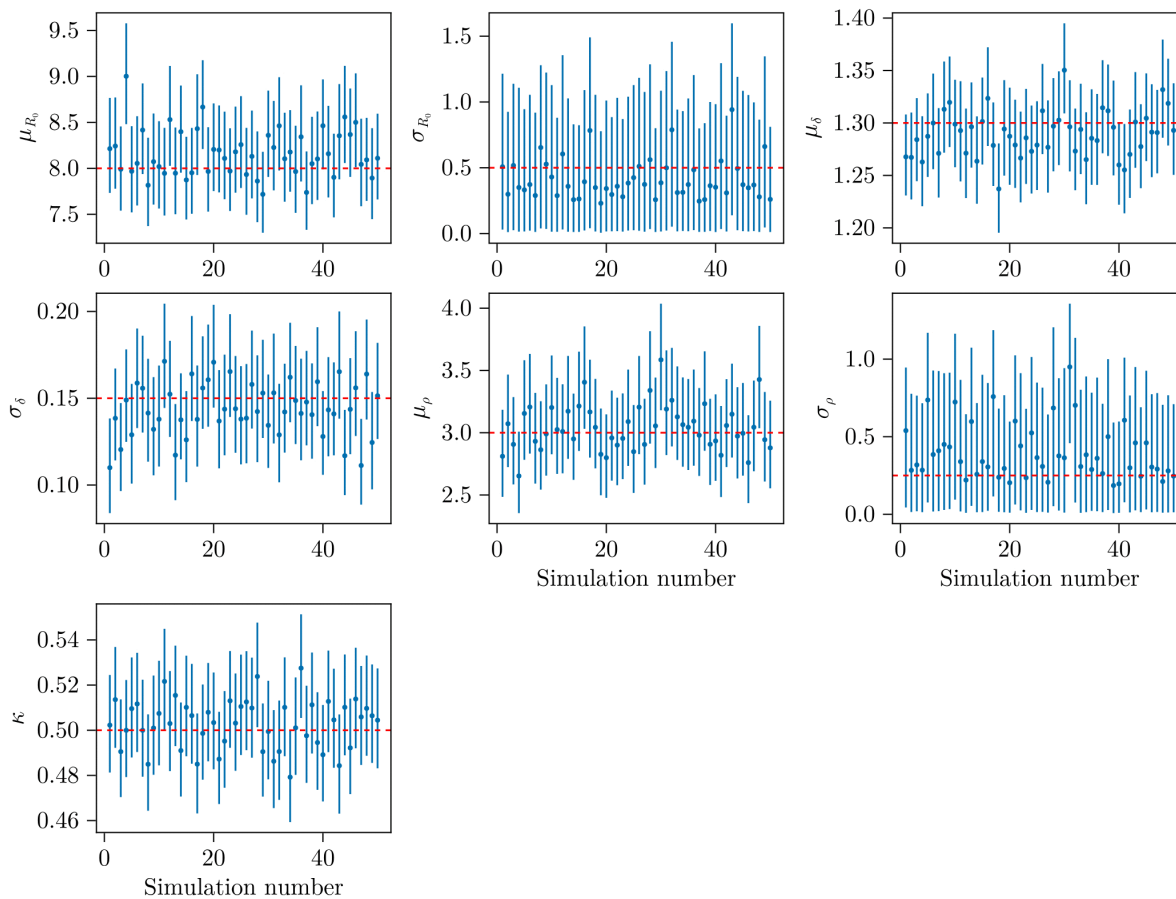
Fig 4 shows the posterior predictive fits for some example individuals from the synthetic dataset shown in Fig 2A. In all cases, the modelled trajectories track the observed data well, demonstrating that the method can reliably recover the viral dynamics in an ideal scenario.

In addition to the trajectories and estimated parameters, by its construction, our model can estimate the posterior predictive time-shift for individuals. Fig 5 compares samples from the the prior (orange) and posterior (blue) time shift distributions for a single individual. We see that the prior predictive distributions are quite wide and highly varied. Some of the density functions have large amounts of mass concentrated above 1-2 days and have long tails (in both directions).

**Table 4. Numerical summary of 200 posterior fits for the shared parameters.** Coverage is the proportion of simulations (out of 200) in which the true parameter value lies within the 95% credible interval. The average interval width is reported across simulations, with a 95% confidence interval. Relative bias is also reported with its corresponding 95% confidence interval. See Sect 7 of S1 text for further details of the scoring metrics.

Parameter	Coverage	Interval width	Relative bias
$\mu_{R_0}$	0.95	0.958 [0.862, 1.096]	0.016 [−0.032, 0.072]
$\sigma_{R_0}$	1.00	1.05 [0.808, 1.233]	−0.109 [−0.422, 0.525]
$\mu_\delta$	0.95	0.086 [0.078, 0.095]	−0.007 [−0.037, 0.025]
$\sigma_\delta$	0.95	0.056 [0.05, 0.063]	−0.021 [−0.215, 0.146]
$\mu_\rho$	0.97	0.745 [0.654, 0.858]	0.020 [−0.093, 0.134]
$\sigma_\rho$	0.98	0.826 [0.774, 1.443]	0.585 [−0.106, 2.067]
$\kappa$	0.96	0.044 [0.041, 0.046]	0.006 [−0.036, 0.044]

<https://doi.org/10.1371/journal.pcbi.1013775.t004>



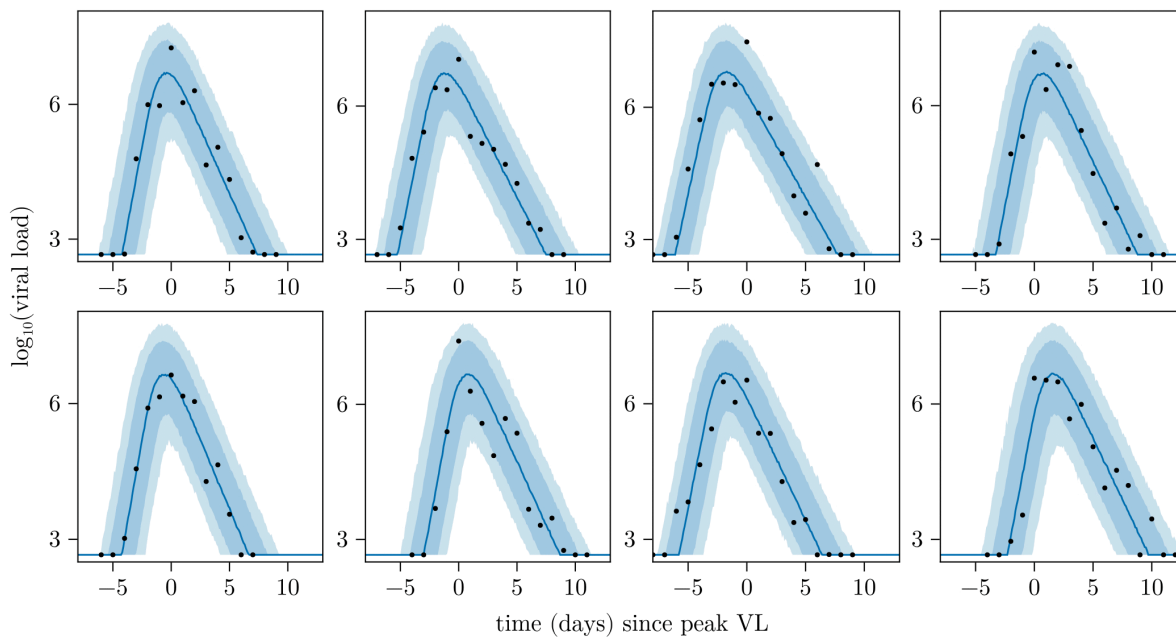
**Fig 3. 95% credible interval plots of the population parameters for 50 of the 200 simulated studies.** Blue lines indicate the 95% credible interval for a given simulation, with blue points indicating the median. The horizontal red dashed line shows the true value of the parameters used throughout the simulations.

<https://doi.org/10.1371/journal.pcbi.1013775.g003>

Contrasting this, the posterior predictive time-shift distributions (blue) are more concentrated, with the distributions roughly resembling one another besides the variation arising from the posterior sampled parameters themselves.

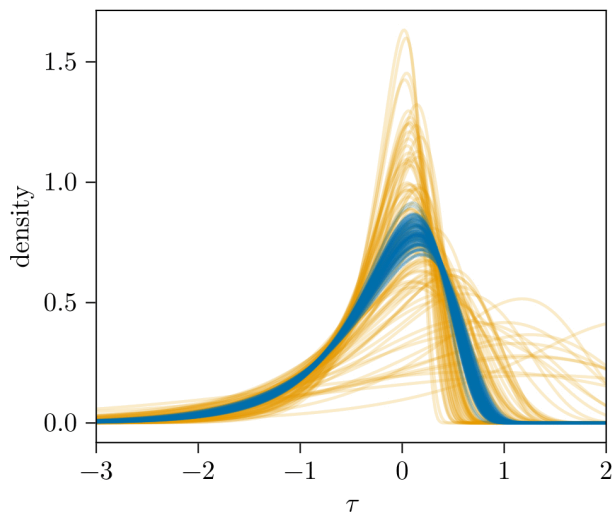
### 3.2 NBA study

The marginal posteriors for the shared parameters are shown in Fig 6. We see that there is large movement in the distributions from the prior (orange) to the posteriors (blue) for all parameters. The model appears to be able to estimate all parameters including the measurement noise  $\kappa$ . Posterior means and 95% credible intervals are provided in Table 5. Our estimates for  $\mu_{R_0}$  are consistent with the results in Table 2. It should be noted that compared to Zitzmann et al. [5, SI] our population estimate is larger, 14.64 compared with 8.2. Our  $R_0$  estimates are consistent with the value of 14.2 reported in Germano et al. [18] from their stochastic model (which also accounts for immune responses). Our population estimate for  $\mu_\rho$  is 3.19 which is slightly larger than, but still consistent with the estimate of 3.07 from [5, SI]. We infer the population mean rate of removal for infectious cells to be 1.15, slightly lower than 1.3 in [5, SI]. There are several possible reasons for the difference in the estimates of  $R_0$  (i.e.  $\mu_{R_0}$ ). The first being that, as our model is stochastic, larger values of the parameters are required to ensure that the model overcomes extinction than what would be estimated by a deterministic model. We are also fitting to a much larger number of individuals (163 vs 25) including those that are unvaccinated



**Fig 4. Posterior predictive fits for a subset of individuals in an example synthetic dataset.** Posterior predictive VL trajectories for a subset of individuals shown in Fig 2. Solid lines represent the median of the posterior predictive fits and the shaded regions represent the 80% and 95% prediction intervals. Black dots indicate observed data.

<https://doi.org/10.1371/journal.pcbi.1013775.g004>

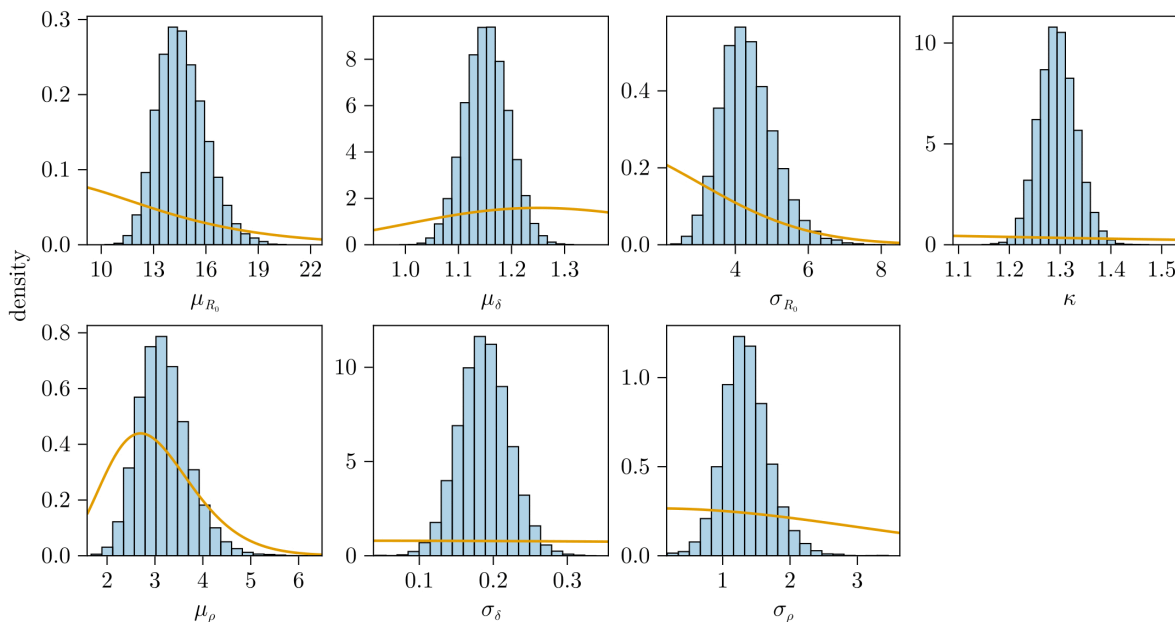


**Fig 5. Time-shift posterior predictive distributions.** Posterior predictive (blue) and prior predictive (orange) time-shifts for a single individual in the synthetic dataset.

<https://doi.org/10.1371/journal.pcbi.1013775.g005>

and have limited data during the growth and decline phases. Additionally, we account for the LOD in our model and make no assumptions as to the time from infection to peak instead allowing our model to infer the infection time.

Fig 7 shows the posterior predictive VL fits (blue) for a subset of individuals in the NBA dataset. The first row of Fig 7 corresponds to data for four individuals also analysed in the SI of Zitzmann et al. [5] alongside their estimated VL



**Fig 6. Histograms of the marginal posterior distributions for the shared parameters (hyper-parameters and the scale of the measurement noise) for the NBA model fit.** Prior distributions are indicated by orange lines and posteriors are shown as blue histograms.

<https://doi.org/10.1371/journal.pcbi.1013775.g006>

**Table 5. Posterior means and 95% credible intervals for the shared parameters.**

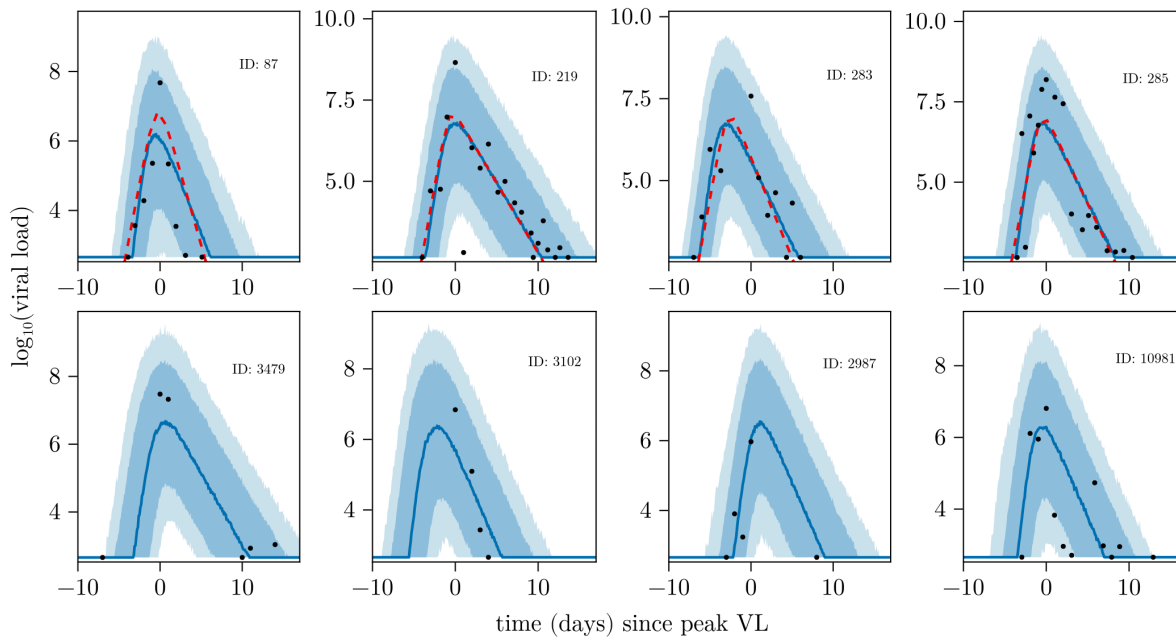
Parameter	Mean (95% CrI)
$\mu_{R_0}$	14.64 [12.23, 17.60]
$\sigma_{R_0}$	4.34 [3.07, 5.98]
$\mu_\delta$	1.15 [1.07, 1.23]
$\sigma_\delta$	0.19 [0.12, 0.26]
$\mu_\rho$	3.19 [2.28, 4.35]
$\sigma_\rho$	1.35 [0.71, 2.10]
$\kappa$	1.29 [1.23, 1.37]

<https://doi.org/10.1371/journal.pcbi.1013775.t005>

trajectories (red dashed line). We see strong agreement between our model and the model in the SI of Zitzmann et al. [5, SI] model, with our median predictions closely tracking their estimated trajectories. The second row of plots corresponds to individuals (visually) selected for having limited data in either the pre- or post-peak phase. The posterior predictive trajectories track the observed data well.

## 4 Discussion

We have presented a novel method for fitting mechanistic within-host viral dynamics models that explicitly account for process noise, providing a more realistic representation of viral load (VL) trajectories. Our approach leverages the concept of a random time-shift—first introduced in Barbour et al. [16] and later developed into a practical tool in Morris et al. [17]—to capture the effects of stochasticity in the WHVDs. Our key contribution in this work is demonstrating how the time-shift distributions can be directly incorporated into the likelihood function, allowing stochasticity to be modelled without the need for computationally expensive stochastic simulations, for example using Gillespie’s algorithm [22] or tau-leaping [23]. Furthermore, this enables efficient inference, making it feasible to fit the model to large datasets using a hierarchical framework. Hierarchical approaches have been widely used in phenomenological studies of within-host viral



**Fig 7. Posterior predictive fits for a subset of individuals from the NBA dataset.** The first row of plots corresponds to a subset of individuals also presented in the SI of Zitzmann et al. [5] for comparison purposes. The second row corresponds to individuals not studied in that work, but that have sparse data that were selected by visual inspection. The observations for each individual are shown in black. The solid lines represent the median of the posterior predictive fits and the shaded regions represent the 80% and 95% prediction intervals. The red dashed lines in the first row of plots indicate the deterministic VL trajectories estimated in [5]. All subplots show the same time period  $[-10, 17]$ .

<https://doi.org/10.1371/journal.pcbi.1013775.g007>

dynamics (WHVDs) [6,11], but to our knowledge, no stochastic WHVD models have incorporated hierarchical structures. Our approach can be viewed as a hybrid method, combining the interpretability and tractability of deterministic models while capturing key elements of stochastic modelling, offering a middle ground between fully stochastic simulation and purely deterministic fitting.

Existing approaches for fitting WHVDs fall into two main categories: simpler phenomenological models, such as piecewise linear functions [4,6,11], and mechanistic models that assume deterministic dynamics [5,12,25,40,44,45]. Phenomenological models offer computational advantages, as their likelihoods are easier to evaluate and they integrate well with probabilistic programming frameworks like Stan [54]. However, they primarily describe viral dynamics in terms of slope parameters (for growth and decline) and a hinge point (for peak VL), limiting their ability to infer biologically meaningful quantities such as the basic reproduction number  $R_0$  or production rate  $\rho$ . Mechanistic models, by contrast, provide a more realistic representation of infection dynamics but are computationally more demanding, as they require repeatedly solving ODEs or simulating stochastic processes. Such ODE-based models can be fit using dedicated mixed-effects tools such as Monolix [55], which implement efficient frequentist approaches for nonlinear hierarchical models but are less flexible than Bayesian frameworks for incorporating stochastic dynamics or propagating uncertainty in a principled way. Parameter identifiability also poses a major challenge, particularly when biological parameters influence viral load in overlapping ways (e.g. relationships among  $R_0$ , the viral production rate, infection timing, and initial conditions) [5,25]. Although some hybrid stochastic–deterministic models have been developed [18,56,57], they remain computationally intensive and have not been widely applied to large, multi-individual viral kinetics datasets. Our method is the first to use a stochastic mechanistic model within a hierarchical framework that scales efficiently to datasets with over a hundred individuals.

We have adopted the simplest viral dynamics model for this work due to parameter identifiability concerns, but the method can easily be applied to other suitable models. These models must satisfy the conditions required for applying the time-shift approximation [16,17], which many biological systems with large populations naturally fulfill [17,58,59]. The main constraint beyond the large population assumption lies in data availability during the pre-exponential growth phase. Systems where early-time dynamics are critical to macroscopic outcomes but where direct observations are limited, such as other within-host models, are particularly well suited for this approach. Examples include the innate response model used in Zitzmann et al. [5] and more complex models, such as those examining the role of MUC1 in reducing influenza severity [14]. Conversely, these conditions also define when our method may not be the most appropriate. In cases where high-quality early-growth-phase data is available, exact stochastic methods may be preferable. Epidemic outbreak data, for instance, typically includes detailed observations during the initial growth phase when cases are few and monitoring efforts are robust, making alternative approaches more suitable.

The primary limitation of our method is its complexity compared to simpler approaches in the literature that leverage standard packages like Stan [54]. Our framework integrates multiple techniques, neural networks, random time-shift computations, Laplace approximations, and numerical solvers, to fit the model. The neural network plays a critical role in enabling the time-shift approximation by learning a smooth mapping from model parameters to the parameters of the time-shift distribution, effectively circumventing the need for expensive optimisation routines during likelihood evaluation. This learned surrogate is then embedded directly within the likelihood, facilitating fast likelihood computations. This complexity means the method is not a black box and requires expertise to debug. If performance issues arise, such as MCMC convergence problems, identifying the source of the issue within the pipeline can be challenging.

Increased method complexity often comes with computational challenges, akin to the trade-offs between importance sampling in particle filters and bootstrap filtering methods [60,61]. To mitigate these issues, we have structured the method into several modular and easily testable steps. The two main approximations—using a neural network to learn time-shift distribution parameters and employing Laplace approximations—can both be replaced with exact computations (at the cost of efficiency) for validation. We recommend validating each component separately before applying the full framework. This includes ensuring that the neural network generalizes well beyond its training set, verifying the accuracy of time-shift distributions (e.g., by comparing against simulations from the stochastic model [17]), and using Laplace approximations only when justified (e.g., for normally distributed observation processes). Profile likelihoods can further confirm that the method accurately captures the shape of the exact likelihood, as demonstrated in Sect 4 of [S1 text](#).

Another implementation consideration when applying to new models is that the calculation of time-shift distributions are model-specific. Our previous work [17] provides a package for computing the time-shift distributions given the branching process results that can be used for this. Additionally, the neural network must be redesigned and retrained for any new VL kinetics model as the underlying branching process structure is altered. While these challenges are not insurmountable, they are important factors when deciding whether to use this approach. However, our work here provides a general template for implementation making it substantially easier to extend the method to other models going forward.

Simulation-based validation of viral kinetics models has received relatively little attention in the literature, with most models being fitted directly to data and rarely tested on synthetic datasets. This is likely due to the challenge of generating biologically plausible simulations that capture realistic viral dynamics and experimental designs. In this work, we provide a preliminary validation study using multiple synthetic datasets generated from a single biologically plausible parameter set—drawn from literature estimates and consistent with our real dataset—and show that the true population parameters can be reasonably recovered. Validation and robustness of an inference method has many other dimensions, including model misspecification, varying sample sizes (both within and across individuals), population variability, and measurement error. Exploring inference performance across these factors is an important direction for future work, which would further assess method reliability and contribute to broader validation of viral kinetics models.

A key advantage of our approach is its ability to model heterogeneity in viral trajectories, which is particularly relevant when initial infections involve a small number of cells. Beyond within-host modelling, incorporating stochasticity also

has implications for between-host transmission studies [62–65]. Small variations in VL trajectories can influence an individual's infectiousness, affecting transmission dynamics in settings like household studies [39,66,67]. A more realistic representation of within-host variability could lead to improved estimates of key epidemiological parameters, which is an area of ongoing research.

This paper introduces a methodological advancement for working with within-host viral dynamics (WHVDs). Rather than proposing a new model, we focus on developing computational tools that enable efficient Bayesian inference for existing stochastic VL kinetics models. Our interdisciplinary approach, combining applied mathematics, machine learning, and computational statistics, makes it feasible to fit these models to large datasets using biologically motivated mechanistic models. By leveraging modern computational techniques, we present a flexible framework that can accommodate the intricacies of viral dynamics and scale efficiently. A key enabler of this work is the use of the Julia programming language [42], which allows for seamless integration of model specification, automatic differentiation, and inference within a single environment. This capability makes it possible to conduct complex analyses on local hardware, reducing the reliance on dedicated high-performance computing clusters. Overall, the methods introduced here offer a powerful toolset for advancing the study of WHVDs and supporting future research into more realistic and scalable modelling frameworks.

## Supporting information

**S1 text. Supplementary text with additional methodological details.** This text includes the derivation of the deterministic approximation and the key branching process results underlying the time-shift distributions. We provide a full derivation of the Laplace approximation and compare it with the exact calculation using profile likelihoods (Fig 1 in S1 text). Further details are given on the model and inference method, including the MCMC routine, as well as the construction of the neural network used for amortised optimisation. (PDF)

## Author contributions

**Conceptualization:** Dylan J. Morris, Andrew J. Black.

**Formal analysis:** Dylan J. Morris.

**Methodology:** Dylan J. Morris, Lauren Kennedy, Andrew J. Black.

**Software:** Dylan J. Morris.

**Supervision:** Lauren Kennedy, Andrew J. Black.

**Validation:** Dylan J. Morris.

**Visualization:** Dylan J. Morris.

**Writing – original draft:** Dylan J. Morris, Lauren Kennedy, Andrew J. Black.

**Writing – review & editing:** Dylan J. Morris, Lauren Kennedy, Andrew J. Black.

## References

1. Perelson AS. Modelling viral and immune system dynamics. *Nat Rev Immunol*. 2002;2(1):28–36. <https://doi.org/10.1038/nri700> PMID: 11905835
2. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. Kinetics of influenza A virus infection in humans. *J Virol*. 2006;80(15):7590–9. <https://doi.org/10.1128/JVI.01623-05> PMID: 16840338
3. Canini L, Perelson AS. Viral kinetic modeling: State of the art. *J Pharmacokinetic Pharmacodyn*. 2014;41(5):431–43. <https://doi.org/10.1007/s10928-014-9363-3> PMID: 24961742
4. Kissler SM, Fauver JR, Mack C, Olesen SW, Tai C, Shiue KY, et al. Viral dynamics of acute SARS-CoV-2 infection and applications to diagnostic and public health strategies. *PLoS Biol*. 2021;19(7):e3001333. <https://doi.org/10.1371/journal.pbio.3001333> PMID: 34252080

5. Zitzmann C, Ke R, Ribeiro RM, Perelson AS. How robust are estimates of key parameters in standard viral dynamic models?. *PLoS Comput Biol*. 2024;20(4):e1011437. <https://doi.org/10.1371/journal.pcbi.1011437> PMID: 38626190
6. Kissler SM, Fauver JR, Mack C, Tai CG, Breban MI, Watkins AE, et al. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated persons. *N Engl J Med*. 2021;385(26):2489–91. <https://doi.org/10.1056/NEJMc2102507> PMID: 34941024
7. Serena L, Marzolla M, D'Angelo G, Ferretti S. A review of multilevel modeling and simulation for human mobility and behavior. *Simul Modell Pract Theory*. 2023;127:102780. <https://doi.org/10.1016/j.simpat.2023.102780>
8. Ke R, Martinez PP, Smith RL, Gibson LL, Mirza A, Conte M, et al. Daily longitudinal sampling of SARS-CoV-2 infection reveals substantial heterogeneity in infectiousness. *Nat Microbiol*. 2022;7(5):640–52. <https://doi.org/10.1038/s41564-022-01105-z> PMID: 35484231
9. Néant N, Lingas G, Le Hingrat Q, Ghosn J, Engelmann I, Lepiller Q, et al. 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*. 2021;118(8):e2017962118. <https://doi.org/10.1073/pnas.2017962118> PMID: 33536313
10. Ashcroft P, Lehtinen S, Bonhoeffer S. Test-trace-isolate-quarantine (TTIQ) intervention strategies after symptomatic COVID-19 case identification. *PLoS One*. 2022;17(2):e0263597. <https://doi.org/10.1371/journal.pone.0263597> PMID: 35148359
11. Challenger JD, Foo CY, Wu Y, Yan AWC, Marjaneh MM, Liew F, et al. Modelling upper respiratory viral load dynamics of SARS-CoV-2. *BMC Med*. 2022;20(1):25. <https://doi.org/10.1186/s12916-021-02220-0> PMID: 35022051
12. Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. *Proc Natl Acad Sci U S A*. 2021;118(49):e2111477118. <https://doi.org/10.1073/pnas.2111477118> PMID: 34857628
13. Ciupe SM, Heffernan JM. In-Host Modeling. *Infectious Disease Modelling*. 2017;2:188–202. doi:10.1016/j.idm.2017.04.002.
14. Li K, Cao P, McCaw JM. Modelling the effect of MUC1 on influenza virus infection kinetics and macrophage dynamics. *Viruses*. 2021;13(5):850. <https://doi.org/10.3390/v13050850> PMID: 34066999
15. Bai F, Allen LJS. Probability of a major infection in a stochastic within-host model with multiple stages. *Appl Math Lett*. 2019;87:1–6. <https://doi.org/10.1016/j.aml.2018.07.022>
16. Barbour AD, Hamza K, Kaspi H, Klebaner FC. Escape from the boundary in Markov population processes. *Adv Appl Probabil*. 2015;47(4):1190–211. <https://doi.org/10.1239/aap/1449859806>
17. Morris D, Maclean J, Black AJ. Computation of random time-shift distributions for stochastic population models. *J Math Biol*. 2024;89(3). <https://doi.org/10.1007/s00285-024-02132-6>
18. Germano DPJ, Zarebski AE, Hautphenne S, Moss R, Flegg JA, Flegg MB. A hybrid framework for compartmental models enabling simulation-based inference; 2024.
19. Saenz RA, Quinlivan M, Elton D, Macrae S, Blunden AS, Mumford JA, et al. Dynamics of influenza virus infection and pathology. *J Virol*. 2010;84(8):3974–83. <https://doi.org/10.1128/JVI.02078-09> PMID: 20130053
20. Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science*. 2021;373(6552):eabh0635. <https://doi.org/10.1126/science.abh0635> PMID: 34083451
21. Odaka M, Inoue K. Modeling viral dynamics in SARS-CoV-2 infection based on differential equations and numerical analysis. *Heliyon*. 2021;7(10):e08207. <https://doi.org/10.1016/j.heliyon.2021.e08207> PMID: 34693064
22. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. 1977;81(25):2340–61. <https://doi.org/10.1021/j100540a008>
23. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys*. 2001;115(4):1716–33. <https://doi.org/10.1063/1.1378322>
24. Canini L, Carrat F. Population modeling of influenza A/H1N1 virus kinetics and symptom dynamics. *J Virol*. 2011;85(6):2764–70. <https://doi.org/10.1128/JVI.01318-10> PMID: 21191031
25. Ciupe SM, Tuncer N. Identifiability of parameters in mathematical models of SARS-CoV-2 infections in humans. *Sci Rep*. 2022;12(1):14637. <https://doi.org/10.1038/s41598-022-18683-x> PMID: 36030320
26. Nguyen VK, Klawonn F, Mikolajczyk R, Hernandez-Vargas EA. Analysis of practical identifiability of a viral infection model. *PLoS One*. 2016;11(12):e0167568. <https://doi.org/10.1371/journal.pone.0167568> PMID: 28036339
27. Yan AWC, Cao P, McCaw JM. On the extinction probability in models of within-host infection: The role of latency and immunity. *J Math Biol*. 2016;73(4):787–813. <https://doi.org/10.1007/s00285-015-0961-5> PMID: 26748917
28. Ross SM. *Introduction to Probability Models*. Academic Press; 2014.
29. Durrett R. *Essentials of stochastic processes*. Springer texts in statistics. Cham: Springer International Publishing; 2016.
30. Begon M, Bennett M, Bowers RG, French NP, Hazel SM, Turner J. A clarification of transmission terms in host-microparasite models: Numbers, densities and areas. *Epidemiol Infect*. 2002;129(1):147–53. <https://doi.org/10.1017/s0950268802007148> PMID: 12211582
31. Smith AP, Moquin DJ, Bernhauerova V, Smith AM. Influenza virus infection model with density dependence supports biphasic viral decay. *Front Microbiol*. 2018;9.
32. Diekmann O, Heesterbeek JA, Metz JA. On the definition and the computation of the basic reproduction ratio  $R_0$  in models for infectious diseases in heterogeneous populations. *J Math Biol*. 1990;28(4):365–82. <https://doi.org/10.1007/BF00178324> PMID: 2117040
33. van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci*. 2002;180:29–48. [https://doi.org/10.1016/s0025-5564\(02\)00108-6](https://doi.org/10.1016/s0025-5564(02)00108-6) PMID: 12387915

34. Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface*. 2010;7(47):873–85. <https://doi.org/10.1098/rsif.2009.0386> PMID: 19892718
35. Curran-Sebastian J, Pellis L, Hall I, House T. Calculation of epidemic first passage and peak time probability distributions. *SIAM/ASA J Uncertain Quantif*. 2024;12(2):242–61. <https://doi.org/10.1137/23m1548049>
36. Amos B. Tutorial on amortized optimization. *FNT Mach Learn*. 2023;16(5):592–732. <https://doi.org/10.1561/2200000102>
37. Marino J, Cvitkovic M, Yue Y. A general method for amortizing variational filtering. In: *Advances in neural information processing systems*. vol. 31. Curran Associates, Inc.; 2018.
38. Gershman S, Goodman N. Amortized inference in probabilistic reasoning. In: *Proceedings of the annual meeting of the cognitive science society*; 2014.
39. Marcató AJ, Black AJ, Walker CR, Morris D, Meagher N, Price DJ, et al. Learnings from the Australian first few X household transmission project for COVID-19. *Lancet Reg Health West Pac*. 2022;28:100573. <https://doi.org/10.1016/j.lanwpc.2022.100573> PMID: 36089928
40. Li K, McCaw JM, Cao P. Enhanced viral infectivity and reduced interferon production are associated with high pathogenicity for influenza viruses. *PLoS Comput Biol*. 2023;19(2):e1010886. <https://doi.org/10.1371/journal.pcbi.1010886> PMID: 36758109
41. MacKay DJ. *Information theory, inference and learning algorithms*. Cambridge University Press; 2003.
42. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A fresh approach to numerical computing. *SIAM Rev*. 2017;59(1):65–98. <https://doi.org/10.1137/141000671>
43. Revels J, Lubin M, Papamarkou T. *Forward-mode automatic differentiation in Julia*; 2016.
44. Hernandez-Vargas EA, Velasco-Hernandez JX. In-host mathematical modelling of COVID-19 in humans. *Annu Rev Control*. 2020;50:448–56. <https://doi.org/10.1016/j.arcontrol.2020.09.006> PMID: 33020692
45. Ejima K, Kim KS, Ludema C, Bento AI, Iwanami S, Fujita Y, et al. Estimation of the incubation period of COVID-19 using viral load data. *Epidemics*. 2021;35:100454. <https://doi.org/10.1016/j.epidem.2021.100454> PMID: 33773195
46. Sunagawa J, Park H, Kim KS, Komorizono R, Choi S, Ramirez Torres L, et al. Isolation may select for earlier and higher peak viral load but shorter duration in SARS-CoV-2 evolution. *Nat Commun*. 2023;14(1):7395. <https://doi.org/10.1038/s41467-023-43043-2> PMID: 37989736
47. Robert CP, Casella G. *Monte Carlo statistical methods*. 2nd ed. New York: Springer; 2004.
48. Rackauckas C, Nie Q. *DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia*. *JORS*. 2017;5(1):15. <https://doi.org/10.5334/jors.151>
49. Innes M, Saba E, Fischer K, Gandhi D, Rudilosso MC, Joy NM. Fashionable modelling with flux. *CoRR*. 2018. <https://doi.org/abs/1811.01457>
50. Gelman A. *Bayesian data analysis*. City: CRC Press; 2013.
51. McCormack CP, Yan AWC, Brown JC, Sukhova K, Peacock TP, Barclay WS, et al. Modelling the viral dynamics of the SARS-CoV-2 delta and omicron variants in different cell types. *J R Soc Interface*. 2023;20(205):20230187. <https://doi.org/10.1098/rsif.2023.0187> PMID: 37553993
52. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-normalization, folding, and localization: An improved  $\hat{R}$  for assessing convergence of MCMC (with discussion). *Bayesian Anal*. 2021;16(2). <https://doi.org/10.1214/20-ba1221>
53. Bédard M. Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process Applic*. 2008;118(12):2198–222. <https://doi.org/10.1016/j.spa.2007.12.005>
54. Stan Development Team. *Stan: A probabilistic programming language*; 2024. <https://mc-stan.org>
55. Simulations PL, Plus. *Monolix 2024R1*; 2024.
56. Rebuli NP, Bean NG, Ross JV. Hybrid Markov chain models of S-I-R disease dynamics. *J Math Biol*. 2017;75(3):521–41. <https://doi.org/10.1007/s00285-016-1085-2> PMID: 28013336
57. Kreger J, Komarova NL, Wodarz D. A hybrid stochastic-deterministic approach to explore multiple infection and evolution in HIV. *PLoS Comput Biol*. 2021;17(12):e1009713. <https://doi.org/10.1371/journal.pcbi.1009713> PMID: 34936647
58. Jäger W, Rost H, Tautu P. *Density dependent Markov population processes*. Biological growth and spread. Berlin, Heidelberg: Springer; 1980. p. 36–49.
59. Allen LJS. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infect Dis Model*. 2017;2(2):128–42. <https://doi.org/10.1016/j.idm.2017.03.001> PMID: 29928733
60. Black AJ. Importance sampling for partially observed temporal epidemic models. *Stat Comput*. 2018;29(4):617–30. <https://doi.org/10.1007/s11222-018-9827-1>
61. McKinley TJ, Ross JV, Deardon R, Cook AR. Simulation-based Bayesian inference for epidemic models. *Computat Stat Data Anal*. 2014;71:434–47. <https://doi.org/10.1016/j.csda.2012.12.012>
62. Wang X, Wang S, Wang J, Rong L. A Multiscale model of COVID-19 dynamics. *Bull Math Biol*. 2022;84(9):99. <https://doi.org/10.1007/s11538-022-01058-8> PMID: 35943625
63. Goyal A, Reeves DB, Schiffer JT. Multi-scale modelling reveals that early super-spreader events are a likely contributor to novel variant predominance. *J R Soc Interface*. 2022;19(189):20210811. <https://doi.org/10.1098/rsif.2021.0811> PMID: 35382576
64. Almcocera AES, Nguyen VK, Hernandez-Vargas EA. Multiscale model within-host and between-host for viral infectious diseases. *J Math Biol*. 2018;77(4):1035–57. <https://doi.org/10.1007/s00285-018-1241-y> PMID: 29737396

65. Marion G, Hadley L, Isham V, Mollison D, Panovska-Griffiths J, Pellis L, et al. Modelling: Understanding pandemics and how to control them. *Epidemics*. 2022;39:100588. <https://doi.org/10.1016/j.epidem.2022.100588> PMID: 35679714
66. Black AJ, Geard N, McCaw JM, McVernon J, Ross JV. Characterising pandemic severity and transmissibility from data collected during first few hundred studies. *Epidemics*. 2017;19:61–73. <https://doi.org/10.1016/j.epidem.2017.01.004> PMID: 28189386
67. House T, Riley H, Pellis L, Pouwels KB, Bacon S, Eidukas A, et al. Inferring risks of coronavirus transmission from community household data. *Stat Methods Med Res*. 2022;31(9):1738–56. <https://doi.org/10.1177/09622802211055853> PMID: 36112916