A theoretical framework for transitioning from patient-level to population-scale epidemiological dynamics: influenza A as a case study

AUTHORS

W.S. Hart[1]*, P.K. Maini[1], C.A. Yates[2], R.N. Thompson[1,3]

*Correspondence to: william.hart@keble.ox.ac.uk

AFFILIATIONS

[1]Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

[2]Centre for Mathematical Biology, University of Bath, Claverton Down, Bath BA2 7AY, UK

[3]Christ Church, University of Oxford, St Aldates, Oxford OX1 1DP, UK

ABSTRACT

Multi-scale epidemic forecasting models have been used to inform population-scale predictions with within-host models and/or infection data collected in longitudinal cohort studies. However, most multi-scale models are complex and require significant modelling expertise to run. We formulate an alternative multi-scale modelling framework using a compartmental model with multiple infected stages. In the large-compartment limit, our easy-to-use framework generates identical results compared to previous more complicated approaches. We apply our framework to the case study of influenza A in humans. By using a viral dynamics model to generate synthetic patient-level data, we explore the effects of limited and inaccurate patient data on the accuracy of population-scale forecasts. If infection data are collected daily, we find that a cohort of at least 40 patients is required for a mean population-scale forecasting error below 10%. Forecasting errors may be reduced by including more patients in future cohort studies or by increasing the frequency of observations for each patient. Our work therefore provides not only an accessible epidemiological modelling framework, but also insight into the data required for accurate forecasting using multi-scale models.

KEYWORDS

epidemiological model; infectious disease outbreak forecasting; multi-scale model; nested model; longitudinal study; cohort study

1. INTRODUCTION

Infectious disease epidemics in humans, animals and plants have severe impacts [1–7]. Mathematical models are increasingly used to forecast the future dynamics of outbreaks

47 [7–9] and to plan interventions [10–13], while within-host models are used to understand
48 the spread of infection at the individual host-level [14–17]. Standard population-scale
49 epidemiological models assume that the infectiousness of each host is constant over
50 the course of the infectious period [4], but in reality infectiousness will vary as the
51 infection progresses through the host due to changing pathogen loads [18,19] and other
52 factors including behavioural responses to infection [18,20].
53
54 Multi-scale models have been used to connect epidemiological dynamics at the patient-
55 level (within-host; we refer to “patients” throughout but similar ideas can be applied to
56 pathogens of animals or plants) to those at the population-scale (between-host) [21–34].
57 These models (sometimes referred to as nested models [19,30]) tend to assume a
58 specified relationship between the level of infection within a patient and the rate at
59 which the patient transmits the pathogen to susceptible individuals [18,19,35]. A within-
60 host model, parameterised by fitting to patient data, is then used to determine the
61 parameters of a population-scale model incorporating time-dependent infectiousness
62 [19,35]. In addition to patient-level dynamics affecting population-scale transmission,
63 there may be reciprocal feedback from the population-scale to the patient-level [19] – for
64 example, if there are multiple co-circulating strains of the pathogen [24].
65
66 A recent review concluded that, while numerous multi-scale epidemiological modelling
67 studies exist, relatively few include substantial use of data [29]. While one reason for
68 this is the lack of widely available datasets [18,36], we contend that another contributing
69 factor is that previous multi-scale modelling frameworks have been complex, making

them challenging to implement other than by highly specialist mathematical modellers. Such frameworks have often employed integro-differential equations (IDEs) [19,24,27,30,31,33,35,37], although alternatives such as individual-based stochastic models [12,23,25,26,38] have also been considered. IDEs are challenging to solve, requiring bespoke numerical methods [28]. Some studies using IDEs have involved explicit simulation of the full multi-scale model [24,31,37]. However, others have either only used the multi-scale framework to derive quantities such as the basic reproduction number of the pathogen rather than predicting temporal epidemic dynamics [30,33], or have made simplifying assumptions such as taking a within-host model to be in equilibrium [21,22,27,32]. Although an assumption that the pathogen load in each infected host is not changing (or changes only a limited number of times) might be appropriate for chronic infections, it leads to an approximate population-scale model that does not explicitly account for time-dependent infectiousness or other potentially complex patient-level dynamics.

In most previous studies that have used IDEs to transition from within- to between-host, the progression of infection through all patients has been assumed to be identical [35]. Patient-level dynamics are therefore characterised by a within-host model in which the values of model parameters (describing factors such as pathogen replication as well as immune responses) are the same for all patients. These parameters have either assumed values that have not been derived rigorously from data [29], or have been obtained by fitting the model to data collected in longitudinal cohort studies from a small number of patients [27,33]. Within-host parameters are, in fact, likely to vary between

individuals [14,39,40], for example due to differences in immune responses [41], while measurement error may also lead to inaccurate parameter estimates particularly given limited numbers of observations [35,42]. As we show, if patient-level data are only available from a limited number of patients, then predictions of population-scale epidemic dynamics may be inaccurate.

In this paper, we introduce a novel framework for transitioning from within- to between-host epidemiological dynamics straightforwardly. Our method involves using a compartmental model with a large number of infected compartments to predict the population-scale dynamics. Compartmental models, comprising systems of ordinary differential equations (ODEs), can be solved easily using standard numerical routines and software packages [4,28,43,44], are straightforward to adapt to include further biological detail [4,20,28], and are widely used for epidemic modelling [4,45]. We show rigorously that our modelling framework is equivalent to a more complex IDE approach, in the large-compartment limit of our method. Since the number of compartments is simply a choice for the user to make, our easy-to-use method can generate results that are as accurate as those from more complex approaches.

To demonstrate our framework, we consider modelling an outbreak due to the influenza A virus. We use a previously parameterised within-host model [14] to generate a synthetic dataset representative of real patient data (figure S1), incorporating variability in the viral load time series between patients due to factors such as differences in immune responses. Since the magnitude of this variability has been chosen to match

data from a previous cohort study [14,46], our dataset is comparable to obtaining data from a cohort study, but with the advantage that we can test our approach using many different possible cohorts of any size (from small cohort sizes up to very large cohort sizes that generate idealised data). We explore the effects of both the number of patients from which data are available, and the extent of measurement error in patient data, on population-scale predictions. Our work therefore provides insight into the data required for accurate forecasting using multi-scale epidemic models, as well as an accessible modelling framework that can be used for forecasting during future epidemics of a range of infectious diseases.

2. RESULTS

Transitioning from within- to between-host influenza dynamics

We have developed a new compartmental framework for transitioning from within- to between-host epidemic dynamics. In our approach, a within-host model is fitted to data from individual patients, to estimate the pathogen load of each measured patient at every time since infection (figure 1*a*). As with other multi-scale epidemic models [18,19,35], by assuming a functional relationship between pathogen load and infectiousness, the expected infectiousness, $\beta(\tau)$, of any host is then estimated at each time since infection, τ days (figure 1*b*). We will call $\beta(\tau)$ the *expected infectiousness curve*. In previous approaches, the Kermack and McKendrick (K&M) IDE model [47] (see Section S1 of the Supplementary Material) has then been used to calculate the population-scale dynamics (figure 1*c*). We instead use the expected infectiousness

curve to parameterise a multi-stage compartmental model with a large number of infected compartments (figure 1*d*), which can also be used to predict the population-scale dynamics (figure 1*e*). For details on the compartmental and IDE approaches, see Methods. In the limit of infinitely many compartments in our framework, the two approaches are mathematically equivalent (we prove this rigorously in Section S2).

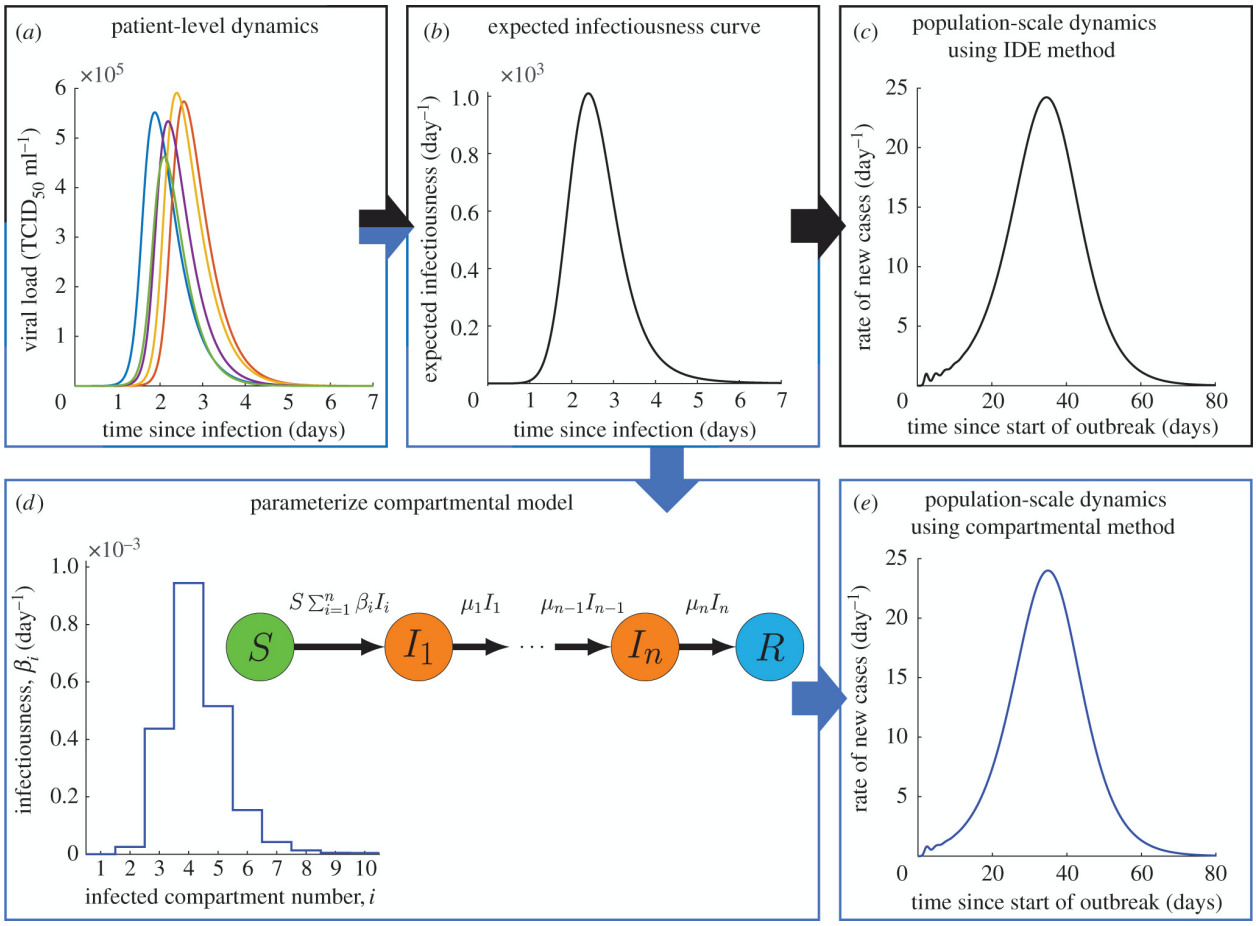

Figure 1. Schematic demonstrating methods for transitioning from patient-level to population-scale epidemiological dynamics. In standard approaches (*a*-*c*), measurements from within each patient (*a*) are used to parameterise a within-host model, giving rise to an averaged infectiousness curve (*b*). These patient-level dynamics can then be nested in an IDE model (e.g. the K&M model [47]) used to predict the population-scale dynamics (*c*). However, IDE models are challenging to solve. In contrast, in

our approach (*a*,*b*,*d*,*e*) the expected infectiousness curve is instead used to parameterise a compartmental model (*d*) that can be used to predict population-scale dynamics straightforwardly (*e*). Early-epidemic oscillations in panels (*c*) and (*e*) occur because the expected infectiousness of an infected host is close to zero in the first day of infection, leading to delays before successive generations of newly infected hosts begin to transmit the pathogen.

To illustrate our framework in a concrete setting, we considered the specific case of influenza A infection in humans. We used the target cell-limited (TCL) within-host model, which has previously been fitted to data from a cohort study of influenza infection [14], to generate synthetic data from a large number of patients (see Methods). The data were used to calculate the expected infectiousness curve (figure 2*a*) under the assumptions of a linear relationship between viral load and infectiousness [18,33,35,38,39] and a basic reproduction number (defined to be the expected number of secondary cases arising from a single infected host in an otherwise entirely susceptible population [4]) of 1.5 [8] (see Methods), although we consider other assumptions and values of the reproduction number later (Sections S8 and S9).

Both our compartmental approach using a large number of infected compartments (n = 1000) and the previously used IDE method (i.e. the K&M model) were then used to predict the population-scale outbreak dynamics, initially assuming the expected infectiousness curve was known exactly (figure 2*b*). We considered a population of size N = 1000 and assumed a single newly infected individual was introduced into an entirely susceptible population. The two approaches produced almost identical results – the error in the predicted population-scale dynamics when the compartmental method was

used, calculated as a proportional error relative to the dynamics predicted using the IDE method (see Methods), was only 0.2%. We explored how many compartments are required in our framework to ensure accurate population-scale forecasts, finding that in general, the error in predictions scales with $1/n$ as the number of compartments, n, becomes large (Section S5). When the infectiousness curve shown in figure 2*a* was used to transition to population-scale dynamics, we found that $n = 24$ compartments are sufficient for an error in population-scale predictions of 10% or less (figure S2*b*).

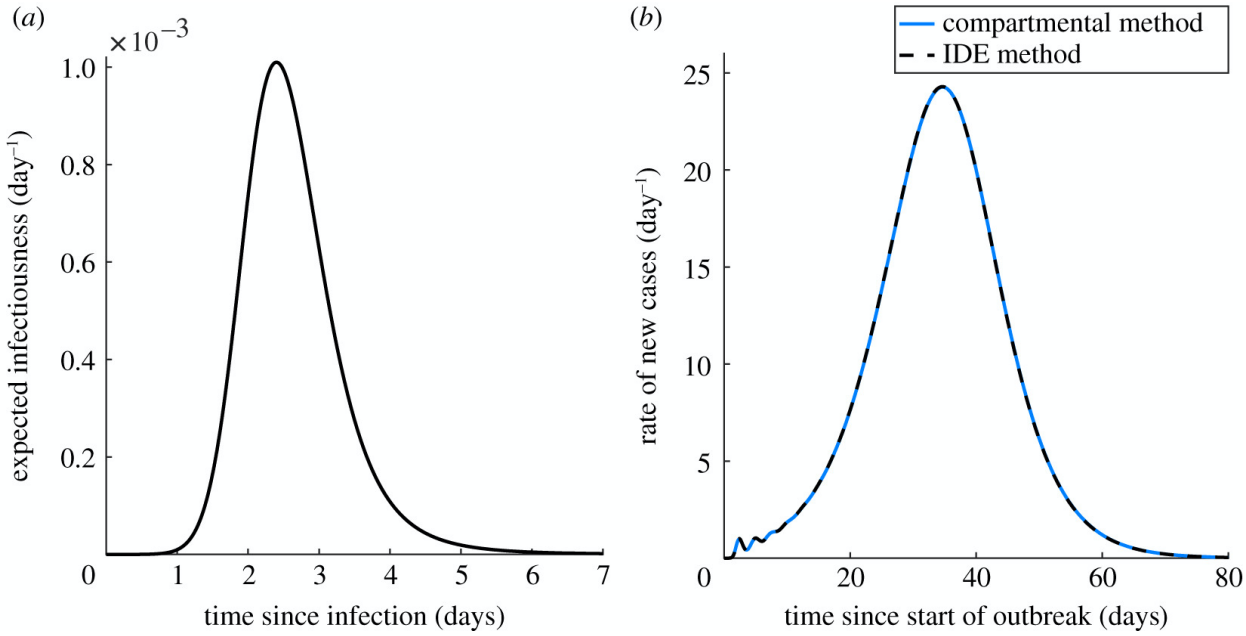

Figure 2. Transitioning from within- to between-host influenza dynamics using the compartmental and IDE methods. (*a*) The expected infectiousness curve, $\beta(\tau)$, when the patient-level dynamics are perfectly characterised. (*b*) The population-scale dynamics, using our compartmental approach with n = 1000 infected compartments (blue), and using the IDE method (black dashed), for the infectiousness curve shown in panel (*a*).

The effect of limited and inaccurate patient-level data on population-scale predictions

We considered the effect of two quantities on population-scale predictions: the number of patients from which individual patient data are available, and the extent of measurement error in patient-level data. Initially, we considered these two factors in isolation, before testing their combined effects. We defined error metrics to quantify the errors that arose in the patient-level dynamics (the within-host error) and in the population-scale dynamics (the between-host error), using proportional errors in order to enable comparison between errors at the two scales (see Methods).

Number of patients

In most cohort studies used to inform multi-scale models, data are only available from a small number of patients [27,33] and within-host parameters may vary significantly between patients [14,39,40]. To investigate the error in population-scale predictions that inadequate data may generate, we supposed that data were only available from d randomly chosen patients (see Methods). To isolate the effect of variability between hosts rather than measurement error, the exact viral load of each patient was initially assumed to be known at every time since infection. We used the available data to estimate the expected infectiousness curve (figure 3*a*) and calculated the approximate population-scale dynamics using our compartmental framework with $n = 1000$ infected compartments (figure 3*b*). For a fixed cohort size, d, significantly different predictions of

population-scale dynamics are possible, depending on which patients are included in the study (figure 3*c*). Therefore, we calculated the distributions of errors in both the patient-level and population-scale dynamics, relative to the case in which the patient-level dynamics were perfectly characterised, over 5000 repeats for each of a range of patient cohort sizes, *d* (figure 3*d*). Equivalent results using the IDE method rather than our compartmental approach are shown in figure S3*a*.

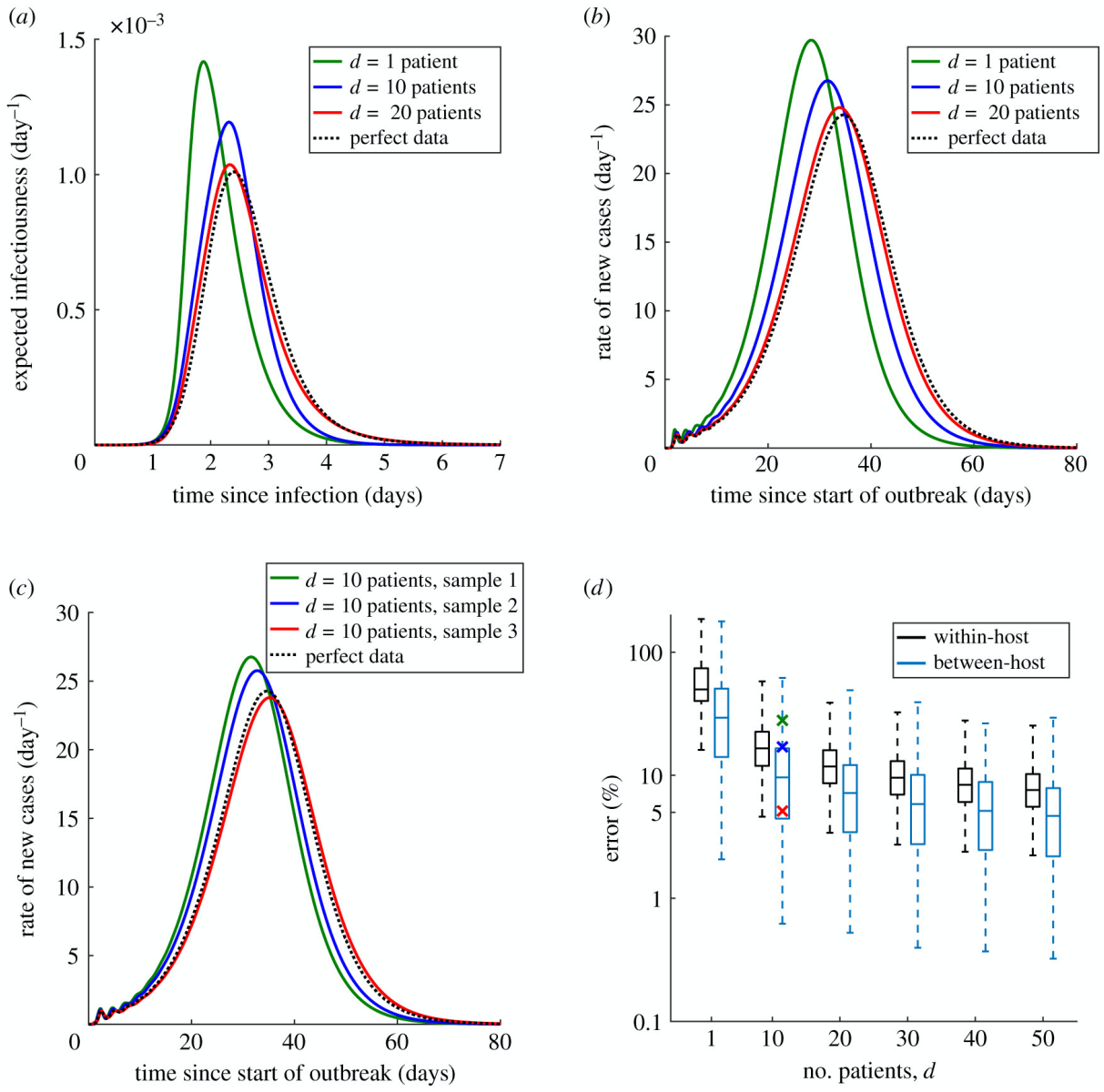

Figure 3. How many patients need to supply data for accurate population-scale predictions? (*a*) Examples of approximate expected infectiousness curves when (exact and continuous) data are available from $d = 1$ (green), $d = 10$ (blue) or $d = 20$ (red) randomly chosen patients, and the expected infectiousness curve when the patient-level dynamics are perfectly characterised (black dotted). (*b*) The predicted population-scale dynamics for each infectiousness curve in panel (*a*), using $n = 1000$ compartments in our framework. (*c*) Three examples of possible approximate population-scale dynamics, when data are available from $d = 10$ patients. (*d*) Box-and-whisker plots indicating the distributions of within-host (black) and between-host (blue) errors for different groups of patients randomly chosen in the study cohort, for a range of values of the number of patients, d. The boxes indicate the lower quartile, median and upper quartile, and the maximum length of each whisker is 1.5 times the interquartile range. The crosses represent the between-host errors corresponding to the curves of the same colour in panel (*c*) (these are at values of 28%, 17% and 5% error).

As the number of patients is increased, the errors at patient-level and population-scale both decrease in general (depending on precisely which patients are included in the study cohort), but at a decreasing rate. The magnitude of the population-scale error is generally smaller than that of the patient-level error. Therefore, limited data do not necessarily preclude accurate population-scale predictions, even when there is a large amount of variability between different patients. In this case – when there are exact and continuous data available from each patient – a cohort size of $d = 20$ patients is sufficient for the between-host error to be 10% or less on average (figure 3*d*). However, since the errors are affected by the precise patients included in the cohort, more patients are required for a greater certainty of a small between-host error. For example, $d = 30$ patients are required to ensure that the upper quartile of between-host errors for cohorts of that size is less than 10%.

246

247 *Extent of measurement error*

248

249 In longitudinal cohort studies, data are only collected from each patient at a limited
250 number of time points. For studies of influenza infections, data may be collected daily
251 (for example [14,39]) over the course of infection, which lasts approximately one week
252 [48]. However, there can be significant measurement error whenever the viral load is
253 recorded [35,42].

254

255 We considered viral load values recorded daily for each patient for a week after
256 infection, although we also considered the effect of more frequent observations (Section
257 S7). To incorporate measurement error in the synthetic data, a normally distributed error
258 with standard deviation σ was applied to the logarithm of each measurement. To
259 estimate the patient-level dynamics, we fitted the TCL model to the data for each patient
260 (figure 4*a*, see Methods for details).

261

262 To demonstrate the effect of measurement error on population-scale predictions, we
263 assumed that data were available from $d = 10$ randomly chosen patients, and compared
264 estimates of the expected infectiousness curve, first under the assumption that the viral
265 load of each host was known exactly at all times during infection, and second when
266 there was measurement error in daily recordings of the viral load (figure 4*b*). Our
267 compartmental framework with $n = 1000$ compartments was then used to predict the
268 population-scale dynamics in both cases (figure 4*c*). We calculated the within-host and

between-host errors that arose directly due to measurement error, by taking the “true” dynamics to be those when exact and continuous data were available from the same 10 hosts. The distributions of these errors, each time calculated over 5000 repeats for a range of values of σ, are shown in figure 4*d* (for equivalent results obtained using the IDE method, see figure S3*b*).

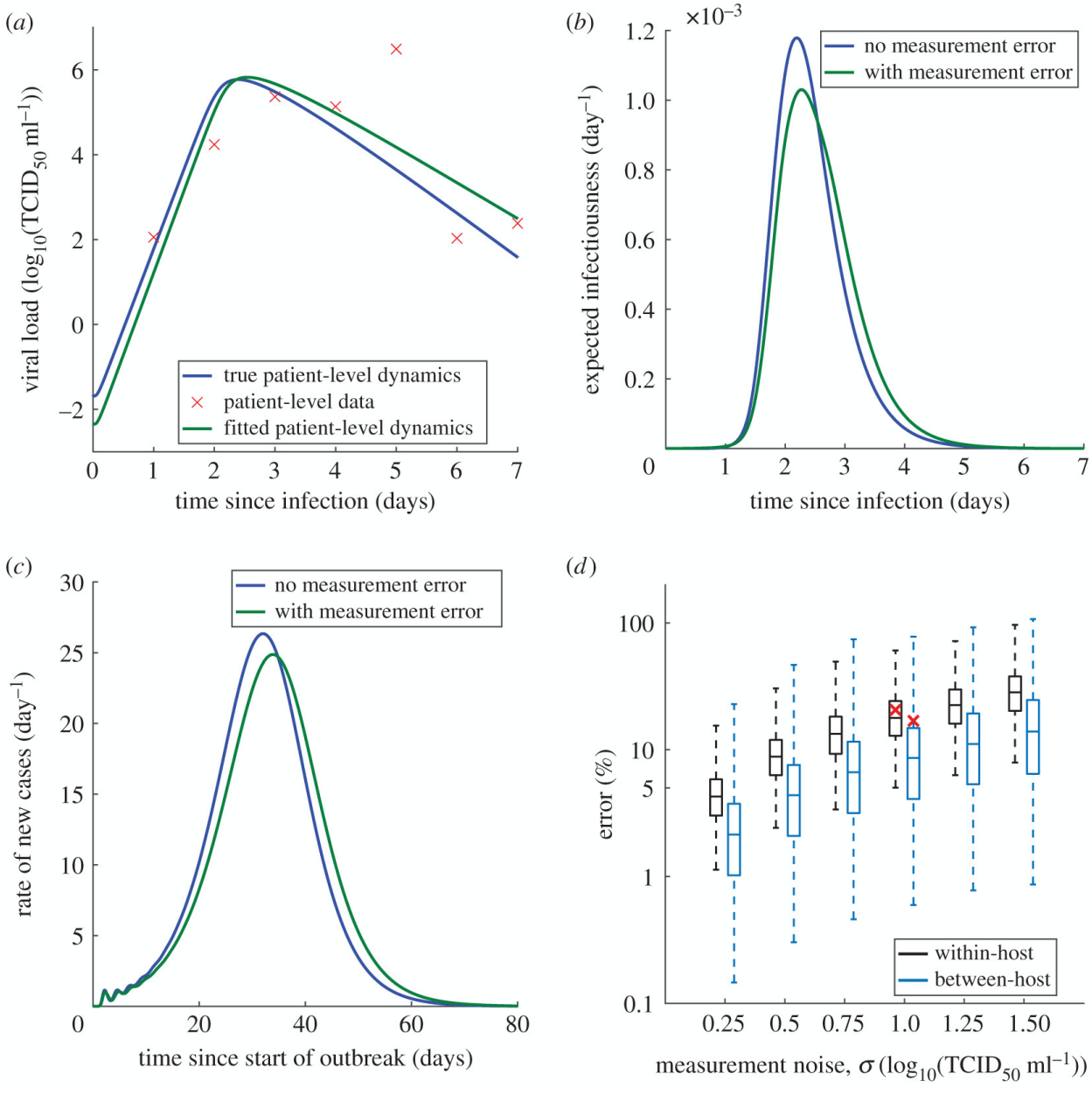

Figure 4. The effect on population-scale predictions of measurement error in patient-level data. (*a*) Example of synthetic data for a single patient: the true viral load of the patient against time since infection (blue), daily synthetic data with a measurement noise level of σ = 1 $\log_{10}$(TCID50/ml) (red crosses), and the viral load against time when the TCL model is fitted to the data (green). In figure S1, synthetic viral load data generated using the TCL model are compared to the real data that were used to parameterise the TCL model [14]. (*b*) Examples of expected infectiousness curves, without measurement error (blue) and with measurement error (green), for d = 10 patients. (*c*) The population-scale dynamics for each infectiousness curve in (*b*), using n = 1000 compartments in our framework. (*d*) Box-and-whisker plots indicating the distributions of within-host (black) and between-host (blue) errors arising directly due to measurement error, for a range of values of the extent of measurement error, σ. The red crosses represent the within-host error corresponding to panel (*b*) and the between-host error corresponding to panel (*c*) (these are at values of 21% and 17% error, respectively).

The errors at patient-level and at population-scale both increase with the measurement noise level, σ. For values of σ of 1 $\log_{10}(\text{TCID}_{50}/\text{ml})$ or higher, the mean population-scale error is over 10%. In that case, when a cohort of only d = 10 patients is used, measurement error alone is likely to prevent accurate population-scale forecasts, even if there is no additional error contribution due to within-host parameter variability.

Overall error

So far, we have described our analyses considering the separate effects of patient cohort size and measurement noise on the characterisation of within-host viral load time series, as well as the resulting impact on population-scale outbreak predictions. However, in reality, both these sources of error would be present simultaneously. Errors

would also occur if a small number of compartments is used in our multi-scale modelling approach, although this can be avoided by simply choosing a large number of compartments in the model. Nonetheless, we also conducted an analysis in which all three potential sources of error were included: (i) number of patients; (ii) measurement error; (iii) number of compartments.

When we investigated the combined effect of these potential sources of error, we considered a measurement noise level of σ = 1 $\log_{10}(\mathrm{TCID}_{50}/\mathrm{ml})$, since this generated synthetic data comparable to those recorded in cohort studies (figure S1). Assuming that data were available from *d* randomly sampled patients, our compartmental framework with *n* infected compartments was used to estimate the population-scale dynamics. We repeated this analysis 10,000 times each for different pairs of values of *d* and *n*, each time calculating the within-host and between-host errors, relative to the case in which the patient-level dynamics were perfectly characterised and the IDE method was used (equivalent to using infinitely many compartments in our compartmental framework). The distributions of within-host and between-host errors, when a large number of compartments (*n* = 1000) is used in our framework, are plotted for different numbers of patients (*d*) in figure 5*a*. Equivalent results using the IDE method are shown in figure S3*c*. When either the compartmental or IDE approach is used, data from *d* = 40 patients are required for an average between-host error of 10% or below (compared to 20 patients if data are recorded exactly, i.e. with no measurement error – see figure 3*d* and figure S3*a*).

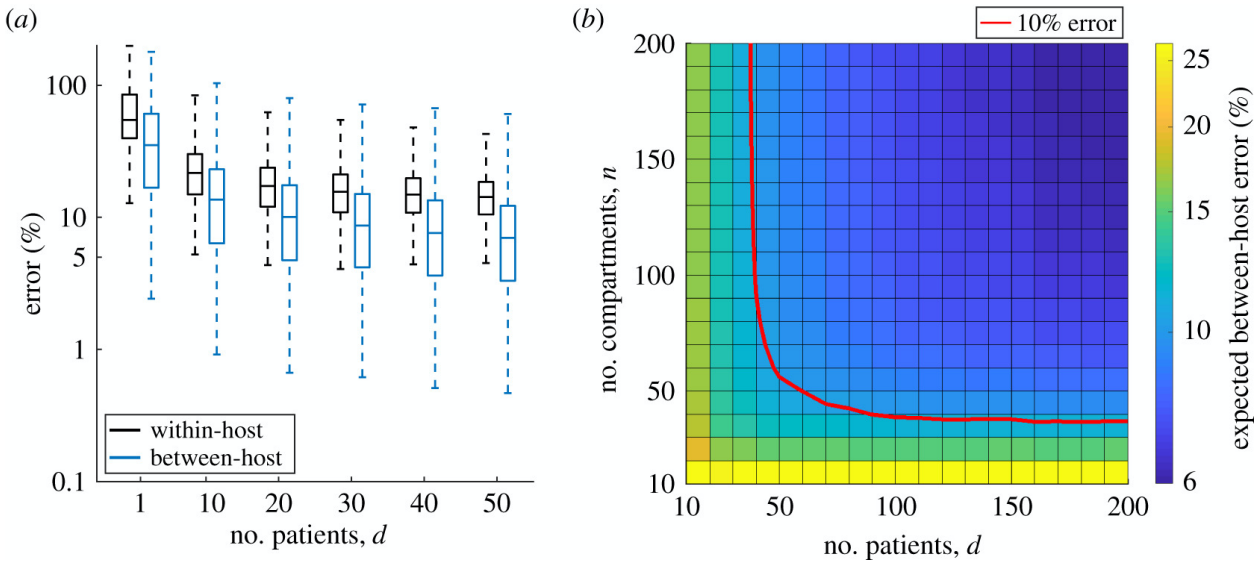

Figure 5. The effects of the number of patients and number of compartments on population-scale predictions. (*a*) Box-and-whisker plots indicating the distributions of within-host (black) and between-host (blue) errors for different patients chosen in the study cohort when n = 1000 compartments are used in our framework, assuming a measurement noise level of σ = 1 $\log_{10}(\text{TCID}_{50}/\text{ml})$, for a range of values of the number of patients, d. (*b*) The expected error in the population-scale dynamics, for different values of the number of compartments, n, and the number of patients, d. The red line indicates where the error is 10%.

The mean error in the population-scale dynamics, for different numbers of patients (d) and numbers of compartments (n), is shown in figure 5*b*. In the case of d = 40 patients, each sampled once daily, the user should choose at least n = 60 compartments in our approach to ensure a mean error of 10% or less. As few as 40 compartments are needed if data are available from a large number (more than 60) of patients. Nonetheless, since the number of compartments to use is simply a choice for the user – rather than requiring any more data to be collected – we suggest that any user of our framework simply chooses a very large number of compartments. We note, however,

that the benefit of using more compartments becomes negligible when more than n = 100 compartments are included in our approach (figure 5*b*).

Whereas in figure 5 we assumed data were collected once daily for a week from each patient, we also conducted supplementary analyses of the between-host error when data were instead collected twice per day from each patient (Section S7) – in this case, data are only required from 20 patients for a mean between-host error of 10% or less (figure S4*b*). If instead the total number of measurements that can be taken is fixed, then it might be necessary to choose between sampling a large number of patients infrequently, or a small number of patients frequently. We explored this in Section S7, and found that sampling patients more than twice per day tended to lead to less accurate population-scale predictions when the total number of measurements was fixed at values below 1000 (figure S5). For realistic cohort sizes, population-scale errors were similar when data were collected either once daily from 2d patients or twice daily from d patients (for example, when d = 20, the respective errors are both 10%).

We examined the robustness of our results to our assumptions when transitioning from within- to between-host (Section S8), finding similar results to those shown in figure 5 in two alternative cases, in which the infectiousness of each patient either scales with the logarithm of their viral load (figure S6*c*) or saturates at high viral loads (figure S6*f*). In addition, we considered the effect of the assumed value of the basic reproduction number, R_0, on our results (Section S9), and also repeated our analyses in figure 5 for different values of the measurement noise level, σ (Section S10), and for different levels

of variability in the within-host parameter values corresponding to different patients (Section S11). When R_0, σ, or the level of variability in within-host parameter values, exceeded the values considered in figure 5, a cohort size larger than 40 hosts was found to be required to ensure a mean between-host error of 10% or below (figures S7-S9) – for example, data from 70 patients are required if $R_0 = 3$ (figure S7*i*).

3. DISCUSSION

In this article, we have introduced a novel, easy-to-use, compartmental framework for nesting patient-level data in population-scale epidemic models. In the large-compartment limit, our method is mathematically equivalent to more complicated approaches that involve IDEs (Section S2). However, our method has the advantage that it can be used straightforwardly, allowing it to be applied widely in future. We have provided adaptable computing code alongside this article to facilitate future use of our approach (see Data Accessibility).

To illustrate our method, we considered the example of influenza A infection in humans. A viral dynamics model [14] was used to generate a synthetic dataset describing changing viral loads in a cohort of patients, which is representative of real patient data (figure S1). We showed how our compartmental framework can be used to predict the population-scale epidemic dynamics, and compared our predictions to forecasts using the more complicated K&M IDE model. The population-scale predictions from our framework closely matched those obtained using the IDE model, provided that a

386 sufficient number of compartments was employed in our approach (figure 2*b* and figure
387 S2).
388
389 The amount of data used in modelling studies of within-host influenza dynamics has
390 varied widely, with some studies using data from fewer than 10 patients [14] but others
391 more than 40 patients [39]. While multi-scale models have often been parameterised
392 using either no or limited data [29], drawing robust population-scale conclusions from
393 cohort studies involving a small number of patients is likely to be challenging, since
394 patient-level dynamics display significant variability between different individuals
395 [14,39,40]. We therefore assessed the errors that arise in predicted population-scale
396 dynamics as a result of limited patient data, as well as considering measurement errors
397 that can beset parameter inference from patient-level data [42]. We first investigated
398 these effects separately (figures 3 and 4), before considering both these effects in a
399 single combined analysis (figure 5). When patient data were collected once daily, we
400 found that data from at least 40 patients were required for a mean population-scale
401 error of 10% or smaller when either our compartmental approach or the IDE method
402 was used (figure 5*a* and figure S3*c*). However, since the precise value of the
403 population-scale error depended on the exact subset of patients that was included in the
404 study, the error could be either larger or smaller than 10% even when data were
405 available from 40 patients (figure 5*a*). As a result, larger numbers of patients can
406 increase the confidence that the error is below a pre-specified threshold value (figure
407 5*a*). We considered daily measurements of pathogen load, since this frequency of data
408 acquisition is common to a number of previous longitudinal studies of influenza

infections (e.g. [14,39]). However, the accuracy of population-scale predictions depends on the frequency with which data are collected (Section S7), so ensuring regular data collection from each patient in future cohort studies is important for accurate population-scale forecasting.

Our approach was motivated by earlier studies in which compartmental models with multiple latent and infectious stages were employed so that the standard assumption of exponentially distributed latent and infectious periods was relaxed [49–53]. The use of multiple stages allows for gamma distributed latent and infectious periods (the so-called "linear chain trick" [52] or "method of stages" [53]), and gamma distributions have been shown to characterise epidemiological periods accurately [49,51]. However, in those studies [49–53], the level of infectiousness is assumed constant throughout the infectious period. We were therefore also inspired by previous research in which time-dependent infectiousness was incorporated into multi-stage compartmental models, in cases where the compartments correspond to clearly distinct phases of infection (for example, studies of HIV [54] and Ebola [20]) or convenient time periods [55]. Our approach is more similar to a method used to include experimental data in models of plant disease [28], but differs from previous literature [20,28,54,55] due to our use of a large number of infected compartments corresponding to different infection rates in order to provide an easy method to transition from within- to between-host that is accurate for any patient-level infection dynamics.

We focussed on the case study of influenza A in humans because compartmental models are frequently used to model both patient-level and population-scale influenza dynamics, while there has also been significant interest in developing models linking the dynamics at the two scales [35]. In principle, however, our approach could be extended to model outbreaks of a range of other pathogens for which patient-level dynamics are well characterised. This would require careful consideration of the functional relationship between pathogen load and infectiousness, since this is likely to differ between pathogens [18]. In particular, the mode of transmission may be an important factor in determining suitable relationships for different pathogens.

To describe individual patient-level influenza dynamics, we used the simple TCL within-host model. More detailed within-host models exist, and involve features including a delay before target cells begin to shed virus (an eclipse phase) [14] or explicit modelling of innate and adaptive immune responses [56]. While the TCL model was sufficient to demonstrate our approach here, the expected infectiousness curve in our framework could be generated using a within-host model with any level of complexity. Alternatively, if patient-level infection dynamics are not well characterised, then an expected infectiousness curve that is estimated from transmission data [12,57], rather than within-host data, could also be embedded within our framework.

In order to generate synthetic patient-level data, we assumed that two within-host parameters varied between patients, using previous parameter estimates to determine the level of parameter variability [14]. We incorporated measurement error by adding a

normally distributed random variate to daily observations of the logarithm of the viral load (although we also considered other frequencies of data collection in Section S7). Differences in both the extent of measurement error, and the extent of parameter variability between patients, can lead to significant differences in population-scale errors (figures S8 and S9). Therefore, when our modelling framework is used to determine how many patients should be included in future cohort studies, careful consideration of the measurement error and the variability in pathogen load time series between patients is important.

The TCL model was fitted to the data from each patient using a basic least squares estimation approach, since the precise method of parameter inference is not central to our modelling framework. However, it would be straightforward to extend our approach to consider different error structures and methods for fitting models to patient-level data. In particular, a non-linear mixed effects modelling approach – amounting to a partial pooling of the data between individuals – could be used. This would enable robust parameter estimation in a real dataset, particularly in settings in which the numbers of data points per patient are small, and both the frequency and timing of data collection may vary between patients [58,59]. Going forward, we will use such a method to explore further whether or not there is an optimal balance between the number of patients and the frequency of measurements per patient, if total resources are limited (see Section S7).

In our main analyses, we made the common assumption that the infectiousness of an influenza-infected host is proportional to their viral load [18,33,35,38,39], although we also obtained similar results in two alternative cases in which infectiousness either scales with the logarithm of the viral load [33,38] or saturates at high pathogen loads [21,55] (Section S8). However, more complex possibilities could easily be incorporated into our framework. For example, future studies may also incorporate varying symptoms during infection into our approach [20,23,39], in order to account for dependency of transmissibility on behavioural factors in addition to pathogen load [18].

While we considered errors in population-scale predictions arising due to variability between different infected patients when data are limited, our results were obtained using a population-scale model in which the population was assumed to be homogeneous and well-mixed. Variability between different patients was assumed to be random, so that all infected hosts could effectively be assumed to follow the same averaged infectiousness curve. In Section S3, we provide mathematical justification for this averaging in the population-scale dynamics (see also [60]). We sought to develop our framework for transitioning from within- to between-host using the simplest possible population-scale model, but our compartmental approach could be generalised, for example, to models incorporating age structure, spatial effects, social contact networks or stochasticity [4]. In an age-structured model, different within-host parameter values (or even different models) could be used to describe patient-level dynamics in the different age groups, since there may be substantial differences in within-host dynamics between patients of different ages [61].

In summary, we have introduced a novel compartmental framework for nesting patient data in population-scale epidemiological models. We have demonstrated our easy-to-use approach in the context of influenza. Not only can our modelling approach be used to inform population-scale predictions with data from patients, but it can also be used to design cohort studies by determining which data need to be collected. As a result, clear communication between clinical epidemiologists who conduct cohort studies and epidemiological modellers will allow for optimal study design. Including patient-level dynamics in population-scale epidemiological models as proposed here has the potential to improve epidemic forecasts; we hope that the simplicity of our approach will facilitate its use for forecasting in a wide range of future outbreaks.

4. METHODS

Within-host model

The TCL model of viral dynamics, which has previously been used to model influenza infections [14,15,62], is given by

$$\begin{aligned} \frac{dT}{d\tau} &= -\beta TV, \\ \frac{dI}{d\tau} &= \beta TV - \delta I, \\ \frac{dV}{d\tau} &= pI - cV, \end{aligned} \tag{4.1}$$

where $T(\tau)$ is the number of susceptible target cells, $I(\tau)$ is the number of infected target cells, $V(\tau)$ $TCID_{50}$/ml is the quantity of free virus, and τ days is the time since infection.

The model has previously been parametrised [14] for influenza A infection in humans (Table 1).

Table 1. Estimated parameter values and initial conditions for the TCL within-host model [14].

Parameter	Definition	Value
β	Infection rate of susceptible cells by virus	2.7×10^{-5} $(TCID_{50}/ml)^{-1}$ day^{-1}
δ	Death rate of infected cells	4.0 day^{-1}
p	Viral shedding rate by infected cells	1.2×10^{-2} $(TCID_{50}/ml)$ day^{-1}
c	Clearance rate of free virus	3.0 day^{-1}
$T(0)$	Initial number of susceptible cells	4×10^{8}
$I(0)$	Initial number of infected cells	0
$V(0)$	Initial quantity of free virus	9.3×10^{-2} $TCID_{50}/ml$

We used the TCL model to generate synthetic data from different patients. To incorporate variability between patients, we assumed that the parameters δ and $V(0)$ in the TCL model vary between individuals. This represents variation in the strength of the immune response and in the initial viral load. For each patient, $\log_{10}(\delta)$ was sampled from a normal distribution with mean 0.60 $\log_{10}(day^{-1})$ and standard deviation 0.25 $\log_{10}(day^{-1})$, and $\log_{10}(V(0))$ was sampled from a normal distribution with mean −1.03 $\log_{10}(TCID_{50}/ml)$ and standard deviation 1.12 $\log_{10}(TCID_{50}/ml)$. These values were chosen to match variability in previous individual parameter estimates [14], while the lognormal distribution was used to guarantee positivity. All other parameters were fixed at the values given in Table 1.

We considered analyses in which viral load was assumed to be observed exactly and continuously throughout infection, as well as analyses in which measurements of the viral load were recorded once daily for one week after infection. In the latter case, we incorporated measurement error by applying a normally distributed random variate with standard deviation σ to the logarithm of each measurement. We fitted the TCL model to the daily data from each patient using least squares estimation – in particular, the values of the parameters δ and $V(0)$ were chosen to minimise the sum of squares distance between the logarithm of the viral load in the model and in the data, while all other parameter values were assumed to be known exactly and were fixed at the values in Table 1. To avoid unrealistically large estimates of the initial viral load, we imposed $V(0) \leq 10^3$ $TCID_{50}$/ml when we fitted the parameters. An example of synthetic data generated for a single host, in addition to the fitted TCL model, is given in figure 4*a*.

The SI$_n$R model

The population-scale SI$_n$R model [28,63] of pathogen transmission in a population of N hosts is given by

$$\begin{aligned}
\frac{dS}{dt} &= -S\sum_{j=1}^{n}\beta_j I_j\,,\\
\frac{dI_1}{dt} &= S\sum_{j=1}^{n}\beta_j I_j - \mu_1 I_1,\\
\frac{dI_i}{dt} &= \mu_{i-1}I_{i-1} - \mu_i I_i, \qquad for\ i = 2,\dots,n,\\
\frac{dR}{dt} &= \mu_n I_n,
\end{aligned} \tag{4.2}$$

where $S(t)$ is the number of susceptible individuals, $I_i(t)$ is the number of individuals in the i^{th} infected compartment, and t days is the time since the start of the outbreak. Individuals in the i^{th} infected compartment infect susceptible hosts at total rate $\beta_i I_i S$ per day, and progress to the next infected compartment (or recover, if $i = n$) at total rate $\mu_i I_i$ per day. The basic reproduction number of this model is [63]

$$R_0 = N \sum_{i=1}^{n} \frac{\beta_i}{\mu_i}. \quad (4.3)$$

From within- to between-host

We used both an existing IDE approach (steps A-C below) and a new compartmental framework (steps A-B and D-E below) to transition from patient-level to population-scale dynamics. The two methods are outlined below, and a schematic is shown in figure 1.

A. Fit a within-host model to longitudinally sampled data on patient-level dynamics, to estimate the pathogen load of each individual patient at every time since infection.
B. Estimate the expected infectiousness curve, $\beta(\tau)$, at each time since infection, τ days, by assuming that the infectiousness of each host depends on the pathogen load according to a pre-specified relationship between these quantities.

Then either C:

C. Solve the K&M IDE model, with infectiousness curve $\beta(\tau)$, to calculate the population-scale dynamics (details of the K&M model are given in Section S1).

Or D-E:

D. Parameterise the SI_nR model: choose the number of infected compartments, *n*, where *n* is assumed to be large. Then find *T* such that $\beta(\tau)$ is zero or very small for τ > *T* days, and choose the parameters in the SI_nR model to be

$$\mu_i = \frac{n}{T},$$

$$\beta_i = \frac{n}{T}\int_{(i-1)T/n}^{iT/n} \beta(\tau)d\tau\,, \qquad \text{for } i = 1, \dots, n-1,$$

$$\beta_n = \frac{n}{T}\int_{(n-1)T/n}^{\infty} \beta(\tau)d\tau. \tag{4.4}$$

Explanation of these parameter choices is given in Section S2.

E. Solve the SI_nR model numerically to approximate the population-scale dynamics.

In most of our analyses, we assumed a linear relationship between the viral load and infectiousness of each influenza-infected host, although two alternative possibilities are considered in Section S8. In particular, in our main analyses we assumed that

$$\beta^{(i)}(\tau) = kV^{(i)}(\tau), \tag{4.5}$$

for constant *k,* where *i* represents the particular host under consideration. Therefore, the expected infectiousness, $\beta(\tau)$, was given in terms of the expected viral load, $V(\tau)$, of a host at time τ days since infection (calculated over a large number of realisations of the within-host model), by

$$\beta(\tau) = kV(\tau). \tag{4.6}$$

We fixed the constant *k* by assuming that the basic reproduction number,

$$R_0 = N\int_0^{\infty} \beta(\tau)\, d\tau, \tag{4.7}$$

was known. In our main analyses, we fixed R_0 = 1.5, which is consistent with estimates for influenza A infection [8] (different values of R_0 are considered in Section S9). The expected infectiousness could therefore be calculated using the formula

$$\beta(\tau) = \frac{R_0}{N \int_0^\infty V(x)dx} V(\tau). \quad (4.8)$$

To calculate the "true" expected infectiousness curve, $\beta(\tau)$, we computed the expected viral load over 10,000 realisations of the within-host model. We also considered analyses in which data were only available from a smaller number of patients, *d*. In such cases, we simulated the within-host model *d* times to calculate the exact patient-level dynamics corresponding to each patient, and used the data to estimate first $V(\tau)$ and then $\beta(\tau)$. In analyses where we also incorporated measurement error, we used the patient-level dynamics estimated by fitting the within-host model to daily observations of the viral load for each patient, in order to estimate $\beta(\tau)$.

Both the compartmental and IDE methods were then used to predict the population-scale dynamics. To parameterise the SI_nR model, we took T = 7 days, since the expected infectiousness was found to be very small after a week since infection. We considered a population of size N = 1000, and assumed that there was initially a single newly infected individual, with all others susceptible. These initial conditions were implemented in the SI_nR model by taking $I_1(0) = 1$ and $S(0) = 999$, with all other compartments containing zero hosts initially.

Errors at patient-level and population-scale

We defined error metrics in order to quantify the errors that arise in the patient-level dynamics and in the population-scale dynamics. These were defined as proportional errors, so as to enable comparison between errors at the different scales.

First, we defined the within-host error, E_{wh}, to be the difference between the exact and approximate infectiousness curves, integrated over the entire course of infection, as a proportion of the area of the exact infectiousness curve. Therefore,

$$E_{wh} = \frac{\int_0^\infty \left|\beta_{\text{approx}}(\tau) - \beta_{\text{exact}}(\tau)\right| d\tau}{\int_0^\infty \beta_{\text{exact}}(\tau) d\tau}, \tag{4.9}$$

where $\beta_{\text{exact}}(\tau)$ and $\beta_{\text{approx}}(\tau)$ are the exact and approximate infectiousness curves, respectively.

Similarly, if $S_{\text{exact}}(t)$ and $S_{\text{approx}}(t)$ are the exact and approximate numbers of susceptible individuals at time t days since the start of the epidemic, then we defined the between-host error, E_{bh}, in terms of the rate of new cases per day throughout the epidemic, i.e.

$$E_{bh} = \frac{\int_0^\infty \left|\dot{S}_{\text{approx}}(t) - \dot{S}_{\text{exact}}(t)\right| dt}{\int_0^\infty -\dot{S}_{\text{exact}}(t) dt}, \tag{4.10}$$

where the dot denotes differentiation with respect to time.

AUTHOR CONTRIBUTIONS

RNT conceived the research; All authors designed the study; WSH carried out the research; RNT, PKM and CAY supervised the research; WSH and RNT drafted the manuscript; All authors revised the manuscript and gave final approval for publication.

ACKNOWLEDGEMENTS

We would like to thank members of the Wolfson Centre for Mathematical Biology at the University of Oxford for helpful discussions about this work. We would also like to thank Nik Cunniffe for suggestions about the multi-compartment SIR model, and the anonymous reviewers whose suggestions helped us to improve the manuscript.

DATA ACCESSIBILITY

All analyses were performed in MATLAB. Code is available for running the models, and is available at https://github.com/will-s-hart/WithinBetweenHostCompartmental. Our approach can also be recoded and adapted straightforwardly in other computing languages.

FUNDING

WSH was funded by an EPSRC Excellence Award for his doctoral studies. RNT was funded by a Junior Research Fellowship from Christ Church, Oxford. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Coburn BJ, Wagner BG, Blower S. 2009 Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Med.* **7**, 30.
2. Daszak P, Cunningham AA, Hyatt AD. 2000 Emerging infectious diseases of

662 wildlife---threats to biodiversity and human health. *Science* **287**, 443–449.
663 3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P.
664 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993.
665 4. Keeling MJ, Rohani P. 2011 *Modeling infectious diseases in humans and animals*.
666 Princeton: Princeton University Press. (doi:10.1016/s1473-3099(08)70147-6)
667 5. Morens DM, Folkers GK, Fauci AS. 2004 The challenge of emerging and re-
668 emerging infectious diseases. *Nature* **430**, 242–249.
669 6. Taylor LH, Latham SM, Woolhouse MEJ. 2001 Risk factors for human disease
670 emergence. *Philos. T. Roy. Soc. B* **356**, 983–989.
671 7. Thompson RN, Brooks-Pollock E. 2019 Detection, forecasting and control of
672 infectious disease epidemics: modelling outbreaks in humans, animals and plants.
673 *Philos. T. Roy. Soc. B* **374**.
674 8. Fraser C *et al.* 2009 Pandemic potential of a strain of influenza A (H1N1): early
675 findings. *Science* **324**, 1557–1561.
676 9. WHO Ebola Response Team. 2014 Ebola virus disease in West Africa - the first 9
677 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–95.
678 10. Bussell EH, Dangerfield CE, Gilligan CA, Cunniffe NJ. 2019 Applying optimal
679 control theory to complex epidemiological models to inform real-world disease
680 management. *Philos. T. Roy. Soc. B* **374**, 20180284.
681 11. Ferguson NM, Donnelly CA, Anderson RM. 2001 The foot-and-mouth epidemic in
682 Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–
683 1160.
684 12. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A,

685 Iamsirithaworn S, Burke DS. 2005 Strategies for containing an emerging influenza
686 pandemic in Southeast Asia. *Nature* **437**, 209–214.
687 13. Thompson RN, Gilligan CA, Cunniffe NJ. 2018 Control fast or control smart: when
688 should invading pathogens be controlled? *PLoS Comput. Biol.* **14**, 1–21.
689 14. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. 2006 Kinetics
690 of influenza A virus infection in humans. *J. Virol.* **80**, 7590–7599.
691 15. Canini L, Perelson AS. 2014 Viral kinetic modeling: state of the art. *J.*
692 *Pharmacokinet. Pharmacodyn.* **41**, 431–443.
693 16. Perelson AS. 2002 Modelling viral and immune system dynamics. *Nat. Rev.*
694 *Immunol.* **2**, 28–36.
695 17. Zitzmann C, Kaderali L. 2018 Mathematical analysis of viral replication dynamics
696 and antiviral treatment strategies: from basic models to age-based multi-scale
697 modeling. *Front. Microbiol.* **9**, 1546.
698 18. Handel A, Rohani P. 2015 Crossing the scale from within-host infection dynamics
699 to between-host transmission fitness: a discussion of current assumptions and
700 knowledge. *Philos. T. Roy. Soc. B* **370**.
701 19. Mideo N, Alizon S, Day T. 2008 Linking within- and between-host dynamics in the
702 evolutionary epidemiology of infectious diseases. *Trends Ecol. Evol.* **23**, 511–517.
703 20. Hart WS, Hochfilzer LFR, Cunniffe NJ, Lee H, Nishiura H, Thompson RN. 2019
704 Accurate forecasts of the effectiveness of interventions against Ebola may require
705 models that account for variations in symptoms during infection. *Epidemics* **29**,
706 100371.
707 21. Almocera AES, Nguyen VK, Hernandez-Vargas EA. 2018 Multiscale model

within-host and between-host for viral infectious diseases. *J. Math. Biol.* **77**, 1035–1057.

22. Boldin B, Diekmann O. 2008 Superinfections can induce evolutionarily stable coexistence of pathogens. *J. Math. Biol.* **56**, 635–672.

23. Lukens S *et al.* 2014 A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC Public Health* **14**, 1019.

24. Lythgoe KA, Pellis L, Fraser C. 2013 Is HIV short-sighted? Insights from a multistrain nested model. *Evolution* **67**, 2769–2782.

25. Nguyen VK, Mikolajczyk R, Hernandez-Vargas EA. 2018 High-resolution epidemic simulation using within-host infection and contact data. *BMC Public Health* **18**, 1–11.

26. Yamin D, Gertler S, Ndeffo-Mbah ML, Skrip LA, Fallah M, Nyenswah TG, Altice FL, Galvani AP. 2015 Effect of Ebola progression on transmission and control in Liberia. *Ann. Intern. Med.* **162**, 11–17.

27. Coombs D, Gilchrist MA, Ball CL. 2007 Evaluating the importance of within- and between-host selection pressures on the evolution of chronic pathogens. *Theor. Popul. Biol.* **72**, 576–591.

28. 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.

29. Childs LM, El Moustaid F, Gajewski Z, Kadelka S, Nikin-Beers R, Smith, Jr JW, Walker M, Johnson LR. 2019 Linked within-host and between-host models and data for infectious diseases: a systematic review. *PeerJ* **7**, e7057.

30. Day T, Alizon S, Mideo N. 2011 Bridging scales in the evolution of infectious disease life histories: theory. *Evolution* **65**, 3448–3461.
31. Gandolfi A, Pugliese A, Sinisgalli C. 2015 Epidemic dynamics and host immune response: a nested approach. *J. Math. Biol.* **70**, 399–435.
32. Gilchrist MA, Coombs D. 2006 Evolution of virulence: interdependence, constraints, and selection using nested models. *Theor. Popul. Biol.* **69**, 145–153.
33. Handel A, Brown J, Stallknecht D, Rohani P. 2013 A multi-scale analysis of influenza A virus fitness trade-offs due to temperature-dependent virus persistence. *PLoS Comput. Biol.* **9**.
34. Legros M, Bonhoeffer S. 2016 A combined within-host and between-hosts modelling framework for the evolution of resistance to antimalarial drugs. *J. R. Soc. Interface* **13**, 20160148.
35. Murillo LN, Murillo MS, Perelson AS. 2013 Towards multiscale modeling of influenza infection. *J. Theor. Biol.* **332**, 267–290.
36. Alizon S, Van Baalen M. 2008 Acute or chronic? Within-host models with immune dynamics, infection outcome, and parasite evolution. *Am. Nat.* **172**, E244–E256.
37. Magal P, McCluskey CC, Webb GF. 2010 Lyapunov functional and global asymptotic stability for an infection-age model. *Appl. Anal.* **89**, 1109–1140.
38. Chao DL, Halloran ME, Obenchain VJ, Longini IM. 2010 FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput. Biol.* **6**.
39. Canini L, Carrat F. 2011 Population modeling of influenza A/H1N1 virus kinetics and symptom dynamics. *J. Virol.* **85**, 2764–2770.
40. Vegvari C, Hadjichrysanthou C, Cauët E, Lawrence E, Cori A, De Wolf F,

754 Anderson RM. 2016 How can viral dynamics models inform endpoint measures in
755 clinical trials of therapies for acute viral infections? *PLoS One* **11**, 1–13.
756 41. Nguyen VK, Hernandez-Vargas EA. 2017 Windows of opportunity for Ebola virus
757 infection treatment and vaccination. *Sci. Rep.* **7**, 8975.
758 42. Nguyen VK, Klawonn F, Mikolajczyk R, Hernandez-Vargas EA. 2016 Analysis of
759 practical identifiability of a viral infection model. *PLoS One* **11**, 1–16.
760 43. Frost S. In press. epirecipes. See http://epirecip.es/epicookbook (accessed on 11
761 September 2019).
762 44. Jenness SM, Goodreau SM, Morris M. 2018 Epimodel: An R package for
763 mathematical modeling of infectious disease over networks. *J. Stat. Softw.* **84**, 1–
764 47.
765 45. Roberts M, Andreasen V, Lloyd A, Pellis L. 2015 Nine challenges for deterministic
766 epidemic models. *Epidemics* **10**, 49–53.
767 46. Murphy BR *et al.* 1980 Evaluation of influenza A/Hong Kong/123/77 (H1N1) ts-
768 1A2 and cold-adapted recombinant viruses in seronegative adult volunteers.
769 *Infect. Immun.* **29**, 348–355.
770 47. Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of
771 epidemics. *P. Roy. Soc. A* **115**, 700–721.
772 48. Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, Valleron
773 AJ. 2008 Time lines of infection and disease in human influenza: a review of
774 volunteer challenge studies. *Am. J. Epidemiol.* **167**, 775–785.
775 49. Lloyd AL. 2009 Sensitivity of model-based epidemiological parameter estimation
776 to model assumptions. In *Mathematical and Statistical Estimation Approaches in*

Epidemiology (eds G Chowell, JM Hyman, LMA Bettencourt, C Castillo-Chavez), pp. 123–141. Dordrecht: Springer Netherlands.

50. Mitchell L, Ross J V. 2016 A data-driven model for influenza transmission incorporating media effects. *Roy. Soc. Open Sci.* **3**.

51. Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the management of infectious diseases. *PLoS Med.* **2**, 0621–0627.

52. Blythe SP, Anderson RM. 1988 Distributed incubation and infectious periods in models of the transmission dynamics of the Human Immunodeficiency Virus (HIV). *Math. Med. Biol.* **5**, 1–19.

53. Lloyd AL. 2001 Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *P. Roy. Soc. B* **268**, 985–993.

54. Hethcote HW, Van Ark JW, Longini IM. 1991 A simulation model of AIDS in San Francisco: I. Model formulation and parameter estimation. *Math. Biosci.* **106**, 203–222.

55. Christofferson RC, Mores CN, Wearing HJ. 2014 Characterizing the likelihood of dengue emergence and detection in naïve populations. *Parasites and Vectors* **7**, 282.

56. Beauchemin CAA, Handel A. 2011 A review of mathematical models of influenza A infections within a host or cell culture: lessons learned and challenges ahead. *BMC Public Health* **11**, S7.

57. Nishiura H, Eichner M. 2007 Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiol. Infect.* **135**, 1145–1150.

58. Best K, Guedj J, Madelain V, de Lamballerie X, Lim S-Y, Osuna CE, Whitney JB, Perelson AS. 2017 Zika plasma viral dynamics in nonhuman primates provides insights into early infection and antiviral strategies. *P. Natl. Acad. Sci. USA* **114**, 8847–8852.

59. Thompson RN, Wymant C, Spriggs RA, Raghwani J, Fraser C, Lythgoe KA. 2019 Link between the numbers of particles and variants founding new HIV-1 infections depends on the timing of transmission. *Virus Evol.* **5**.

60. Fraser C. 2007 Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One* **2**.

61. Hernandez-Vargas EA *et al.* 2014 Effects of aging on influenza virus infection dynamics. *J. Virol.* **88**, 4123–4131.

62. Smith AM, Perelson AS. 2011 Influenza A virus infection kinetics: quantitative data and models. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **3**, 429–445.

63. Allen LJS, Lahodny GE. 2012 Extinction thresholds in deterministic and stochastic epidemic models. *J. Biol. Dyn.* **6**, 590–611.