Supplementary Material

S1 The K&M model

The Kermack and McKendrick (K&M) model [1] is an integro-differential equation (IDE) epidemic model, in which the infectiousness of each infected host varies continuously in time. We considered a form of the K&M model given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S(t)\left[\int_0^t \beta(\tau)\frac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}\mathrm{d}\tau - \int_t^\infty \beta(\tau)g(\tau-t)\mathrm{d}\tau\right], \tag{S1.1}$$

where $S(t)$ is the number of susceptible individuals at time t days since the start of the outbreak, $\beta(\tau)$ day^{-1} is the expected transmission rate of an individual who has been infected for time τ days (the expected infectiousness curve), and $g(\tau)$ represents the initial density of individuals who have been infected for time τ days. The basic reproduction number of this model is [2]

$$R_0 = N\int_0^\infty \beta(\tau)\mathrm{d}\tau, \tag{S1.2}$$

where N is the (constant) total population size. Latent periods, as well as any distribution of latent and infectious periods between hosts, can be incorporated into the expected infectiousness curve, $\beta(\tau)$.

If there is initially a single newly infected individual, with the remainder of the population susceptible, then $g(\tau) = \delta(\tau)$, where $\delta(\tau)$ is the Dirac delta function. In this case, the K&M model reduces to

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S(t)\left[\int_0^t \beta(\tau)\frac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}\mathrm{d}\tau - \beta(t)\right]. \tag{S1.3}$$

The SI$_n$R model (see Methods in the main text) is in fact obtained as a special case of the K&M model [3, 4]. The corresponding infectiousness curve, $\beta(\tau)$, is the expected transmission rate of an individual at time τ days since infection in the analogous stochastic SI$_n$R model. In this case, the expected infectiousness curve is non-zero for all times since infection $\tau > 0$, even if latent compartments are explicitly included in the SI$_n$R model. This is because the time that a given patient spends in each infected compartment is exponentially

distributed, and so may be arbitrarily short or long.

Numerical solution

To solve the K&M model numerically in the main text (figure $2b$) and in Section S6 of the Supplementary Material, we used a forward Euler method. In particular, we chose a time-step δt, and approximated the solution of the K&M model (in the form given by equation (S1.3)) using the finite difference scheme

$$S_{i+1} = S_i \left[1 + \delta t \left(\sum_{j=1}^{i} \beta_j (S_{i+1-j} - S_{i-j}) - \beta_i \right) \right], \tag{S1.4}$$

where $t_i = i\delta t, S_i = S(t_i), \beta_i = \beta(t_i)$. The time-step was chosen to be sufficiently small to ensure that the error in the numerical solution was negligible compared to other sources of error considered.

S2 Proof of equivalence between the compartmental and IDE methods

In this section, we first explain how the parameter values in the $\mathrm{SI}_n\mathrm{R}$ model can be chosen in order to approximate the population-scale dynamics if the expected infectiousness curve, $\beta(\tau)$, is known. We then prove that, for these parameter choices, the K&M IDE model is obtained in the limit as the number of infected compartments, $n \to \infty$.

Choosing parameter values in the $\mathrm{SI}_n\mathrm{R}$ model

We assume that the rates of progression through the infected compartments in the $\mathrm{SI}_n\mathrm{R}$ model, μ_i, all take the same value (where this value will depend on the number of compartments), so that we can write

$$\mu_1 = \mu_2 = \cdots = \mu_n = 1/\lambda(n), \tag{S2.1}$$

where $\lambda(n)$ is the mean time that a given host spends in each infected compartment (for a given number of compartments, n). We will choose $\lambda(n)$ so that $\lambda(n) \to 0$ as $n \to \infty$, in order to ensure that the infectiousness of each host varies continuously in the limit $n \to \infty$ (an explicit choice for $\lambda(n)$ is given below).

To choose the transmission rates, β_i, we consider an individual who spends exactly the mean time, $\lambda(n)$, in each infected compartment. At time τ since infection, such an individual will have infectiousness given by

$$\beta(\tau; n) = \begin{cases} \beta_1, & \text{for } 0 \leq \tau < \lambda(n), \\ \beta_2, & \text{for } \lambda(n) \leq \tau < 2\lambda(n), \\ \vdots & \\ \beta_n, & \text{for } (n-1)\lambda(n) \leq \tau < n\lambda(n), \\ 0, & \text{for } \tau \geq n\lambda(n). \end{cases} \tag{S2.2}$$

We wish to choose the β_i so that $\beta(\tau; n) \to \beta(\tau)$ as $n \to \infty$ with $\lambda(n) \to 0$. This will ensure that if a host spends exactly the mean time in each infected compartment, then their infectiousness will tend to $\beta(\tau)$ at every time since infection, τ, as $n \to \infty$. Therefore, since infectiousness also varies continuously when $n \to \infty$, each infected host will have an individual infectiousness curve given exactly by $\beta(\tau)$ in the limit. In general, $\beta(\tau)$ may have unbounded support, so we will also need to choose $\lambda(n)$ so that $n\lambda(n) \to \infty$ as $n \to \infty$. However, in practice we may expect there to be some T such that $\beta(\tau) = 0$ for $\tau > T$, in which case we may fix $n\lambda(n) = T$ (i.e., we can take $\lambda(n) = T/n$).

In particular, we choose the transmission rates to be

$$\beta_i = \frac{1}{\lambda(n)} \int_{(i-1)\lambda(n)}^{i\lambda(n)} \beta(\tau) \mathrm{d}\tau, \qquad i = 1, \ldots, n-1, \tag{S2.3}$$

$$\beta_n = \frac{1}{\lambda(n)} \int_{(n-1)\lambda(n)}^{\infty} \beta(\tau) \mathrm{d}\tau, \tag{S2.4}$$

so that β_i is the average value of $\beta(\tau)$ between times $(i-1)\lambda(n)$ and $i\lambda(n)$ since infection, and β_n is the average of $\beta(\tau)$ over times since infection greater than $(n-1)\lambda(n)$. These choices ensure that, for each n, the $\mathrm{SI}_n\mathrm{R}$ model has the same basic reproduction number as the K&M model with infectiousness curve $\beta(\tau)$.

The integro-differential equation and compartmental methods are equivalent in the limit $n \to \infty$

To show that the K&M model is obtained in the limit of infinitely many compartments when the SI$_n$R model is parameterised as outlined above, we first define an approximation to the density of infected individuals who have been infected for time τ in the SI$_n$R model,

$$I(t,\tau;n) \coloneqq \begin{cases} I_1/\lambda(n), & \text{for } 0 \le \tau < \lambda(n), \\ I_2/\lambda(n), & \text{for } \lambda(n) \le \tau < 2\lambda(n), \\ \vdots & \\ I_n/\lambda(n), & \text{for } (n-1)\lambda(n) \le \tau < n\lambda(n). \end{cases} \tag{S2.5}$$

We can rewrite the SI$_n$R equations in terms of $\beta(\tau;n)$ and $I(t,\tau;n)$ as

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S\int_0^{n\lambda(n)} \beta(\tau;n) I(t,\tau;n)\mathrm{d}\tau, \tag{S2.6}$$

$$\frac{\partial I(t,\tau;n)}{\partial t} = \begin{cases} -\dfrac{1}{\lambda(n)}\left(\dfrac{\mathrm{d}S}{\mathrm{d}t} + I(t,\tau;n)\right), & \text{for } 0 \le \tau < \lambda(n), \\ \dfrac{1}{\lambda(n)}\left(I\left(t,\tau-\lambda(n);n\right) - I(t,\tau;n)\right), & \text{for } \lambda(n) \le \tau < n\lambda(n). \end{cases} \tag{S2.7}$$

In the limit where $n \to \infty$, with $\lambda(n) \to 0$ (also taking $n\lambda(n) \to \infty$ if $\beta(\tau)$ has unbounded support), we find that $S(t)$ and $I(t,\tau) = I(t,\tau;\infty)$ satisfy

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S(t)\int_0^{\infty} \beta(\tau) I(t,\tau)\mathrm{d}\tau, \tag{S2.8}$$

$$\frac{\partial I(t,\tau)}{\partial t} + \frac{\partial I(t,\tau)}{\partial \tau} = 0, \quad \tau > 0, \tag{S2.9}$$

$$I(t,0) = -\frac{\mathrm{d}S}{\mathrm{d}t}. \tag{S2.10}$$

The PDE (S2.9), subject to boundary condition (S2.10) and a general initial condition $I(0,\tau) = g(\tau)$, has solution given by

$$I(t,\tau) = \begin{cases} g(\tau - t), & \text{for } t < \tau, \\ -\dfrac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}, & \text{for } t > \tau. \end{cases} \tag{S2.11}$$

Substituting this solution into equation (S2.8), we find that

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S(t)\left[\int_0^t \beta(\tau)\frac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}\mathrm{d}\tau - \int_t^\infty \beta(\tau)g(\tau-t)\mathrm{d}\tau\right], \tag{S2.12}$$

recovering the K&M model.

Initial conditions

Whenever we solved the $\mathrm{SI}_n\mathrm{R}$ model numerically starting with a single newly infected host (e.g. the blue line in figure 2*b* of the main text), we took the initial conditions $I_1(0) = 1$, $I_i(0) = 0$ for $i = 2, \ldots, n$, and $S(0) = N - 1$ (where N is the total population size). This choice of initial conditions gives

$$I(0,\tau;n) \coloneqq \begin{cases} \dfrac{1}{\lambda(n)}, & \text{for } 0 \leq \tau < \lambda(n), \\ 0, & \text{otherwise.} \end{cases} \tag{S2.13}$$

Taking the limit $n \to \infty$ with $\lambda(n) \to 0$, we find that $g(\tau) = I(0,\tau) = 0$ for $\tau > 0$, with $\int_0^\infty I(0,\tau)\mathrm{d}\tau = 1$. Therefore, $g(\tau) = \delta(\tau)$, where $\delta(\tau)$ is the Dirac delta function.

Note on the computational efficiency of the two methods

We consider the case where $\beta(\tau) = 0$ for $\tau > T$, and suppose that the $\mathrm{SI}_n\mathrm{R}$ model with parameter choices as in equations (S2.1) and (S2.3)–(S2.4) is discretised using an Euler method with a time-step $\delta t = \lambda(n) = T/n$. It is straightforward to see that this numerical scheme is essentially equivalent to the finite difference scheme given for the numerical solution of the K&M model in equation (S1.4), up to slight differences in the transmission rates. This therefore suggests there is no innate difference in computational efficiency between the compartmental and IDE approaches (rather, this will depend on the precise numerical methods used to solve each model). The major advantage of our compartmental approach is its ease of use, while remaining similarly computationally efficient to the IDE approach.

S3 Explicitly including variability between hosts in the population-scale model

In this section we consider a $\mathrm{SI}_{m,n}\mathrm{R}$ compartmental model, which explicitly incorporates variability in the time-course of infection between different hosts. We show that the K&M model is also obtained in the limit of infinitely many infected compartments when the $\mathrm{SI}_{m,n}\mathrm{R}$ model is used to predict population-scale dynamics.

The $\mathrm{SI}_{m,n}\mathrm{R}$ model

We consider a multi-stage compartmental model in which there are n possible "types" of infection, which are acquired with probabilities p_j, $j = 1 \dots n$. An infected individual with infection type j progresses through m infected compartments, $I_{i,j}$, $i = 1, \dots, m$. The model is given by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S \sum_{k=1}^{m} \sum_{l=1}^{n} \beta_{k,l} I_{k,l}, \tag{S3.1}$$

$$\frac{\mathrm{d}I_{1,j}}{\mathrm{d}t} = p_j S \sum_{k=1}^{m} \sum_{l=1}^{n} \beta_{k,l} I_{k,l} - \mu_{1,j} I_{1,j}, \tag{S3.2}$$

$$\frac{\mathrm{d}I_{i,j}}{\mathrm{d}t} = \mu_{i-1,j} I_{i-1,j} - \mu_{i,j} I_{i,j}, \quad i = 2, \dots, m, \tag{S3.3}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \sum_{l=1}^{n} \mu_{m,l} I_{m,l} \tag{S3.4}$$

where $\beta_{i,j} I_{i,j} S$ is the total rate at which individuals in class $I_{i,j}$ infect susceptibles, and $\mu_{i,j} I_{i,j}$ is the rate at which such hosts progress to the next infected compartment (or recover, if $i = m$).

In the limit of a continuum of possible infection types, denoted by the real variable y, we obtain the equations

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S \sum_{k=1}^{m} \int_{-\infty}^{\infty} \beta_k(y) I_k(t, y) \mathrm{d}y, \tag{S3.5}$$

$$\frac{\partial I_1(t, y)}{\partial t} = f(y) S \sum_{k=1}^{m} \int_{-\infty}^{\infty} \beta_k(y) I_k(t, y) \mathrm{d}y - \mu_1(y) I_1(t, y), \tag{S3.6}$$

$$\frac{\partial I_i(t,y)}{\partial t} = \mu_{i-1}(y)I_{i-1}(t,y) - \mu_i(y)I_i(t,y), \quad i = 2,\ldots,m, \tag{S3.7}$$

where $f(y)$ is the probability density of acquiring a type y infection, $I_i(t,y)$ is the density of individuals with a type y infection and in the i^{th} infected stage, $\beta_i(y)$ is the transmission rate of such individuals, and $\mu_i(y)$ is the rate of transition into stage $(i+1)$. This easily generalises to a multi-parameter distribution of infection types.

Approximating population-scale dynamics in the $\text{SI}_{m,n}\text{R}$ framework when there are multiple infection types

Suppose now that there are n possible infection types, acquired with probabilities p_j, and that each host with a type j infection follows a known infectiousness curve, $\beta_j(\tau)$. In this case, we can approximate the population-scale dynamics within the $\text{SI}_{m,n}\text{R}$ framework by taking

$$\mu_{ij} = 1/\lambda(m), \tag{S3.8}$$

$$\beta_{ij} = \frac{1}{\lambda(m)} \int_{(i-1)\lambda(m)}^{i\lambda(m)} \beta_j(\tau)\mathrm{d}\tau, \qquad i = 1,\ldots,m-1, \tag{S3.9}$$

$$\beta_{mj} = \frac{1}{\lambda(m)} \int_{(m-1)\lambda(m)}^{\infty} \beta_j(\tau)\mathrm{d}\tau, \tag{S3.10}$$

where we will again take $\lambda(m) \to 0$ as the number of compartments, $m \to \infty$.

If we have a continuum of infection types, denoted by y, with corresponding infectiousness curves $\beta(\tau, y)$, then we take

$$\mu_i(y) = 1/\lambda(m), \tag{S3.11}$$

$$\beta_i(y) = \frac{1}{\lambda(m)} \int_{(i-1)\lambda(m)}^{i\lambda(m)} \beta(\tau,y)\mathrm{d}\tau, \qquad i = 1,\ldots,m-1, \tag{S3.12}$$

$$\beta_m(y) = \frac{1}{\lambda(m)} \int_{(m-1)\lambda(m)}^{\infty} \beta(\tau,y)\mathrm{d}\tau, \tag{S3.13}$$

in equations (S3.5)–(S3.7).

The limit of infinitely many compartments

Now, suppose there are n possible infectiousness curves, and we parameterise the $\mathrm{SI}_{m,n}\mathrm{R}$ model as outlined above. Analogously to our definitions in Section S2, we define $\beta_j(\tau;m)$ as the piecewise constant approximation to the j^{th} infectiousness curve, and $I_j(t,\tau;m)$ as the approximate density of individuals with a type j infection who have been infected for time τ, so that the $\mathrm{SI}_{m,n}\mathrm{R}$ model can be written in the form

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S\sum_{l=1}^{n}\int_0^{m\lambda(m)} \beta_l(\tau;m)I_l(t,\tau;m)\mathrm{d}\tau, \tag{S3.14}$$

$$\frac{\partial I_j(t,\tau;n)}{\partial t} = \begin{cases} -\dfrac{1}{\lambda(m)}\left(p_j\dfrac{\mathrm{d}S}{\mathrm{d}t} + I_j(t,\tau;m)\right), & \text{for } 0\le\tau<\lambda(m), \\ \dfrac{1}{\lambda(m)}\left(I_j(t,\tau-\lambda(m);m) - I_j(t,\tau;m)\right), & \text{for } \lambda(m)\le\tau<m\lambda(m). \end{cases} \tag{S3.15}$$

In the limit $m\to\infty$, we obtain

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -S\sum_{l=1}^{n}\int_0^{\infty} \beta_l(\tau)I_l(t,\tau)\mathrm{d}\tau, \tag{S3.16}$$

$$\frac{\partial I_j(t,\tau)}{\partial t} + \frac{\partial I_j(t,\tau)}{\partial \tau} = 0, \quad \tau>0, \tag{S3.17}$$

$$I_j(t,0) = -p_j\frac{\mathrm{d}S}{\mathrm{d}t}. \tag{S3.18}$$

The PDE (S3.17), subject to boundary condition (S3.18) general initial condition $I_j(0,\tau) = g_j(\tau)$, has solution

$$I_j(t,\tau) = \begin{cases} g_j(\tau-t), & \text{for } t<\tau, \\ -p_j\dfrac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}, & \text{for } t>\tau. \end{cases} \tag{S3.19}$$

We then find that

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S\sum_{j=1}^{n}\left[\int_0^t p_j\beta_j(\tau)\frac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}\mathrm{d}\tau - \int_t^{\infty}\beta_j(\tau)g_j(\tau-t)\mathrm{d}\tau\right]. \tag{S3.20}$$

If we now define the expected infectiousness of an individual who has been infected for time τ to be

$$\beta(\tau) \coloneqq \sum_{j=1}^{n} p_j\beta_j(\tau), \tag{S3.21}$$

then we have that

$$\frac{\mathrm{d}S}{\mathrm{d}t} = S\left[\int_0^t \beta(\tau)\frac{\mathrm{d}S(t-\tau)}{\mathrm{d}t}\mathrm{d}\tau - \sum_{j=1}^{n}\int_t^\infty \beta_j(\tau)g_j(\tau-t)\mathrm{d}\tau\right]. \tag{S3.22}$$

Assuming that the term generated by the initial conditions has little effect once the outbreak has taken off, then this is simply the standard K&M model.

Therefore, the population-scale dynamics may be calculated by assuming that all infected individuals follow infectiousness curve $\beta(\tau)$. In the limit of a continuum of possible infection types, this analysis still holds, and equation (S3.21) becomes

$$\beta(\tau) = \int_{-\infty}^{\infty} f(y)\beta(\tau, y)\mathrm{d}y. \tag{S3.23}$$

This may be generalised further to multi-parameter distributions of infection types. In our example where the patient-level dynamics are described by the TCL model, the variation of the within-host parameters δ and $V(0)$ leads to a two-parameter distribution of possible infectiousness curves.

S4 Real and synthetic patient-level data

To demonstrate that the synthetic data that we generated (see Methods in the main text) was comparable to real data, we plotted real data from 6 patients collected in a cohort study of influenza infection (figure S1*a*–*f*) [5, 6] alongside synthetic data for 6 patients (figure S1*g*–*l*).

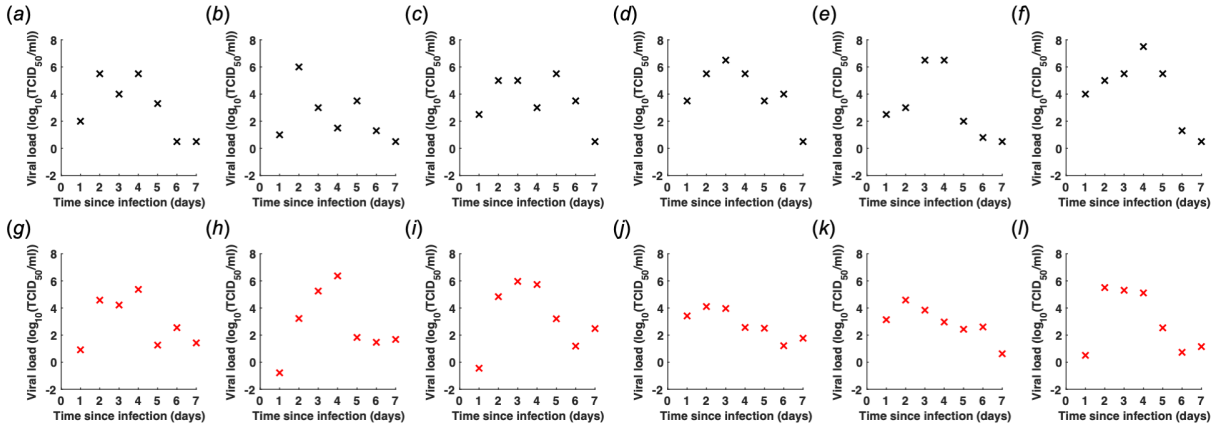

Figure S1: Real and synthetic patient-level data. (a–f) Viral load data against time since infection (black crosses) for 6 patients, taken from a cohort study of influenza infection and previously used to parameterise the TCL within-host model [5, 6]. (g–l) Synthetic viral load data against time since infection for 6 patients (red crosses), generated using the TCL model (see Methods).

S5 How many compartments are required for accurate population-scale predictions?

In this section, we consider the errors in predictions that can arise when a finite number of compartments is used in our compartmental framework. For simplicity, we focus on the case considered in the main text where the infectiousness curve has bounded support, so that $\beta(\tau) = 0$ for $\tau > T$, and we can choose the mean time a given host spends in each infected compartment to be $\lambda(n) = T/n$.

When taking the limit $n \to \infty$ in equation (S2.7), the leading order error terms are $\mathcal{O}(\lambda(n)) = \mathcal{O}(1/n)$. In this case of an infectiousness curve with bounded support, we will also have

$$\beta(\tau; n) = \beta(\tau) + \mathcal{O}(1/n), \tag{S5.1}$$

for all $\tau > 0$, so that the error terms when taking the limit in equation (S2.6) will also be $\mathcal{O}(1/n)$. This therefore suggests that the error in the population-scale dynamics (as defined by equation (4.10) in the main text) when the compartmental method is used, relative to the dynamics predicted using the IDE method, will be $\mathcal{O}(1/n)$ as the number of compartments, $n \to \infty$.

In order to confirm this error estimate, and also to investigate how many compartments are required for sufficiently accurate population-scale forecasts in our case study of influenza A, we assumed that the patient-level dynamics were perfectly characterised so that the true expected infectiousness curve (as shown in figure 2a of the main text) was known exactly. We then generated the population-scale dynamics for different values of the number of infected compartments, n, and compared these to the dynamics predicted using the IDE method (figure S2a).

To quantify the improvement of the approximation as the number of compartments is increased, we plotted the error in the population-scale dynamics against the number of compartments, n (figure S2b). In this case, $n = 24$ compartments were required for an error of 10% or below, while 47 compartments were needed for a 5% error (figure S2b). We also plotted the error on a log-log scale for a wider range of n, in order to confirm that the error scales with $1/n$ as the number of compartments, n, becomes large (figure S2c).

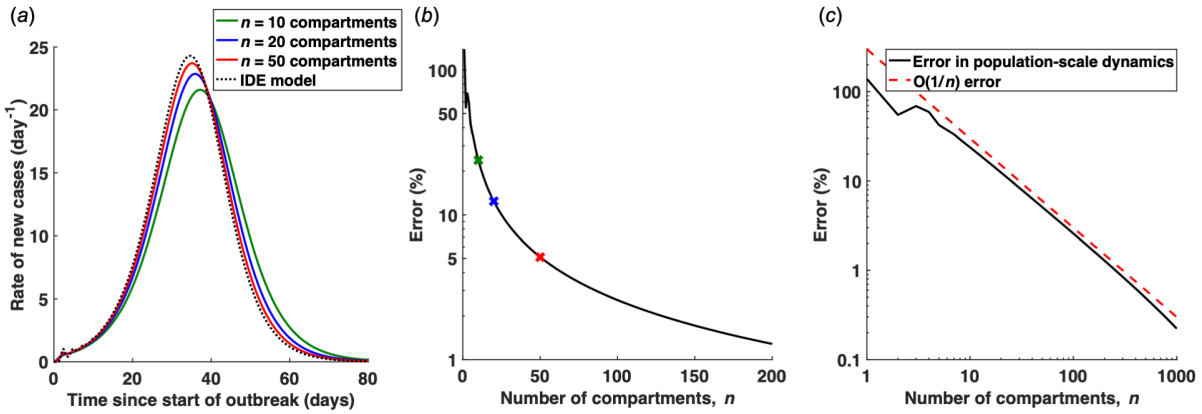

Figure S2: How many compartments are required for accurate population-scale predictions? (a) The population-scale dynamics, using our compartmental approach with n =10 (green), 20 (blue) and 50 (red) infected compartments, and using the IDE method (black dotted), when the infectiousness curve is as in figure 2a of the main text. (b) The error in the population-scale dynamics against the number of compartments used in our framework, n, plotted on a log-scale. The crosses represent the errors corresponding to the curves of the same colour in panel (a) (these are at values of 24%, 12% and 5% error). (c) The error in the population-scale dynamics for a larger range of n (black), plotted on a log-log scale and compared to an error proportional to $1/n$ (red dashed).

S6 Results using the IDE method to transition from patient-level to population-scale dynamics

We repeated the analyses in figures $3d$, $4d$ and $5a$ of the main text, but using the IDE method rather than our compartmental framework to transition from patient-level to population-scale dynamics (figure S3), finding that the results we obtained using the two approaches were almost identical.

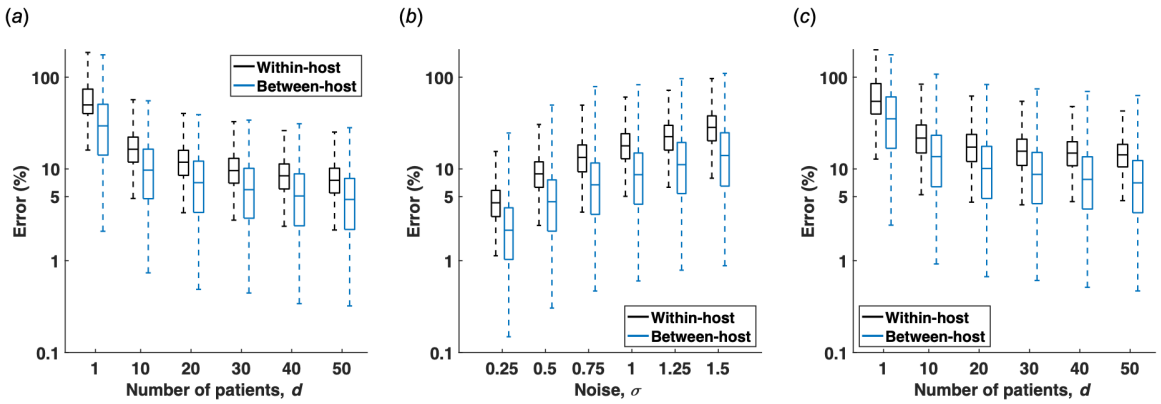

Figure S3: Results using the IDE method to transition from patient-level to population-scale dynamics. Panels (a–c) are equivalent to figures $3d$, $4d$ and $5a$ in the main text, respectively, but using the IDE method instead of our compartmental approach to calculate population-scale dynamics.

S7 Increased frequency of data collection

In figure 5 of the main text, we assumed that patient-level data were collected once daily for a week from each patient. If data are instead collected twice daily, then estimates of individual patient-level dynamics may be more accurate (figure S4a). We repeated our analyses in figure 5, but assuming twice daily data collection from each patient (figure S4b–c), finding that in this case data are only required from 20 patients for a 10% mean error (figure S4b).

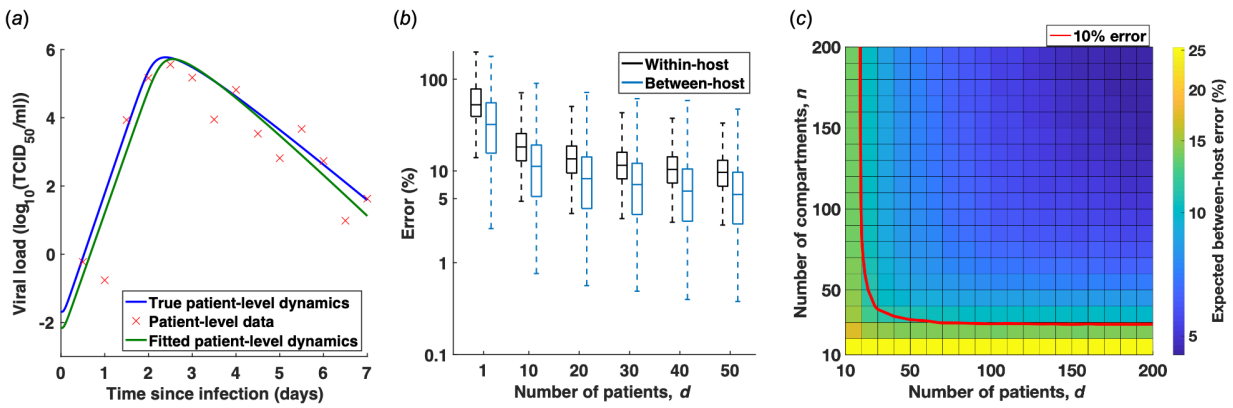

Figure S4: Results assuming patient-level data are collected twice daily from each patient. (a) Example of synthetic data for a single patient: the true viral load of the patient against time since infection (blue), twice daily data with measurement noise level $\sigma = 1$ $\log_{10}$(TCID$_{50}$/ml) (red crosses), and the viral load against time when the TCL model is fitted to the data (green). (b) Box-and-whisker plots indicating the distributions of within-host (black) and between-host (blue) errors for different patients chosen in the study cohort when $n = 1000$ compartments are used in our framework, assuming a measurement noise level of $\sigma = 1$ $\log_{10}$(TCID$_{50}$/ml), for a range of values of the number of patients, d. (c) The expected error in the population-scale dynamics when the compartmental method is used, against the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%.

We also considered a potential trade-off between the number of patients, d, and the number of daily observations per patient, f, if the total number of measurements, $M = 7fd$, is fixed. In particular, we considered values of M that were multiples of 84, so that data acquisition was possible at $f = 1, 2, 3$ and 4 daily measurements from each patient. In figure S5, the expected error in the population-scale dynamics (each time calculated over 10,000 repeats, using $n = 1000$ infected compartments in our framework) is plotted against M for $f = 1, 2, 3$ and 4 (with $d = M/(7f)$ in each case). At larger total numbers of measurements, the expected errors were found to be similar when data were collected either once or twice daily from each patient, whereas data collection 3 or 4 times daily (from a smaller number of patients) could led to larger errors for a given value of M.

Frequencies of observation smaller than once daily could also be considered, in order to find an optimal value of f (or d) for each M. We plan to explore this in future, using a modelling approach in which the data from different individuals are partially pooled (such as a non-linear mixed effects model), in order to ensure accurate parameter estimation in cases where the numbers of data points per patient are very small. Such an approach could also be used

to explore different timings of measurements.

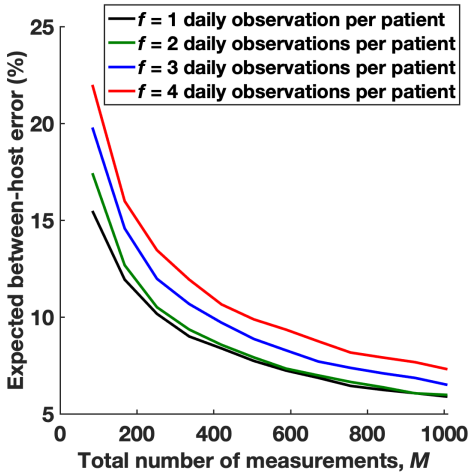

Figure S5: Trading off the number of patients, d, and number of daily observations per patient, f, when the total number of measurements, $M = 7fd$, is fixed. The expected error in the population scale dynamics is plotted against M for $f = 1$ (black), 2 (green), 3 (blue) and 4 (red), with $d = M/(7f)$ in each case. A measurement noise level of $\sigma = 1$ $\log_{10}(\text{TCID}_{50}/\text{ml})$ is assumed.

S8 Alternative relationships between pathogen load and infectiousness

In the main text, we assumed that the infectiousness, $\beta^{(i)}(\tau)$, of an influenza-infected host (where i represents the particular host under consideration) was proportional to their viral load, $V^{(i)}(\tau)$, at any time since infection, τ days. In figure S6, we considered two alternative cases in which we instead assumed

$$\beta^{(i)}(\tau) = k \times F\left(V^{(i)}(\tau)\right), \tag{S8.1}$$

where $F(V)$ is a known function of viral load, V, and k is a constant. In particular, we considered the following two possibilities:

i) A case where infectiousness scales with the logarithm of the viral load, in which we took

$$F(V) = \max\{\log_{10}(V/V^*), 0\}, \tag{S8.2}$$

where we assumed that $V^* = 10^3$ TCID_{50}/ml.

ii) A case where infectiousness saturates at high viral loads, in which we took

$$F(V) = \frac{V}{K+V}, \tag{S8.3}$$

where we assumed that $K = 10^5$ TCID_{50}/ml.

Assuming that, R_0, was known, the expected infectiousness curve was therefore given in each case by

$$\beta(\tau) = \frac{R_0}{N\int_0^\infty \overline{F\left(V^{(i)}(\tau)\right)}\mathrm{d}\tau}\overline{F\left(V^{(i)}(\tau)\right)}, \tag{S8.4}$$

where the bar denotes the average (at a given time since infection, τ days), calculated over a large number of realisations of the within-host model. To compute the "true" expected infectiousness curve, $\beta(\tau)$, we calculated the average of $F\left(V^{(i)}(\tau)\right)$ over 10,000 within-host realisations (figures S6a and S6d). In cases where data were available from d patients, each sampled once daily, we fitted the within-host model to the data for each patient in order to estimate $\beta(\tau)$ using equation (S8.4).

For both choices of $F(V)$, assuming perfectly characterised patient-level dynamics so that the infectiousness curve was known exactly (figures S6a and S6d), we used both the compartmental and IDE methods (see Methods in the main text) to transition to population-scale dynamics (figures S6b and S6e). When our compartmental approach was used, we assumed that the expected infectiousness was very small for times since infection greater than $T = 8$ days. We then repeated our analyses in figure 5b of the main text (figures S6c and S6f), finding similar results to those obtained in figure 5b for both choices of $F(V)$—for example, when there were $d = 40$ patients and we used $n = 200$ compartments in our framework, the mean population-scale errors were 9.7% (figure 5b), 9.2% (figure S6c) and 9.5% (figure S6f) under the assumptions of a linear (equation (4.5) in the main text), log-linear (equation (S8.2)) or saturation (equation (S8.3)) relationship, respectively.

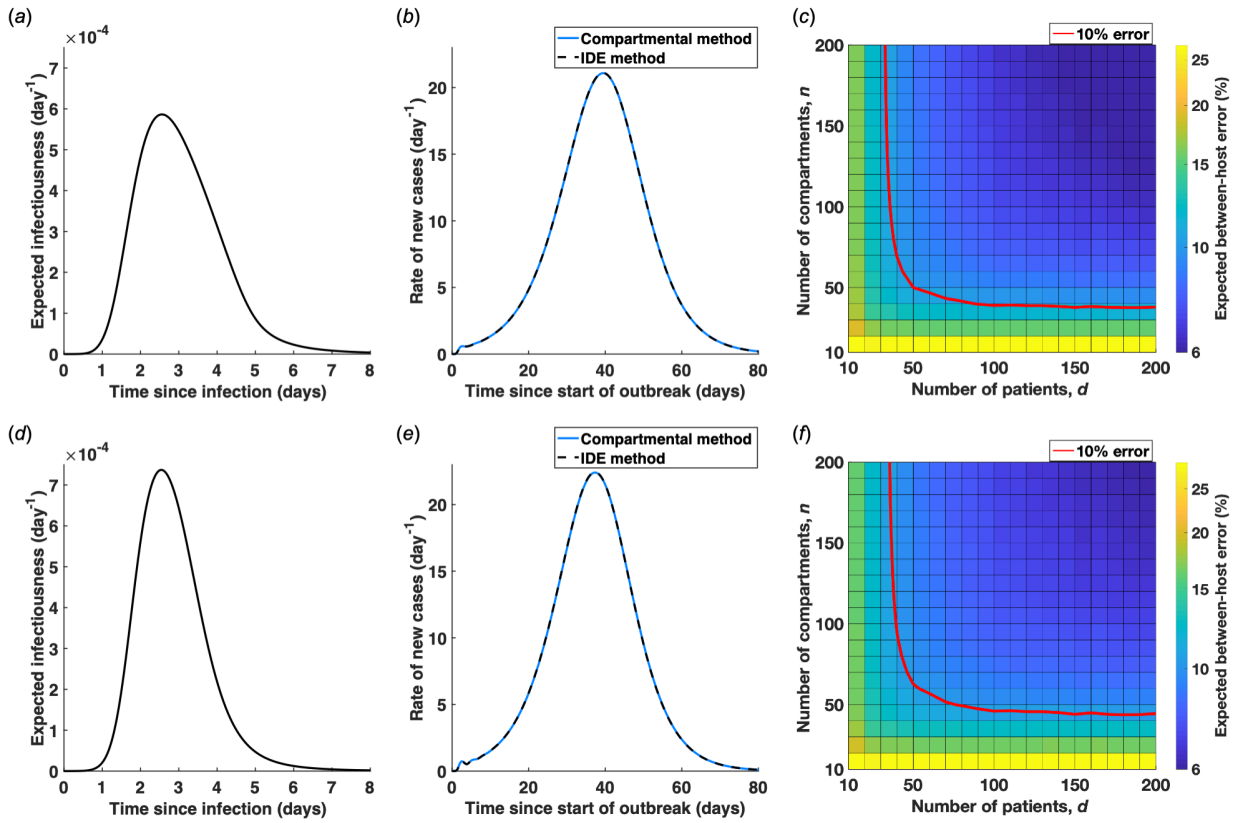

Figure S6: Alternative relationships between pathogen load and infectiousness. (a–c) Results assuming infectiousness scales with the logarithm of the viral load (equation (S8.2)). (a) The expected infectiousness curve, $\beta(\tau)$, when the patient-level dynamics are perfectly characterised. (b) The population-scale dynamics, using our compartmental approach with $n = 1000$ infected compartments (blue), and using the IDE method (black dashed), when the infectiousness curve is as in panel (a). (c) The expected error in the population-scale dynamics when the compartmental method is used, against the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%. A measurement noise level of $\sigma = 1$ $\log_{10}(\text{TCID}_{50}/\text{ml})$ is assumed. (d–f) Equivalent figures to (a–c), but assuming a relationship where infectiousness saturates at high viral loads (equation (S8.3)).

S9 Results for different values of the basic reproduction number, R_0

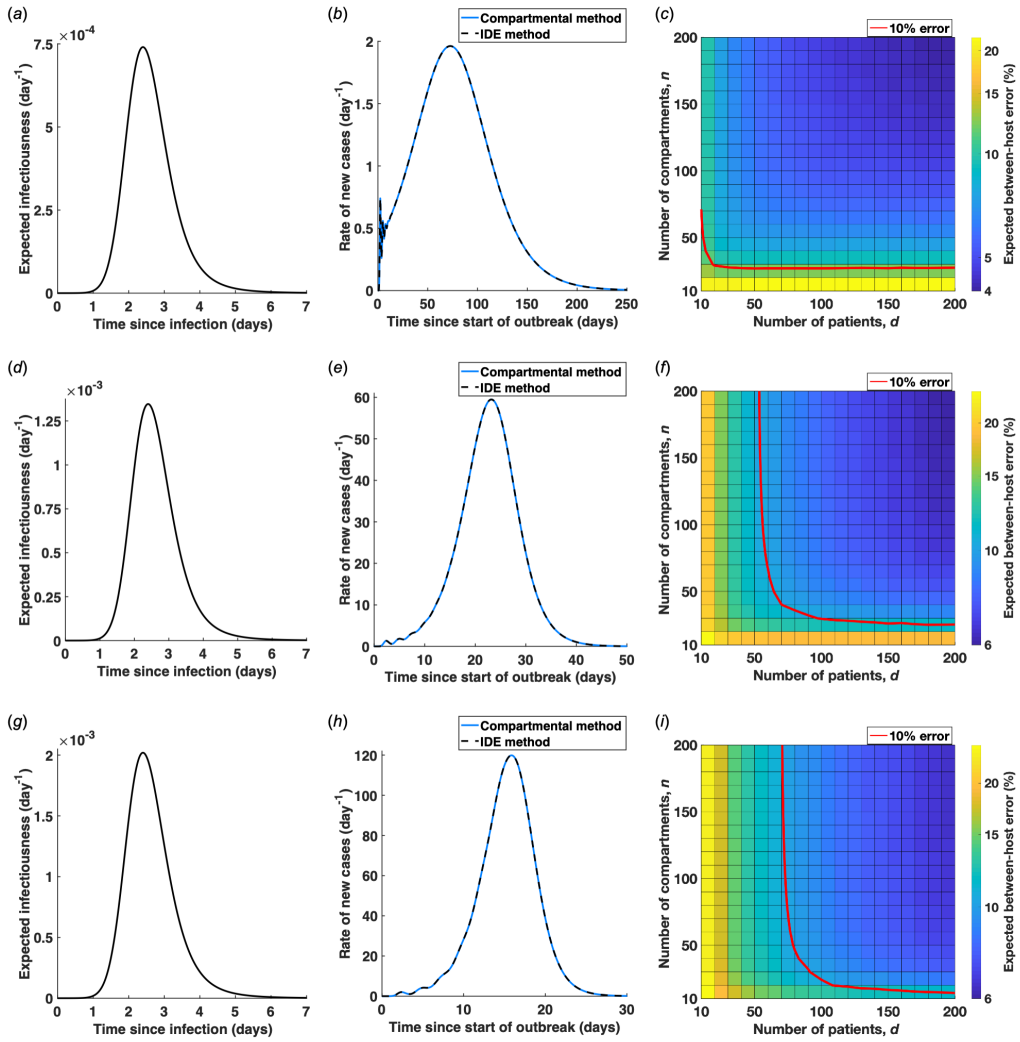

Figure S7: Results for different values of the basic reproduction number, R_0. (a) The expected infectiousness curve, $\beta(\tau)$, when the patient-level dynamics are perfectly characterised, and assuming a basic reproduction number of $R_0 = 1.1$. (b) The population-scale dynamics, using our compartmental approach with $n = 1000$ infected compartments (blue), and using the IDE method (black dashed), when the infectiousness curve is as in panel (a). (c) The expected error in the population-scale dynamics when the compartmental method is used, against the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%. A measurement noise level of $\sigma = 1$ $\log_{10}(\text{TCID}_{50}/\text{ml})$ is assumed. $(d$–$f)$ Equivalent figures to $(a$–$c)$, for $R_0 = 2$. $(g$–$i)$ Equivalent figures to $(a$–$c)$, for $R_0 = 3$.

We considered the effect of the assumed value of the basic reproduction number, R_0, on our results (figure S7). The number of patients required for a mean population-scale error of 10% or below was found to increase with R_0, with $d = 10$ patients required when $R_0 = 1.1$ (figure S7c) and $d = 70$ patients required when $R_0 = 3$ (figure S7i). We also found that fewer compartments were required in our framework to ensure a mean error of 10% or below, for values of R_0 both lower and higher than 1.5 (the value considered in most of our analyses).

S10 Results for different values of the measurement error level, σ

We repeated our analyses in figure 5b of the main text for different values of the measurement noise level, σ, finding that both the number of patients, d, and number of compartments, n, required for a population-scale error of 10% or below increased when there was more measurement error (figure S8).

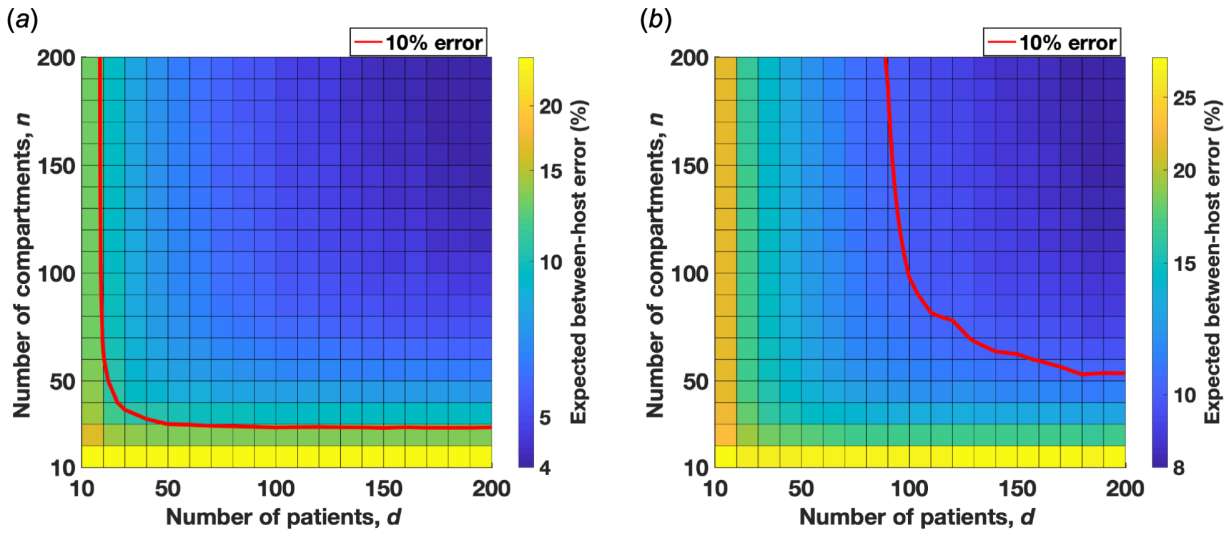

Figure S8: Results for different values of the measurement error level, σ. (a) The expected error in the population-scale dynamics when $\sigma = 0.5$ $\log_{10}(\text{TCID}_{50}/\text{ml})$, against the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%. (b) Equivalent figure to (a), for $\sigma = 1.5$ $\log_{10}(\text{TCID}_{50}/\text{ml})$.

S11 Results for different levels of variability in patient-level parameter values

When we generated patient-level data, we sampled the logarithms of the parameters δ and $V(0)$ in the TCL model from normal distributions with standard deviations 0.25 $\log_{10}(\text{day}^{-1})$ and 1.12 $\log_{10}(\text{TCID}_{50}/\text{ml})$, respectively, for each patient. These values were chosen to match variability in previous individual parameter estimates for 6 patients [5], but the level of variability was not known to high accuracy. Therefore, we also considered changing the standard deviations of the distributions from which we sampled $\log_{10}(\delta)$ and $\log_{10}(V(0))$ to 0.25α $\log_{10}(\text{day}^{-1})$ and 1.12α $\log_{10}(\text{TCID}_{50}/\text{ml})$, respectively, and repeated our analyses in figure 5*b* of the main text for different values of α (figure S9). The number of patients required for a 10% population-scale error increased with α (figure S9).

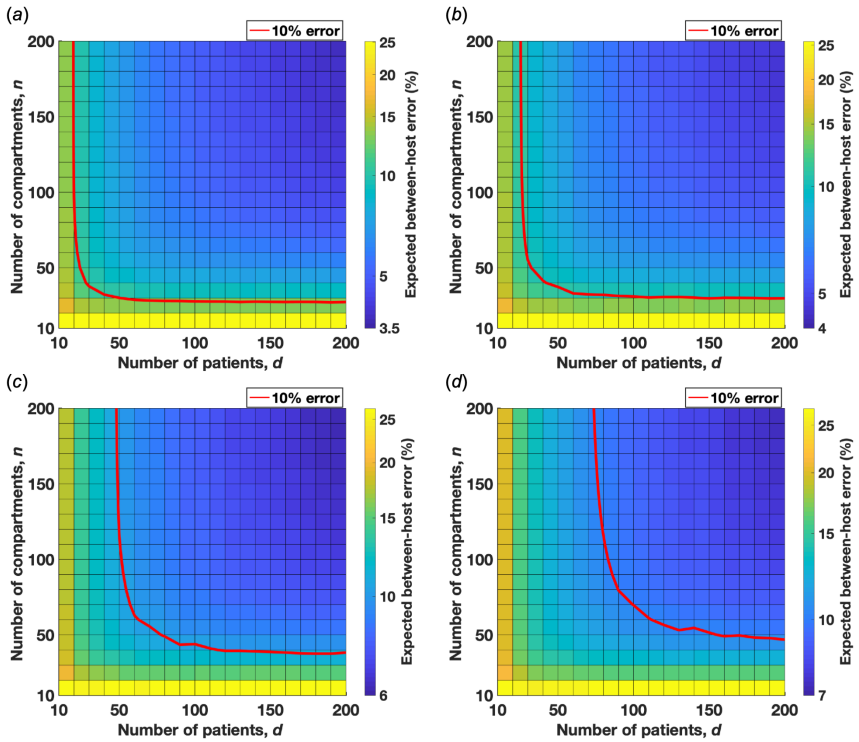

Figure S9: Results for different levels of variability in patient-level parameter values. (*a*) The expected error in the population-scale dynamics when the parameter $\alpha = 0.8$ (see the text in Section S11), against the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%. A measurement noise level of $\sigma = 1$ $\log_{10}(\text{TCID}_{50}/\text{ml})$ is assumed. (*b*) Equivalent figure to (*a*), for $\alpha = 0.9$. (*c*) Equivalent figure to (*a*), for $\alpha = 1.1$. (*d*) Equivalent figure to (*a*), for $\alpha = 1.2$.

References

[1] Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *P. R. Soc. A* **115**, 700–721.

[2] Diekmann O, Heesterbeek JAP. 2000 *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*. Chichester: John Wiley & Sons.

[3] Cunniffe NJ, Stutt ROJH, van den Bosch F, Gilligan CA. 2012 Time-dependent infectivity and flexible latent and infectious periods in compartmental models of plant disease. *Phytopathology* **102**, 365–380.

[4] Diekmann O, Gyllenberg M, Metz JA. 2018 Finite Dimensional State Representation of Linear and Nonlinear Delay Systems. *J. Dyn. Differ. Equations* **30**, 1439–1467.

[5] Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. 2006 Kinetics of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599.

[6] Murphy BR, Rennels MB, Douglas RG, Betts RF, Couch RB, Cate TR, Chanock RM, Kendal AP, Maassab HF, Suwanagool S, Sotman SB, Cisneros LA, Anthony WC, Nalin DR, Levine MM. 1980 Evaluation of influenza A/Hong Kong/123/77 (H1N1) ts-1A2 and cold-adapted recombinant viruses in seronegative adult volunteers. *Infect. Immun.* **29**, 348–355.