1   **A theoretical framework for transitioning from patient-level to population-scale**

2   **epidemiological dynamics: influenza A as a case study**

3

4   **AUTHORS**

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

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

7

8   **AFFILIATIONS**

9   [1]Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford,

10   Woodstock Road, Oxford OX2 6GG, UK

11   [2]Centre for Mathematical Biology, University of Bath, Claverton Down, Bath BA2 7AY, UK

12   [3]Christ Church, University of Oxford, St Aldates, Oxford OX1 1DP, UK

13

14

15

16

17

18

19

20

21

22

23

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.

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

70   them challenging to implement other than by highly specialist mathematical modellers.

71   Such frameworks have often employed integro-differential equations (IDEs)

72   [19,24,27,30,31,33,35,37], although alternatives such as individual-based stochastic

73   models [12,23,25,26,38] have also been considered. IDEs are challenging to solve,

74   requiring bespoke numerical methods [28]. Some studies using IDEs have involved

75   explicit simulation of the full multi-scale model [24,31,37]. However, others have either

76   only used the multi-scale framework to derive quantities such as the basic reproduction

77   number of the pathogen rather than predicting temporal epidemic dynamics [30,33], or

78   have made simplifying assumptions such as taking a within-host model to be in

79   equilibrium [21,22,27,32]. Although an assumption that the pathogen load in each

80   infected host is not changing (or changes only a limited number of times) might be

81   appropriate for chronic infections, it leads to an approximate population-scale model

82   that does not explicitly account for time-dependent infectiousness or other potentially

83   complex patient-level dynamics.

84

85   In most previous studies that have used IDEs to transition from within- to between-host,

86   the progression of infection through all patients has been assumed to be identical [35].

87   Patient-level dynamics are therefore characterised by a within-host model in which the

88   values of model parameters (describing factors such as pathogen replication as well as

89   immune responses) are the same for all patients. These parameters have either

90   assumed values that have not been derived rigorously from data [29], or have been

91   obtained by fitting the model to data collected in longitudinal cohort studies from a small

92   number of patients [27,33]. Within-host parameters are, in fact, likely to vary between

93   individuals [14,39,40], for example due to differences in immune responses [41], while

94   measurement error may also lead to inaccurate parameter estimates particularly given

95   limited numbers of observations [35,42]. As we show, if patient-level data are only

96   available from a limited number of patients, then predictions of population-scale

97   epidemic dynamics may be inaccurate.

98

99   In this paper, we introduce a novel framework for transitioning from within- to between-

100 host epidemiological dynamics straightforwardly. Our method involves using a

101 compartmental model with a large number of infected compartments to predict the

102 population-scale dynamics. Compartmental models, comprising systems of ordinary

103 differential equations (ODEs), can be solved easily using standard numerical routines

104 and software packages [4,28,43,44], are straightforward to adapt to include further

105 biological detail [4,20,28], and are widely used for epidemic modelling [4,45]. We show

106 rigorously that our modelling framework is equivalent to a more complex IDE approach,

107 in the large-compartment limit of our method. Since the number of compartments is

108 simply a choice for the user to make, our easy-to-use method can generate results that

109 are as accurate as those from more complex approaches.

110

111 To demonstrate our framework, we consider modelling an outbreak due to the influenza

112 A virus. We use a previously parameterised within-host model [14] to generate a

113 synthetic dataset representative of real patient data (figure S1), incorporating variability

114 in the viral load time series between patients due to factors such as differences in

115 immune responses. Since the magnitude of this variability has been chosen to match

116 data from a previous cohort study [14,46], our dataset is comparable to obtaining data

117 from a cohort study, but with the advantage that we can test our approach using many

118 different possible cohorts of any size (from small cohort sizes up to very large cohort

119 sizes that generate idealised data). We explore the effects of both the number of

120 patients from which data are available, and the extent of measurement error in patient

121 data, on population-scale predictions. Our work therefore provides insight into the data

122 required for accurate forecasting using multi-scale epidemic models, as well as an

123 accessible modelling framework that can be used for forecasting during future

124 epidemics of a range of infectious diseases.

125

126 <div align="center">**2. RESULTS**</div>

127 **Transitioning from within- to between-host influenza dynamics**

128

129 We have developed a new compartmental framework for transitioning from within- to

130 between-host epidemic dynamics. In our approach, a within-host model is fitted to data

131 from individual patients, to estimate the pathogen load of each measured patient at

132 every time since infection (figure 1*a*). As with other multi-scale epidemic models

133 [18,19,35], by assuming a functional relationship between pathogen load and

134 infectiousness, the expected infectiousness, $\beta(\tau)$, of any host is then estimated at each

135 time since infection, τ days (figure 1*b*). We will call $\beta(\tau)$ the *expected infectiousness*

136 *curve*. In previous approaches, the Kermack and McKendrick (K&M) IDE model [47]

137 (see Section S1 of the Supplementary Material) has then been used to calculate the

138 population-scale dynamics (figure 1*c*). We instead use the expected infectiousness

<div align="center">6</div>

139   curve to parameterise a multi-stage compartmental model with a large number of

140   infected compartments (figure 1*d*), which can also be used to predict the population-

141   scale dynamics (figure 1*e*). For details on the compartmental and IDE approaches, see

142   Methods. In the limit of infinitely many compartments in our framework, the two

143   approaches are mathematically equivalent (we prove this rigorously in Section S2).

144
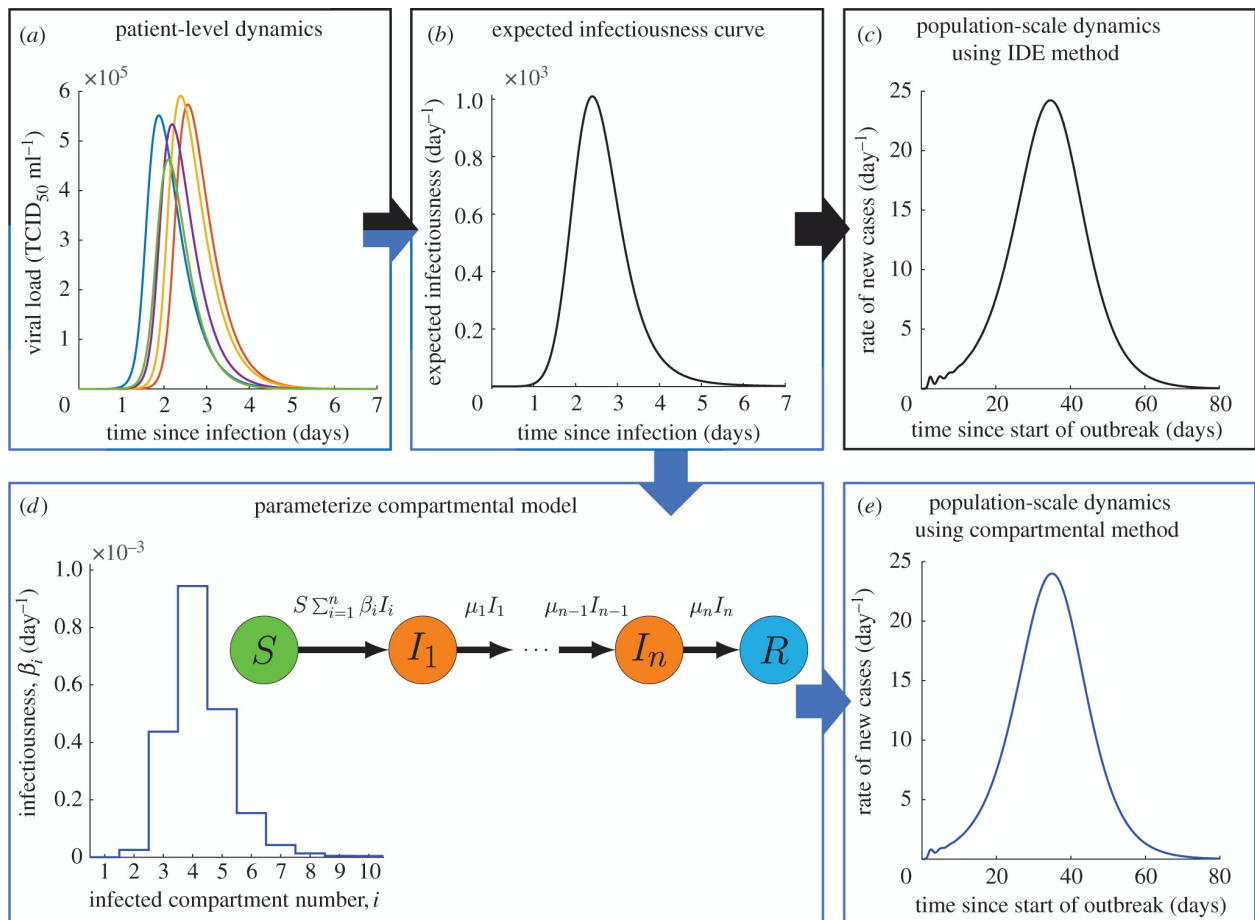
146   **Figure 1. Schematic demonstrating methods for transitioning from patient-level to population-**

147   **scale epidemiological dynamics.** In standard approaches (*a-c*), measurements from within each patient

148   (*a*) are used to parameterise a within-host model, giving rise to an averaged infectiousness curve (*b*).

149   These patient-level dynamics can then be nested in an IDE model (e.g. the K&M model [47]) used to

150   predict the population-scale dynamics (*c*). However, IDE models are challenging to solve. In contrast, in

151   our approach (*a,b,d,e*) the expected infectiousness curve is instead used to parameterise a

152   compartmental model (*d*) that can be used to predict population-scale dynamics straightforwardly (*e*).

153   Early-epidemic oscillations in panels (*c*) and (*e*) occur because the expected infectiousness of an infected

154   host is close to zero in the first day of infection, leading to delays before successive generations of newly

155   infected hosts begin to transmit the pathogen.

156

157   To illustrate our framework in a concrete setting, we considered the specific case of

158   influenza A infection in humans. We used the target cell-limited (TCL) within-host

159   model, which has previously been fitted to data from a cohort study of influenza

160   infection [14], to generate synthetic data from a large number of patients (see Methods).

161   The data were used to calculate the expected infectiousness curve (figure 2*a*) under the

162   assumptions of a linear relationship between viral load and infectiousness

163   [18,33,35,38,39] and a basic reproduction number (defined to be the expected number

164   of secondary cases arising from a single infected host in an otherwise entirely

165   susceptible population [4]) of 1.5 [8] (see Methods), although we consider other

166   assumptions and values of the reproduction number later (Sections S8 and S9).

167

168   Both our compartmental approach using a large number of infected compartments (*n* =

169   1000) and the previously used IDE method (i.e. the K&M model) were then used to

170   predict the population-scale outbreak dynamics, initially assuming the expected

171   infectiousness curve was known exactly (figure 2*b*). We considered a population of size

172   *N* = 1000 and assumed a single newly infected individual was introduced into an entirely

173   susceptible population. The two approaches produced almost identical results – the

174   error in the predicted population-scale dynamics when the compartmental method was

8

175   used, calculated as a proportional error relative to the dynamics predicted using the IDE

176   method (see Methods), was only 0.2%. We explored how many compartments are

177   required in our framework to ensure accurate population-scale forecasts, finding that in

178   general, the error in predictions scales with $1/n$ as the number of compartments, n,

179   becomes large (Section S5). When the infectiousness curve shown in figure 2a was

180   used to transition to population-scale dynamics, we found that $n = 24$ compartments are

181   sufficient for an error in population-scale predictions of 10% or less (figure S2b).

182
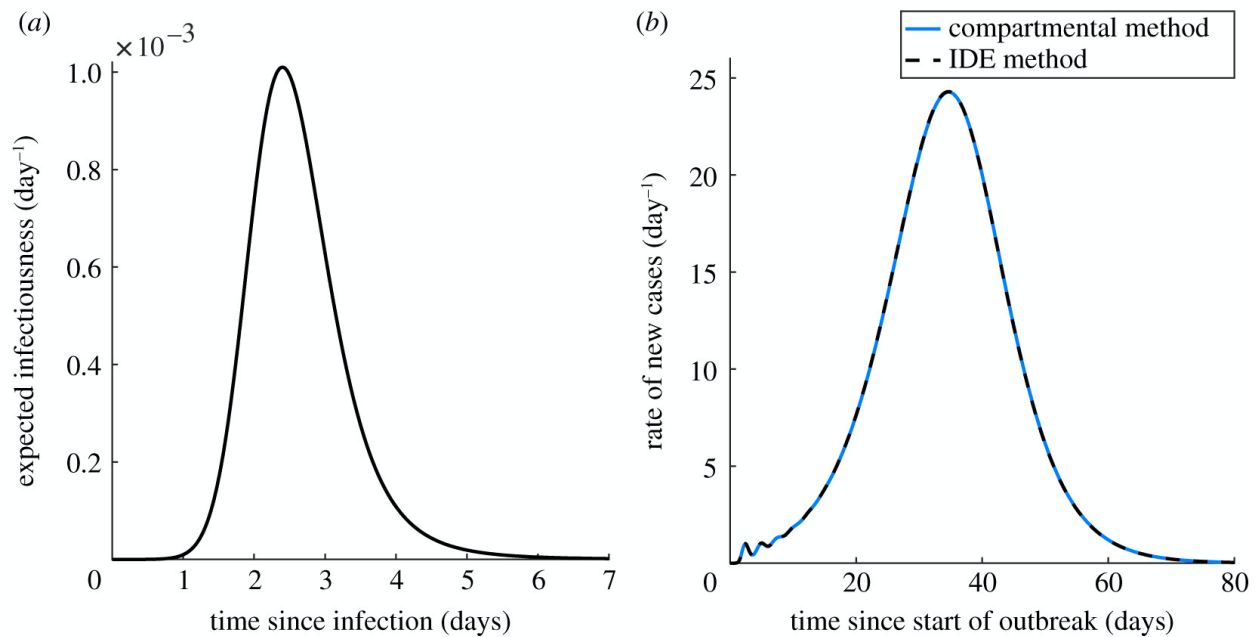
183

184   **Figure 2. Transitioning from within- to between-host influenza dynamics using the compartmental**

185   **and IDE methods.** (a) The expected infectiousness curve, $\beta(\tau)$, when the patient-level dynamics are

186   perfectly characterised. (b) The population-scale dynamics, using our compartmental approach with $n =$

187   1000 infected compartments (blue), and using the IDE method (black dashed), for the infectiousness

188   curve shown in panel (a).

189

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 3a) and calculated the approximate

population-scale dynamics using our compartmental framework with $n = 1000$ infected

compartments (figure 3b). For a fixed cohort size, d, significantly different predictions of

213   population-scale dynamics are possible, depending on which patients are included in

214   the study (figure 3*c*). Therefore, we calculated the distributions of errors in both the

215   patient-level and population-scale dynamics, relative to the case in which the patient-

216   level dynamics were perfectly characterised, over 5000 repeats for each of a range of

217   patient cohort sizes, *d* (figure 3*d*). Equivalent results using the IDE method rather than

218   our compartmental approach are shown in figure S3*a*.

219
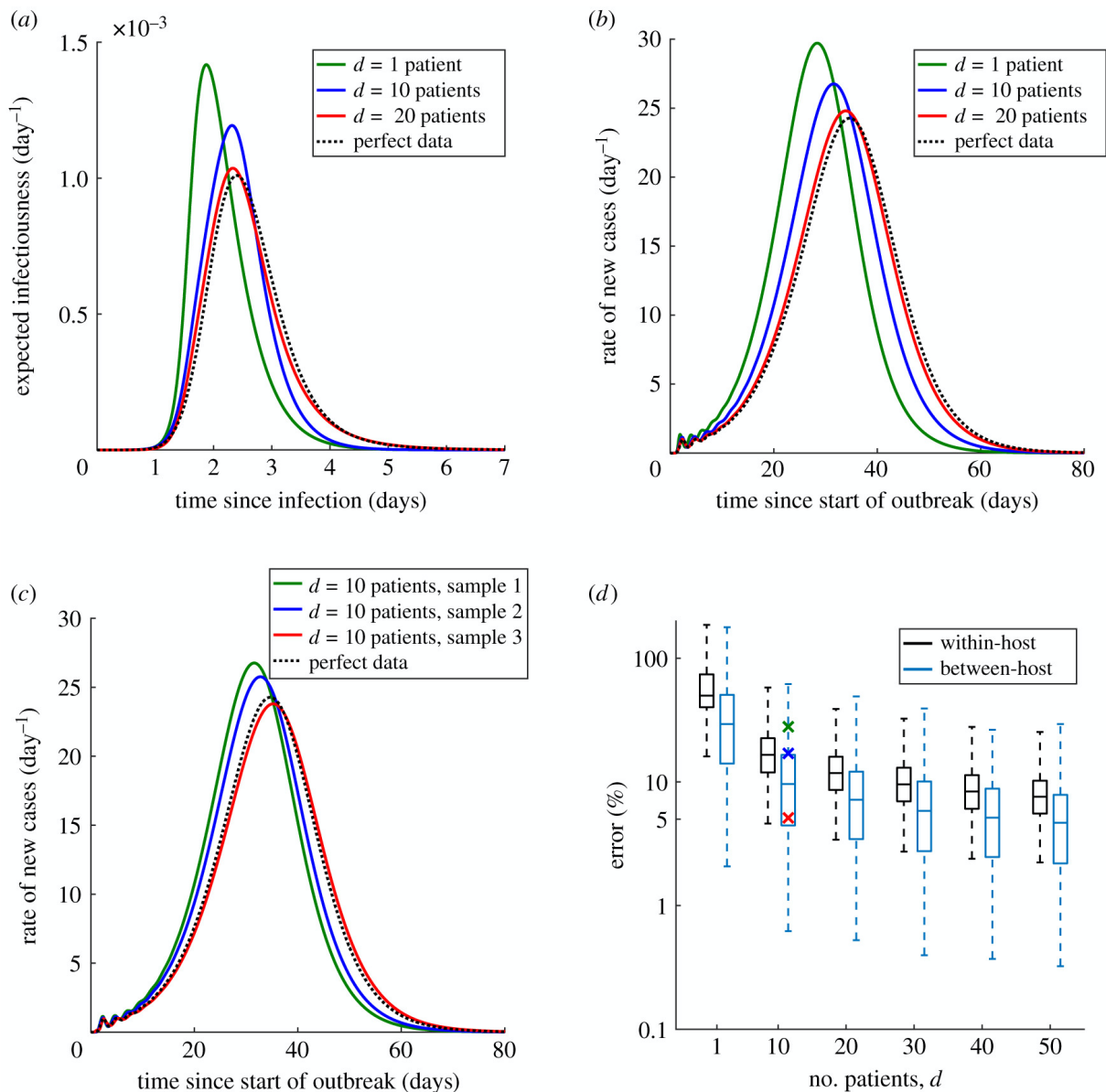
220

221   **Figure 3. How many patients need to supply data for accurate population-scale predictions?** (*a*)

222   Examples of approximate expected infectiousness curves when (exact and continuous) data are available

223   from $d = 1$ (green), $d = 10$ (blue) or $d = 20$ (red) randomly chosen patients, and the expected

224   infectiousness curve when the patient-level dynamics are perfectly characterised (black dotted). (*b*) The

225   predicted population-scale dynamics for each infectiousness curve in panel (*a*), using $n = 1000$

226   compartments in our framework. (*c*) Three examples of possible approximate population-scale dynamics,

227   when data are available from $d = 10$ patients. (*d*) Box-and-whisker plots indicating the distributions of

228   within-host (black) and between-host (blue) errors for different groups of patients randomly chosen in the

229   study cohort, for a range of values of the number of patients, d. The boxes indicate the lower quartile,

230   median and upper quartile, and the maximum length of each whisker is 1.5 times the interquartile range.

231   The crosses represent the between-host errors corresponding to the curves of the same colour in panel

232   (*c*) (these are at values of 28%, 17% and 5% error).

233

234   As the number of patients is increased, the errors at patient-level and population-scale

235   both decrease in general (depending on precisely which patients are included in the

236   study cohort), but at a decreasing rate. The magnitude of the population-scale error is

237   generally smaller than that of the patient-level error. Therefore, limited data do not

238   necessarily preclude accurate population-scale predictions, even when there is a large

239   amount of variability between different patients. In this case – when there are exact and

240   continuous data available from each patient – a cohort size of $d = 20$ patients is

241   sufficient for the between-host error to be 10% or less on average (figure 3*d*). However,

242   since the errors are affected by the precise patients included in the cohort, more

243   patients are required for a greater certainty of a small between-host error. For example,

244    $d = 30$ patients are required to ensure that the upper quartile of between-host errors for

245   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

269   between-host errors that arose directly due to measurement error, by taking the "true"

270   dynamics to be those when exact and continuous data were available from the same 10

271   hosts. The distributions of these errors, each time calculated over 5000 repeats for a

272   range of values of σ, are shown in figure 4d (for equivalent results obtained using the

273   IDE method, see figure S3b).

274

275
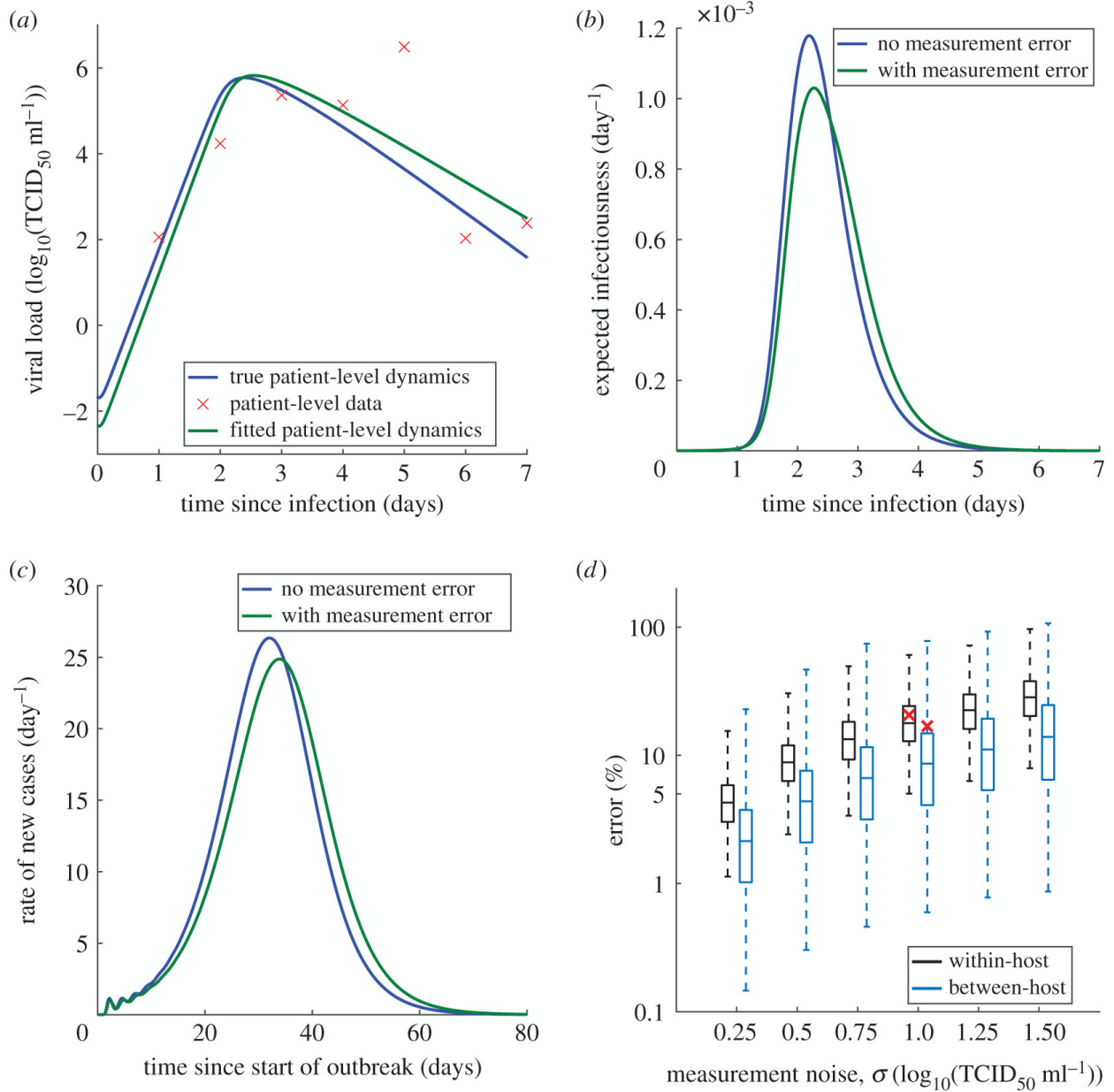
Figure 4. The effect on population-scale predictions of measurement error in patient-level data. (*a*) Example of synthetic data for a single patient: the true viral load of the patient against time since infection (blue), daily synthetic data with a measurement noise level of σ = 1 $\log_{10}$(TCID50/ml) (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 $\log_{10}$(TCID$_{50}$/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

301　would also occur if a small number of compartments is used in our multi-scale modelling

302　approach, although this can be avoided by simply choosing a large number of

303　compartments in the model. Nonetheless, we also conducted an analysis in which all

304　three potential sources of error were included: (i) number of patients; (ii) measurement

305　error; (iii) number of compartments.

306

307　When we investigated the combined effect of these potential sources of error, we

308　considered a measurement noise level of σ = 1 $\log_{10}(\text{TCID}_{50}/\text{ml})$, since this generated

309　synthetic data comparable to those recorded in cohort studies (figure S1). Assuming

310　that data were available from d randomly sampled patients, our compartmental

311　framework with n infected compartments was used to estimate the population-scale

312　dynamics. We repeated this analysis 10,000 times each for different pairs of values of d

313　and n, each time calculating the within-host and between-host errors, relative to the

314　case in which the patient-level dynamics were perfectly characterised and the IDE

315　method was used (equivalent to using infinitely many compartments in our

316　compartmental framework). The distributions of within-host and between-host errors,

317　when a large number of compartments (n = 1000) is used in our framework, are plotted

318　for different numbers of patients (d) in figure 5a. Equivalent results using the IDE

319　method are shown in figure S3c. When either the compartmental or IDE approach is

320　used, data from d = 40 patients are required for an average between-host error of 10%

321　or below (compared to 20 patients if data are recorded exactly, i.e. with no

322　measurement error – see figure 3d and figure S3a).

323

16
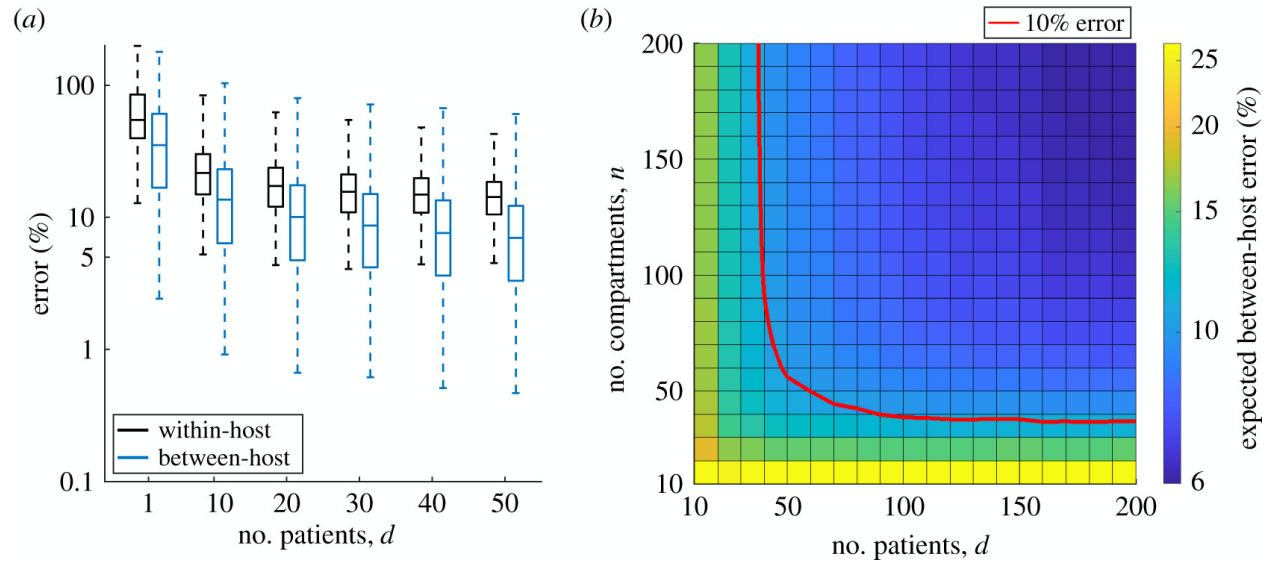
Figure 5. The effects of the number of patients and number of compartments on population-scale predictions. (*a*) Box-and-whisker plots indicating the distributions of within-host (black) and between-host (blue) errors for different patients chosen in the study cohort when n = 1000 compartments are used 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,

341 that the benefit of using more compartments becomes negligible when more than $n =$

342 100 compartments are included in our approach (figure 5*b*).

343

344 Whereas in figure 5 we assumed data were collected once daily for a week from each

345 patient, we also conducted supplementary analyses of the between-host error when

346 data were instead collected twice per day from each patient (Section S7) – in this case,

347 data are only required from 20 patients for a mean between-host error of 10% or less

348 (figure S4*b*). If instead the total number of measurements that can be taken is fixed,

349 then it might be necessary to choose between sampling a large number of patients

350 infrequently, or a small number of patients frequently. We explored this in Section S7,

351 and found that sampling patients more than twice per day tended to lead to less

352 accurate population-scale predictions when the total number of measurements was

353 fixed at values below 1000 (figure S5). For realistic cohort sizes, population-scale errors

354 were similar when data were collected either once daily from 2*d* patients or twice daily

355 from *d* patients (for example, when *d* = 20, the respective errors are both 10%).

356

357 We examined the robustness of our results to our assumptions when transitioning from

358 within- to between-host (Section S8), finding similar results to those shown in figure 5 in

359 two alternative cases, in which the infectiousness of each patient either scales with the

360 logarithm of their viral load (figure S6*c*) or saturates at high viral loads (figure S6*f*). In

361 addition, we considered the effect of the assumed value of the basic reproduction

362 number, R_0, on our results (Section S9), and also repeated our analyses in figure 5 for

363 different values of the measurement noise level, σ (Section S10), and for different levels

364   of variability in the within-host parameter values corresponding to different patients

365   (Section S11). When R_0, σ, or the level of variability in within-host parameter values,

366   exceeded the values considered in figure 5, a cohort size larger than 40 hosts was

367   found to be required to ensure a mean between-host error of 10% or below (figures S7-

368   S9) – for example, data from 70 patients are required if R_0 = 3 (figure S7i).

369

370   **3. DISCUSSION**

371   In this article, we have introduced a novel, easy-to-use, compartmental framework for

372   nesting patient-level data in population-scale epidemic models. In the large-

373   compartment limit, our method is mathematically equivalent to more complicated

374   approaches that involve IDEs (Section S2). However, our method has the advantage

375   that it can be used straightforwardly, allowing it to be applied widely in future. We have

376   provided adaptable computing code alongside this article to facilitate future use of our

377   approach (see Data Accessibility).

378

379   To illustrate our method, we considered the example of influenza A infection in humans.

380   A viral dynamics model [14] was used to generate a synthetic dataset describing

381   changing viral loads in a cohort of patients, which is representative of real patient data

382   (figure S1). We showed how our compartmental framework can be used to predict the

383   population-scale epidemic dynamics, and compared our predictions to forecasts using

384   the more complicated K&M IDE model. The population-scale predictions from our

385   framework closely matched those obtained using the IDE model, provided that a

19

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

409   infections (e.g. [14,39]). However, the accuracy of population-scale predictions depends

410   on the frequency with which data are collected (Section S7), so ensuring regular data

411   collection from each patient in future cohort studies is important for accurate population-

412   scale forecasting.

413

414   Our approach was motivated by earlier studies in which compartmental models with

415   multiple latent and infectious stages were employed so that the standard assumption of

416   exponentially distributed latent and infectious periods was relaxed [49–53]. The use of

417   multiple stages allows for gamma distributed latent and infectious periods (the so-called

418   "linear chain trick" [52] or "method of stages" [53]), and gamma distributions have been

419   shown to characterise epidemiological periods accurately [49,51]. However, in those

420   studies [49–53], the level of infectiousness is assumed constant throughout the

421   infectious period. We were therefore also inspired by previous research in which time-

422   dependent infectiousness was incorporated into multi-stage compartmental models, in

423   cases where the compartments correspond to clearly distinct phases of infection (for

424   example, studies of HIV [54] and Ebola [20]) or convenient time periods [55]. Our

425   approach is more similar to a method used to include experimental data in models of

426   plant disease [28], but differs from previous literature [20,28,54,55] due to our use of a

427   large number of infected compartments corresponding to different infection rates in

428   order to provide an easy method to transition from within- to between-host that is

429   accurate for any patient-level infection dynamics.

430

431   We focussed on the case study of influenza A in humans because compartmental

432   models are frequently used to model both patient-level and population-scale influenza

433   dynamics, while there has also been significant interest in developing models linking the

434   dynamics at the two scales [35]. In principle, however, our approach could be extended

435   to model outbreaks of a range of other pathogens for which patient-level dynamics are

436   well characterised. This would require careful consideration of the functional relationship

437   between pathogen load and infectiousness, since this is likely to differ between

438   pathogens [18]. In particular, the mode of transmission may be an important factor in

439   determining suitable relationships for different pathogens.

440

441   To describe individual patient-level influenza dynamics, we used the simple TCL within-

442   host model. More detailed within-host models exist, and involve features including a

443   delay before target cells begin to shed virus (an eclipse phase) [14] or explicit modelling

444   of innate and adaptive immune responses [56]. While the TCL model was sufficient to

445   demonstrate our approach here, the expected infectiousness curve in our framework

446   could be generated using a within-host model with any level of complexity. Alternatively,

447   if patient-level infection dynamics are not well characterised, then an expected

448   infectiousness curve that is estimated from transmission data [12,57], rather than within-

449   host data, could also be embedded within our framework.

450

451   In order to generate synthetic patient-level data, we assumed that two within-host

452   parameters varied between patients, using previous parameter estimates to determine

453   the level of parameter variability [14]. We incorporated measurement error by adding a

454   normally distributed random variate to daily observations of the logarithm of the viral

455   load (although we also considered other frequencies of data collection in Section S7).

456   Differences in both the extent of measurement error, and the extent of parameter

457   variability between patients, can lead to significant differences in population-scale errors

458   (figures S8 and S9). Therefore, when our modelling framework is used to determine

459   how many patients should be included in future cohort studies, careful consideration of

460   the measurement error and the variability in pathogen load time series between patients

461   is important.

462

463   The TCL model was fitted to the data from each patient using a basic least squares

464   estimation approach, since the precise method of parameter inference is not central to

465   our modelling framework. However, it would be straightforward to extend our approach

466   to consider different error structures and methods for fitting models to patient-level data.

467   In particular, a non-linear mixed effects modelling approach – amounting to a partial

468   pooling of the data between individuals – could be used. This would enable robust

469   parameter estimation in a real dataset, particularly in settings in which the numbers of

470   data points per patient are small, and both the frequency and timing of data collection

471   may vary between patients [58,59]. Going forward, we will use such a method to explore

472   further whether or not there is an optimal balance between the number of patients and

473   the frequency of measurements per patient, if total resources are limited (see Section

474   S7).

475

476   In our main analyses, we made the common assumption that the infectiousness of an

477   influenza-infected host is proportional to their viral load [18,33,35,38,39], although we

478   also obtained similar results in two alternative cases in which infectiousness either

479   scales with the logarithm of the viral load [33,38] or saturates at high pathogen loads

480   [21,55] (Section S8). However, more complex possibilities could easily be incorporated

481   into our framework. For example, future studies may also incorporate varying symptoms

482   during infection into our approach [20,23,39], in order to account for dependency of

483   transmissibility on behavioural factors in addition to pathogen load [18].

484

485   While we considered errors in population-scale predictions arising due to variability

486   between different infected patients when data are limited, our results were obtained

487   using a population-scale model in which the population was assumed to be

488   homogeneous and well-mixed. Variability between different patients was assumed to be

489   random, so that all infected hosts could effectively be assumed to follow the same

490   averaged infectiousness curve. In Section S3, we provide mathematical justification for

491   this averaging in the population-scale dynamics (see also [60]). We sought to develop

492   our framework for transitioning from within- to between-host using the simplest possible

493   population-scale model, but our compartmental approach could be generalised, for

494   example, to models incorporating age structure, spatial effects, social contact networks

495   or stochasticity [4]. In an age-structured model, different within-host parameter values

496   (or even different models) could be used to describe patient-level dynamics in the

497   different age groups, since there may be substantial differences in within-host dynamics

498   between patients of different ages [61].

499

500 In summary, we have introduced a novel compartmental framework for nesting patient

501 data in population-scale epidemiological models. We have demonstrated our easy-to-

502 use approach in the context of influenza. Not only can our modelling approach be used

503 to inform population-scale predictions with data from patients, but it can also be used to

504 design cohort studies by determining which data need to be collected. As a result, clear

505 communication between clinical epidemiologists who conduct cohort studies and

506 epidemiological modellers will allow for optimal study design. Including patient-level

507 dynamics in population-scale epidemiological models as proposed here has the

508 potential to improve epidemic forecasts; we hope that the simplicity of our approach will

509 facilitate its use for forecasting in a wide range of future outbreaks.

510

511   **4. METHODS**

512 **Within-host model**

513

514 The TCL model of viral dynamics, which has previously been used to model influenza

515 infections [14,15,62], is given by

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

516 where $T(\tau)$ is the number of susceptible target cells, $I(\tau)$ is the number of infected target

517 cells, $V(\tau)$ TCID$_{50}$/ml is the quantity of free virus, and τ days is the time since infection.

518   The model has previously been parametrised [14] for influenza A infection in humans

519   (Table 1).

520

521   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	$9.3 \times 10^{-2}\ TCID_{50}/ml$

522

523   We used the TCL model to generate synthetic data from different patients. To

524   incorporate variability between patients, we assumed that the parameters δ and $V(0)$ in

525   the TCL model vary between individuals. This represents variation in the strength of the

526   immune response and in the initial viral load. For each patient, $\log_{10}(\delta)$ was sampled

527   from a normal distribution with mean 0.60 $\log_{10}(day^{-1})$ and standard deviation 0.25

528     $\log_{10}(day^{-1})$, and $\log_{10}(V(0))$ was sampled from a normal distribution with mean −1.03

529     $\log_{10}(TCID_{50}/ml)$ and standard deviation 1.12 $\log_{10}(TCID_{50}/ml)$. These values were

530   chosen to match variability in previous individual parameter estimates [14], while the

531   lognormal distribution was used to guarantee positivity. All other parameters were fixed

532   at the values given in Table 1.

533

534 We considered analyses in which viral load was assumed to be observed exactly and

535 continuously throughout infection, as well as analyses in which measurements of the

536 viral load were recorded once daily for one week after infection. In the latter case, we

537 incorporated measurement error by applying a normally distributed random variate with

538 standard deviation σ to the logarithm of each measurement. We fitted the TCL model to

539 the daily data from each patient using least squares estimation – in particular, the

540 values of the parameters δ and $V(0)$ were chosen to minimise the sum of squares

541 distance between the logarithm of the viral load in the model and in the data, while all

542 other parameter values were assumed to be known exactly and were fixed at the values

543 in Table 1. To avoid unrealistically large estimates of the initial viral load, we imposed

544 $V(0) \leq 10^3$ TCID$_{50}$/ml when we fitted the parameters. An example of synthetic data

545 generated for a single host, in addition to the fitted TCL model, is given in figure 4*a*.

546

547 **The SI$_n$R model**

548

549 The population-scale SI$_n$R model [28,63] of pathogen transmission in a population of *N*

550 hosts is given by

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

551   where $S(t)$ is the number of susceptible individuals, $I_i(t)$ is the number of individuals in

552   the ith infected compartment, and t days is the time since the start of the outbreak.

553   Individuals in the ith infected compartment infect susceptible hosts at total rate $\beta_i I_i S$ per

554   day, and progress to the next infected compartment (or recover, if $i = n$) at total rate $\mu_i I_i$

555   per day. The basic reproduction number of this model is [63]

556
$$R_0 = N \sum_{i=1}^{n} \frac{\beta_i}{\mu_i}. \tag{4.3}$$

557

558   **From within- to between-host**

559

560   We used both an existing IDE approach (steps A-C below) and a new compartmental

561   framework (steps A-B and D-E below) to transition from patient-level to population-scale

562   dynamics. The two methods are outlined below, and a schematic is shown in figure 1.

563

564   A. Fit a within-host model to longitudinally sampled data on patient-level dynamics,

565   to estimate the pathogen load of each individual patient at every time since

566   infection.

567   B. Estimate the expected infectiousness curve, $\beta(\tau)$, at each time since infection, τ

568   days, by assuming that the infectiousness of each host depends on the pathogen

569   load according to a pre-specified relationship between these quantities.

570   Then either C:

571   C. Solve the K&M IDE model, with infectiousness curve $\beta(\tau)$, to calculate the

572   population-scale dynamics (details of the K&M model are given in Section S1).

28

573   Or D-E:

574   D. Parameterise the SI_nR model: choose the number of infected compartments, n,

575   where n is assumed to be large. Then find T such that $\beta(\tau)$ is zero or very small

576   for $\tau > T$ days, and choose the parameters in the SI_nR model to be

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

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

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

580   Explanation of these parameter choices is given in Section S2.

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

582

583   In most of our analyses, we assumed a linear relationship between the viral load and

584   infectiousness of each influenza-infected host, although two alternative possibilities are

585   considered in Section S8. In particular, in our main analyses we assumed that

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

587   for constant k, where i represents the particular host under consideration. Therefore, the

588   expected infectiousness, $\beta(\tau)$, was given in terms of the expected viral load, $V(\tau)$, of a

589   host at time τ days since infection (calculated over a large number of realisations of the

590   within-host model), by

591            $$\beta(\tau) = kV(\tau). \qquad (4.6)$$

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

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

29

594   was known. In our main analyses, we fixed R_0 = 1.5, which is consistent with estimates

595   for influenza A infection [8] (different values of R_0 are considered in Section S9). The

596   expected infectiousness could therefore be calculated using the formula

597
$$\beta(\tau) = \frac{R_0}{N \int_0^\infty V(x)dx} V(\tau). \tag{4.8}$$

598

599   To calculate the "true" expected infectiousness curve, $\beta(\tau)$, we computed the expected

600   viral load over 10,000 realisations of the within-host model. We also considered

601   analyses in which data were only available from a smaller number of patients, d. In such

602   cases, we simulated the within-host model d times to calculate the exact patient-level

603   dynamics corresponding to each patient, and used the data to estimate first $V(\tau)$ and

604   then $\beta(\tau)$. In analyses where we also incorporated measurement error, we used the

605   patient-level dynamics estimated by fitting the within-host model to daily observations of

606   the viral load for each patient, in order to estimate $\beta(\tau)$.

607

608   Both the compartmental and IDE methods were then used to predict the population-

609   scale dynamics. To parameterise the SI_nR model, we took T = 7 days, since the

610   expected infectiousness was found to be very small after a week since infection. We

611   considered a population of size N = 1000, and assumed that there was initially a single

612   newly infected individual, with all others susceptible. These initial conditions were

613   implemented in the SI_nR model by taking $I_1(0)$ = 1 and $S(0)$ = 999, with all other

614   compartments containing zero hosts initially.

615

616   **Errors at patient-level and population-scale**

617

618 We defined error metrics in order to quantify the errors that arise in the patient-level

619 dynamics and in the population-scale dynamics. These were defined as proportional

620 errors, so as to enable comparison between errors at the different scales.

621

622 First, we defined the within-host error, E_{wh}, to be the difference between the exact and

623 approximate infectiousness curves, integrated over the entire course of infection, as a

624 proportion of the area of the exact infectiousness curve. Therefore,

625
$$E_{wh} = \frac{\int_0^\infty \left| \beta_{\text{approx}}(\tau) - \beta_{\text{exact}}(\tau) \right| d\tau}{\int_0^\infty \beta_{\text{exact}}(\tau) d\tau}, \qquad (4.9)$$

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

627 respectively.

628

629 Similarly, if $S_{\text{exact}}(t)$ and $S_{\text{approx}}(t)$ are the exact and approximate numbers of susceptible

630 individuals at time t days since the start of the epidemic, then we defined the between-

631 host error, E_{bh}, in terms of the rate of new cases per day throughout the epidemic, i.e.

632
$$E_{bh} = \frac{\int_0^\infty \left| \dot{S}_{\text{approx}}(t) - \dot{S}_{\text{exact}}(t) \right| dt}{\int_0^\infty -\dot{S}_{\text{exact}}(t) dt}, \qquad (4.10)$$

633 where the dot denotes differentiation with respect to time.

634

635 **AUTHOR CONTRIBUTIONS**

636 RNT conceived the research; All authors designed the study; WSH carried out the

637 research; RNT, PKM and CAY supervised the research; WSH and RNT drafted the

638 manuscript; All authors revised the manuscript and gave final approval for publication.

639

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.

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

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

709   1035–1057.

710   22.   Boldin B, Diekmann O. 2008 Superinfections can induce evolutionarily stable

711   coexistence of pathogens. *J. Math. Biol.* **56**, 635–672.

712   23.   Lukens S *et al.* 2014 A large-scale immuno-epidemiological simulation of

713   influenza A epidemics. *BMC Public Health* **14**, 1019.

714   24.   Lythgoe KA, Pellis L, Fraser C. 2013 Is HIV short-sighted? Insights from a

715   multistrain nested model. *Evolution* **67**, 2769–2782.

716   25.   Nguyen VK, Mikolajczyk R, Hernandez-Vargas EA. 2018 High-resolution

717   epidemic simulation using within-host infection and contact data. *BMC Public*

718   *Health* **18**, 1–11.

719   26.   Yamin D, Gertler S, Ndeffo-Mbah ML, Skrip LA, Fallah M, Nyenswah TG, Altice

720   FL, Galvani AP. 2015 Effect of Ebola progression on transmission and control in

721   Liberia. *Ann. Intern. Med.* **162**, 11–17.

722   27.   Coombs D, Gilchrist MA, Ball CL. 2007 Evaluating the importance of within- and

723   between-host selection pressures on the evolution of chronic pathogens. *Theor.*

724   *Popul. Biol.* **72**, 576–591.

725   28.   Cunniffe NJ, Stutt ROJH, van den Bosch F, Gilligan CA. 2012 Time-dependent

726   infectivity and flexible latent and infectious periods in compartmental models of

727   plant disease. *Phytopathology* **102**, 365–380.

728   29.   Childs LM, El Moustaid F, Gajewski Z, Kadelka S, Nikin-Beers R, Smith, Jr JW,

729   Walker M, Johnson LR. 2019 Linked within-host and between-host models and

730   data for infectious diseases: a systematic review. *PeerJ* **7**, e7057.

731   30.   Day T, Alizon S, Mideo N. 2011 Bridging scales in the evolution of infectious

732   disease life histories: theory. *Evolution* **65**, 3448–3461.

733   31.   Gandolfi A, Pugliese A, Sinisgalli C. 2015 Epidemic dynamics and host immune

734   response: a nested approach. *J. Math. Biol.* **70**, 399–435.

735   32.   Gilchrist MA, Coombs D. 2006 Evolution of virulence: interdependence,

736   constraints, and selection using nested models. *Theor. Popul. Biol.* **69**, 145–153.

737   33.   Handel A, Brown J, Stallknecht D, Rohani P. 2013 A multi-scale analysis of

738   influenza A virus fitness trade-offs due to temperature-dependent virus

739   persistence. *PLoS Comput. Biol.* **9**.

740   34.   Legros M, Bonhoeffer S. 2016 A combined within-host and between-hosts

741   modelling framework for the evolution of resistance to antimalarial drugs. *J. R.*

742   *Soc. Interface* **13**, 20160148.

743   35.   Murillo LN, Murillo MS, Perelson AS. 2013 Towards multiscale modeling of

744   influenza infection. *J. Theor. Biol.* **332**, 267–290.

745   36.   Alizon S, Van Baalen M. 2008 Acute or chronic? Within-host models with immune

746   dynamics, infection outcome, and parasite evolution. *Am. Nat.* **172**, E244–E256.

747   37.   Magal P, McCluskey CC, Webb GF. 2010 Lyapunov functional and global

748   asymptotic stability for an infection-age model. *Appl. Anal.* **89**, 1109–1140.

749   38.   Chao DL, Halloran ME, Obenchain VJ, Longini IM. 2010 FluTE, a publicly

750   available stochastic influenza epidemic simulation model. *PLoS Comput. Biol.* **6**.

751   39.   Canini L, Carrat F. 2011 Population modeling of influenza A/H1N1 virus kinetics

752   and symptom dynamics. *J. Virol.* **85**, 2764–2770.

753   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*

777   *Epidemiology* (eds G Chowell, JM Hyman, LMA Bettencourt, C Castillo-Chavez),

778   pp. 123–141. Dordrecht: Springer Netherlands.

779   50.   Mitchell L, Ross J V. 2016 A data-driven model for influenza transmission

780   incorporating media effects. *Roy. Soc. Open Sci.* **3**.

781   51.   Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the

782   management of infectious diseases. *PLoS Med.* **2**, 0621–0627.

783   52.   Blythe SP, Anderson RM. 1988 Distributed incubation and infectious periods in

784   models of the transmission dynamics of the Human Immunodeficiency Virus

785   (HIV). *Math. Med. Biol.* **5**, 1–19.

786   53.   Lloyd AL. 2001 Destabilization of epidemic models with the inclusion of realistic

787   distributions of infectious periods. *P. Roy. Soc. B* **268**, 985–993.

788   54.   Hethcote HW, Van Ark JW, Longini IM. 1991 A simulation model of AIDS in San

789   Francisco: I. Model formulation and parameter estimation. *Math. Biosci.* **106**, 203–

790   222.

791   55.   Christofferson RC, Mores CN, Wearing HJ. 2014 Characterizing the likelihood of

792   dengue emergence and detection in naïve populations. *Parasites and Vectors* **7**,

793   282.

794   56.   Beauchemin CAA, Handel A. 2011 A review of mathematical models of influenza

795   A infections within a host or cell culture: lessons learned and challenges ahead.

796   *BMC Public Health* **11**, S7.

797   57.   Nishiura H, Eichner M. 2007 Infectiousness of smallpox relative to disease age:

798   estimates based on transmission network and incubation period. *Epidemiol.*

799   *Infect.* **135**, 1145–1150.

800 58. Best K, Guedj J, Madelain V, de Lamballerie X, Lim S-Y, Osuna CE, Whitney JB,

801   Perelson AS. 2017 Zika plasma viral dynamics in nonhuman primates provides

802   insights into early infection and antiviral strategies. *P. Natl. Acad. Sci. USA* **114**,

803   8847–8852.

804 59. Thompson RN, Wymant C, Spriggs RA, Raghwani J, Fraser C, Lythgoe KA. 2019

805   Link between the numbers of particles and variants founding new HIV-1 infections

806   depends on the timing of transmission. *Virus Evol.* **5**.

807 60. Fraser C. 2007 Estimating individual and household reproduction numbers in an

808   emerging epidemic. *PLoS One* **2**.

809 61. Hernandez-Vargas EA *et al.* 2014 Effects of aging on influenza virus infection

810   dynamics. *J. Virol.* **88**, 4123–4131.

811 62. Smith AM, Perelson AS. 2011 Influenza A virus infection kinetics: quantitative

812   data and models. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*

813   **3**, 429–445.

814 63. Allen LJS, Lahodny GE. 2012 Extinction thresholds in deterministic and stochastic

815   epidemic models. *J. Biol. Dyn.* **6**, 590–611.

816