1	A simulation-based approach for estimating the time-dependent reproduction number		
2	from temporally aggregated disease incidence time series data		
3	I Ogi-Gittins <sup>1,2</sup> , WS Hart <sup>3</sup> , J Song <sup>4</sup> , RK Nash <sup>5</sup> , J Polonsky <sup>6</sup> , A Cori <sup>5</sup> , EM Hill <sup>1,2</sup> , RN		
4	Thompson <sup>3</sup> *		
5	Affiliations:		
6	<sup>1</sup> Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK		
7	<sup>2</sup> Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research		
8	(SBIDER), University of Warwick, Coventry, CV4 7AL, UK		
9	<sup>3</sup> Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK		
10	<sup>4</sup> Communicable Disease Surveillance Centre, Health Protection Division, Public Health		
11	Wales, Swansea, SA2 8QA, UK		
12	<sup>5</sup> MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial		
13	College, London, W2 1PG, UK		
14	<sup>6</sup> Geneva Centre of Humanitarian Studies, University of Geneva, Geneva, 1205, Switzerland		
15			
16	*Correspondence to: <u>robin.thompson@maths.ox.ac.uk</u>		

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

17

#### <u>Abstract</u>

Tracking pathogen transmissibility during infectious disease outbreaks is essential for 18 19 assessing the effectiveness of public health measures and planning future control strategies. A 20 key measure of transmissibility is the time-dependent reproduction number, which has been 21 estimated in real-time during outbreaks of a range of pathogens from disease incidence time 22 series data. While commonly used approaches for estimating the time-dependent reproduction 23 number can be reliable when disease incidence is recorded frequently, such incidence data are 24 often aggregated temporally (for example, numbers of cases may be reported weekly rather 25 than daily). As we show, commonly used methods for estimating transmissibility can be 26 unreliable when the timescale of transmission is shorter than the timescale of data recording. 27 To address this, here we develop a simulation-based approach involving Approximate 28 Bayesian Computation for estimating the time-dependent reproduction number from 29 temporally aggregated disease incidence time series data. We first use a simulated dataset 30 representative of a situation in which daily disease incidence data are unavailable and only weekly summary values are reported, demonstrating that our method provides accurate 31 32 estimates of the time-dependent reproduction number under those circumstances. We then 33 apply our method to two previous outbreak datasets consisting of weekly influenza case 34 numbers from 2019-20 and 2022-23 in Wales (in the United Kingdom). Our simple-to-use 35 approach allows more accurate estimates of time-dependent reproduction numbers to be 36 obtained during future infectious disease outbreaks.

37

38 Keywords: Mathematical modelling, Infectious disease epidemiology, Reproduction number,

39 Parameter inference, Serial interval, Approximate Bayesian Computation, EpiEstim,

40 Influenza, Disease incidence

41

### **Introduction**

42 An important challenge for policy makers during infectious disease outbreaks is to devise 43 public health measures that limit transmission without placing an undue burden on the 44 population [1-3]. Central to the decision making process is an ability to monitor changes in 45 pathogen transmissibility in real-time during outbreaks, to determine whether current 46 interventions are sufficient or whether additional restrictions are required. 47 A widely used measure of transmissibility is the time-dependent reproduction number  $(R_t)$ [4–10]. The value of  $R_t$  represents the expected number of infections generated by someone 48 49 infected at time t over the course of their entire infectious period. This quantity changes 50 during an outbreak in response to interventions, variations in host behaviour and depletion of susceptible individuals due to infection-induced immunity. If the value of  $R_t$  is (and remains) 51 52 below one, then the outbreak will decline. On the other hand, if the value of  $R_t$  is (and 53 remains) above one, then the outbreak will grow.

54 Two distinct versions of  $R_t$  exist. First, the "instantaneous" reproduction number [4,5,11–13] 55 represents the expected number of infections generated by someone infected at time t over 56 their infectious period if transmission conditions do not change in future (i.e. assuming that 57 the control interventions in place at time t, and any other factors that affect transmission, are 58 not altered after time t). Second, the "case" reproduction number [12,14] is an analogous 59 quantity but accounts for changes in transmissibility that occur after time t (due to, for 60 example, changes in public health policy). Methods exist for estimating each of these 61 versions of  $R_t$  [15]. However, here we focus on the instantaneous reproduction number as it 62 is more amenable to analyses conducted in real-time during outbreaks when future changes in 63 pathogen transmissibility are unlikely to be known. We therefore refer to the instantaneous 64 reproduction number as  $R_t$  in this article.

65 A commonly used approach for estimating  $R_t$  is the method introduced by Cori *et al.* [4] 66 (hereafter referred to as the Cori method), implemented in the R software package EpiEstim 67 [16] and the online application *EpiEstim App* [17]. This approach is based on a renewal equation model of pathogen transmission (see Methods) and involves estimation of  $R_t$  from 68 disease incidence time series and an estimate of the serial interval distribution (the probability 69 70 distribution characterising the interval between symptom onset times in infector-infectee 71 transmission pairs), both on a daily timescale. The Cori method has been extended in a range 72 of ways following its original development [7], including accounting for imported cases 73 [5,11,18,19], uncertainty in the serial interval distribution [5], superspreading [20,21], 74 multiple pathogen variants [22] and unobserved generations of transmission [23]. 75 However, a challenge that besets estimation of  $R_t$  using the Cori method is temporal 76 aggregation of disease incidence time series data [7,24]. For COVID-19, for example, many 77 public health agencies switched from publishing daily numbers of reported cases to weekly 78 summaries after the height of the pandemic [25]. Often, disease incidence is reported weekly 79 even for pathogens including influenza [26] for which realised serial intervals and generation 80 times are typically only a few days [27–29]. A common workaround when using the Cori 81 method in these scenarios is to match the timescale of the serial interval distribution to the 82 timescale of the incidence data; for example, by supplying a weekly serial interval 83 distribution and applying the Cori method to weekly incidence data. This is problematic not 84 only because it is hard to unpick within-week changes in pathogen transmissibility when data 85 are reported weekly, but also because an assumption of the transmission model underlying the 86 Cori method is that all cases arising at timestep t are generated by infectors from earlier 87 timesteps. In other words, if the Cori method is applied with a weekly timestep, as considered 88 in this study, then it is assumed that an infector and infectee cannot both appear as cases in 89 the disease incidence data in the same week. As the timescale of transmission (as

90 characterised by the serial interval or generation time) of many pathogens is less than one
91 week, this assumption is often incorrect when considering weekly aggregated disease
92 incidence time series data.

93 In this research article, we address this issue by presenting a novel simulation-based method for estimating  $R_t$  from temporally aggregated disease incidence time series data and the serial 94 95 interval distribution. Our approach involves repeated simulation of a renewal equation 96 transmission model for different values of  $R_t$  with a timestep that is smaller than that of the 97 disease incidence data. Using an iterative version of Approximate Bayesian Computation 98 (ABC), we show how  $R_t$  can be estimated in real-time during outbreaks by matching model 99 simulations exactly to the temporally aggregated outbreak data. We apply our approach to simulated data, demonstrating its accuracy and comparing results from our method to those 100 101 obtained using the common workaround of the Cori method applied to temporally aggregated 102 data. We go on to apply our method to real-world outbreak data from the 2019-20 and 2022-103 23 influenza seasons in Wales in the United Kingdom.

104

## **Methods**

105 In our analyses, we consider two possible approaches for estimating  $R_t$  from temporally 106 aggregated disease incidence time series data: a workaround of the widely used Cori method 107 (Approach 1 in Fig 1) and our novel simulation-based method (Approach 2 in Fig 1). Since 108 the simulation-based approach uses a shorter timestep than that of data reporting, this method 109 accounts for the possibility of multiple generations of transmission occurring between dates 110 of data reporting. To provide a concrete setting in which to compare the two methods, we 111 focus on a situation in which disease incidence data are aggregated into weekly timesteps. 112 Below, we describe how the value of  $R_t$  each week can be estimated from the weekly data, 113 first using the Cori method (with a timestep of one week, since the incidence data are

- aggregated into weekly values), and then using our simulation-based approach (using a
- timestep shorter than one week, again using the weekly incidence data).



116



118 involves the application of the commonly used Cori method to weekly aggregated disease incidence time series.

119 Approach 2 (bottom) is the novel simulation-based approach, which involves matching simulations run with a

120 smaller timestep to the weekly aggregated data to estimate  $R_t$ . The second approach relaxes the assumption that

- 121 individuals appearing in the incidence data cannot have infected other individuals appearing in the same week.
- 122 Relaxing this assumption is particularly important during outbreaks in which the timescale of transmission is
- 123 shorter than the temporal aggregation of the data (e.g. if disease incidence time series data are aggregated
- 124 weekly, but serial intervals or generation times can be shorter than one week).

## 125

# 126 <u>The Cori method</u>

- 127 Following previous descriptions of the Cori method [4,5,11], we assume that the expected
- 128 number of cases,  $I_t$ , in week t, is given by

129 
$$\mathbb{E}(I_t|\{I_k\}_{k=1}^{t-1}, R_t, \boldsymbol{w}) = R_t \sum_{s=1}^{t-1} w_s I_{t-s}, \quad (1)$$

130 in which  $w_s$  is the probability that the (weekly discretised) serial interval takes the value *s* 131 weeks. We use the notation **w** to denote the sequence of values of  $w_s$  (s = 1, 2, ...).

The goal of the Cori method is to estimate  $R_t$ , assuming that it takes a constant value during the time period from week  $t - \tau$  to week t. In our analyses, we set  $\tau = 0$  to obtain an estimate of  $R_t$  each week, but we first present the method for general (non-negative integer value)  $\tau$  for consistency with previous presentations of this approach. If the number of cases in week t is drawn from a Poisson distribution, then the probability of observing weekly incidence  $\{l_k\}_{k=t-\tau}^t$  over the time window  $[t - \tau, t]$  (which consists of incidence data from  $\tau + 1$  weeks) is

139 
$$\mathbb{P}(\{I_k\}_{k=t-\tau}^t | \{I_k\}_{k=1}^{t-\tau-1}, R_t, \boldsymbol{w}) = \prod_{k=t-\tau}^t \frac{\left(R_t \sum_{s=1}^{k-1} w_s I_{k-s}\right)^{I_k} \exp\left(-R_t \sum_{s=1}^{k-1} w_s I_{k-s}\right)}{I_k!}$$

140 Assuming that the prior for  $R_t$  is a gamma distribution with shape parameter  $\alpha$  and rate

141 parameter  $\beta$ , then, by applying Bayes' Theorem, the posterior for  $R_t$  is

142 
$$p(R_t|\{I_k\}_{k=1}^t, \boldsymbol{w}) = \text{gamma}\left(R_t, \alpha + \sum_{k=0}^{\tau} I_{t-k}, \beta + \sum_{k=0}^{\tau} \sum_{s=1}^{t-k-1} I_{t-k-s} w_s\right),$$

in which we use the notation  $p(R_t | \{I_k\}_{k=1}^t, w)$  to represent the probability density function of 143 144  $R_t$  conditional on past incidence data and the weekly discretised serial interval distribution. 145 The notation gamma(x, a, b) represents the probability density of a gamma distribution at 146 value x with shape parameter a and rate parameter b. In all of our analyses, as in previous studies [4,5,11], we set  $\alpha = 1$  and  $\beta = 0.2$ . The prior for  $R_t$  therefore has mean and standard 147 148 deviation equal to five. The large standard deviation is chosen so that the prior is relatively 149 uninformative. The high mean ensures that the outbreak is not evaluated as being under control ( $R_t < 1$ ) unless this is very likely to be the case, so that interventions are not relaxed 150 151 erroneously.

152 Throughout the manuscript, we consider estimating individual values of  $R_t$  each week, based 153 on the numbers of new cases observed in that week. In other words, as noted above, we 154 assume that  $\tau = 0$ , in which case the above expression simplifies to

155 
$$p(R_t | \{I_k\}_{k=1}^t, \boldsymbol{w}) = \text{gamma}\left(R_t, \alpha + I_t, \beta + \sum_{s=1}^{t-1} I_{t-s} w_s\right).$$

156 The Cori method can therefore be used to obtain a posterior for  $R_t$  for  $t \ge 2$  weeks.

# 157 <u>Simulation-based inference of $R_t$ </u>

In the renewal equation model underlying the Cori method, the number of cases arising inweek *t* depends on the numbers of cases in previous weeks. Implicit in that approach is an

160	assumption that individuals appearing in the incidence data in any week cannot generate new		
161	cases in the same week. When disease incidence data are temporally aggregated, so that the		
162	timescale of transmission can be shorter than the timestep in the incidence data, this		
163	assumption may be incorrect. To relax this assumption, we consider a novel simulation-based		
164	approach for estimating $R_t$ . The goal of this method is again to estimate the value of $R_t$ for		
165	each week, $t \ge 2$ , but using a renewal equation model with a timestep that is shorter than one		
166	week (e.g., a daily timestep).		
167	Modified renewal equation		
168	In this approach, we consider partitioning the cases in each week into $P$ timesteps, where		
169	each new timestep is $1/P$ weeks. If, for example, $P = 7$ , then we are using a daily timestep in		
170	the simulation-based method. We introduce the following notation:		
171	• $I_{P(t-1)+i}^{(P)}$ represents the number of cases in the <i>i</i> th (partitioned) timestep within week t		
172	$(i=1,2,\ldots,P).$		
173	• $w_s^{(P)}$ represents the probability that the serial interval, discretised into timesteps of		
174	length $1/P$ weeks (see below and Supplementary Material), takes the value s		
175	timesteps.		
176	• $w^{(P)}$ represents the sequence of values of $w_s^{(P)}$ ( $s = 1, 2,$ ).		
177	In forward simulations of the corresponding renewal equation model, we assume that the		
178	number of cases in the $i$ th timestep of week $t$ is drawn from a Poisson distribution with mean		
179	$\mathbb{E}\left(I_{P(t-1)+i}^{(P)} \mid \left\{I_{k}^{(P)}\right\}_{k=1}^{P(t-1)+i-1}, R_{t}, \boldsymbol{w}^{(P)}\right) = R_{t} \sum_{s=1}^{P(t-1)+i-1} w_{s}^{(P)} I_{P(t-1)+i-s'}^{(P)}$		
180	for $i = 1, 2,, P$ . This is analogous to simulating the renewal equation underlying the Cori		
181	method but with a shorter timestep of $1/P$ weeks (rather than with a timestep of length one		
182	week).		

# 183 Inference of $R_t$

184 Inference of  $R_t$  under the simulation-based method involves repeated simulation of the 185 modified renewal equation model, using an iterative version of ABC. In short, the model is 186 simulated repeatedly in each week t, with a different value of  $R_t$  used in each simulation 187 (these  $R_t$  values are sampled independently from the prior, and incidence data for times 188 before week t are sampled from matching simulations from earlier weeks). This process is 189 repeated until a fixed number of simulations (denoted M) have been run in which the 190 simulated number of cases in week t exactly matches the corresponding number of cases in 191 the data,  $I_t$ . The values of  $R_t$  used to generate the matching simulations are then combined 192 into a posterior estimate for  $R_t$ . In all of our analyses using the simulation-based method, a 193 value of M = 1000 was used.

194 This procedure is repeated iteratively, starting with t = 2, then t = 3, and so on. Since this 195 approach only involves obtaining matching simulations for a single week at a time, estimates 196 of  $R_t$  can be obtained relatively quickly (compared to attempting to match an entire 197 simulation run over multiple weeks to the real-world data, as in standard ABC rejection 198 sampling [30]). For a more detailed description of the simulation-based inference method, 199 including an explanation of how cases are distributed between timesteps within the first week 200 in each simulation, see the Supplementary Material. A schematic explaining the steps 201 involved in the inference procedure is shown in Fig S1.

202 <u>Outbreak datasets</u>

We consider three outbreak datasets in our analyses. We first test our approach on a simulated dataset. The use of simulated data not only enables us to compare estimates of  $R_t$  obtained using the simulation-based approach against analogous estimates using the Cori method, but

it also allows us to verify that the simulation-based approach for estimating  $R_t$  generates accurate estimates in a setting in which we know true value of  $R_t$  (i.e., the value used to generate the simulated dataset). We then go on to compare outputs from the simulation-based approach and the Cori method using weekly aggregated disease incidence time series for influenza from 2019-20 and 2022-23 in Wales. *Simulated dataset (Fig 2)* 

212 We generated simulated data using the modified renewal equation, using a very small

timestep so that the discretised serial interval is a close approximation to the continuous serial

- 214 interval. Specifically, a disease incidence time series was generated starting from one initial
- 215 case (in the first timestep) using a timestep of 10 minutes ( $P = 24 \times 7 \times 6 = 1,008$ ). To

216 generate a classic epidemic curve, the simulation was run for 11 weeks with  $R_t = 1.5$  for  $t \le 1.5$ 

217 6 weeks and  $R_t = 0.75$  for t > 6 weeks.

218 Influenza in Wales, 2019-20 (Figs 3,4) and 2022-23 (Figs 5,6)

219 To demonstrate our approach on real-world data, we considered two disease incidence time 220 series datasets provided by Public Health Wales describing estimated numbers of cases of 221 influenza-like illness (ILI) in Wales each week. The original data comprised the clinical 222 consultation rate per 100,000 individuals in sentinel practices in Wales each week [31]. The 223 total number of weekly cases was then estimated by multiplying each value in the original 224 data by 31.075 (i.e. scaling these values based on the population size of Wales, which is 225 3,107,500 [32]). Since the dataset is for ILI, it likely contains some cases that were not 226 influenza. Nonetheless, these data are sufficient to demonstrate and test the methods that we 227 present in our study, and so we assume that the datasets are representative of numbers of 228 influenza cases in Wales. Weekly data were provided from 28 October 2019 to 2 February

2020 (Fig 3A) and 31 October 2022 to 5 February 2023 (Fig 5A). These date ranges each
span 14 weeks with high ILI burden.

## 231 <u>Serial interval</u>

Since we analyse influenza outbreak datasets in this study, we assume throughout that the 232 233 (continuous) serial interval distribution is a gamma distribution with mean 0.37 weeks (2.6 234 days) and standard deviation 0.19 weeks (1.3 days) [33]. While this estimate was derived from household data for pandemic influenza, our focus is on demonstrating the application of 235 236 the simulation-based method rather than precise estimation of the serial interval, and we expect this estimate to be in line with the serial interval for seasonal influenza (i.e., a mean 237 value of less than one week). Denoting the probability density function of the serial interval 238 239 distribution by q(x), then q(x) = gamma(x, 4, 10.8).

240 We discretise this distribution into timesteps of length 1/P weeks to obtain  $w^{(P)}$ . To do this,

241 we adapt the method used by Cori *et al.* [4] in which the serial interval distribution is

242 discretised into timesteps of length one. Specifically, we set

243 
$$w_k^{(P)} = \int_{(k-1)/P}^{(k+1)/P} g(u) \left(1 - P \left|u - \frac{k}{P}\right|\right) du, \quad \text{for } k = 2,3,4, ...$$

as derived in the Supplementary Material. We then choose  $w_1^{(P)}$  so that  $w^{(P)}$  is a valid probability distribution (i.e., the sum of the entries of  $w^{(P)}$  is one). The rationale for normalising  $w^{(P)}$  in this way is that same-timestep cases (i.e. infectors and infectees appearing in the same timestep) are not possible in the renewal equation model. Our approach involves assigning all probability density near zero in g(x) to  $w_1^{(P)}$ , which is the shortest possible serial interval in the model.

## 250

### **Results**

# 251 <u>Simulated dataset</u>

252 We first considered the simulated disease incidence time series dataset in which the incidence data are aggregated into weekly counts (Fig 2A). Since this dataset was generated using a 253 254 serial interval for influenza, transmission occurred on a timescale less than one week. The 255 discretised serial interval is shown in Fig 2B, both with a weekly timestep for use with the 256 Cori method (P = 1; blue) and with a daily timestep for use with the simulation-based 257 method (P = 7; red). Since the renewal equation model underlying both the Cori method and 258 our simulation-based approach does not allow individuals appearing in the incidence data to 259 generate new cases in the same timestep, only the simulation-based approach allows within-260 week realised serial intervals.

We applied both  $R_t$  inference methods to the simulated dataset, finding in this scenario that the simulation-based approach generates more accurate estimates of  $R_t$  than the Cori method (Fig 2C). The percentage error in the estimated value of  $R_t$  each week using the simulationbased approach with P = 7 (compared to value of  $R_t$  used to generate the dataset) is shown in Fig 2D.

In addition to our main analysis shown in Fig 2, we also conducted other analyses using the simulated dataset. We demonstrated that when the simulation-based method is applied with P = 1, the output matches the results obtained when the Cori method is used to estimate  $R_t$ (Fig S2A), as would be expected since the assumptions underlying the two methods are identical in this case. We also considered how  $R_t$  estimates obtained using the simulationbased method change when different values of P are chosen (Fig S2B-D), finding that the

- 272 method can obtain accurate estimates for relatively small values of P (using a value of P = 3
- 273 led to similar errors compared to using P = 7).



275 Figure 2. Estimation of R<sub>t</sub> from the simulated disease incidence time series dataset. A. The simulated 276 outbreak dataset (blue bars), generated with  $R_t = 1.5$  for  $t \le 6$  weeks and  $R_t = 0.75$  for t > 6 weeks (black 277 line). The outbreak was simulated with P = 1,008 starting from one initial case in the first timestep, and new 278 cases were then aggregated into weekly case counts. B. The discretised serial interval, for P = 1 (as used with 279 the Cori method; blue) and P = 7 (red). C. Estimates of  $R_t$  using the Cori method (blue) and the novel 280 simulation-based approach (with P = 7; red). Blue and red lines are the mean estimates, and the shaded regions 281 represent the 95% credible intervals. The value of  $R_t$  underlying the simulation is shown in black. D. The 282 percentage error in the mean estimate of  $R_t$  each week (relative to the true value of  $R_t$  used to generate the 283 dataset) using the simulation-based method with P = 7.

#### 284 Influenza in Wales, 2019-20 and 2022-23

285 We then went on to consider the two Wales influenza outbreak datasets, again using both the 286 Cori and simulation-based methods to estimate  $R_t$ . First, we considered the weekly case 287 counts from the 2019-20 influenza season (Fig 3A). As with the simulated dataset, the 288 simulation-based approach led to different estimates of  $R_t$  than the Cori method;  $R_t$  estimates 289 obtained using the simulation-based approach were typically lower than those from the Cori 290 method during November and December 2019 (the simulation-based approach generally led 291 to  $R_t$  estimates between one and two, whereas the Cori method generated estimates above 292 two in multiple weeks), but then higher than those from the Cori method for most of January 2020 when  $R_t$  was estimated to be less that one (Fig 3B). We computed the percentage error 293 294 in the  $R_t$  estimate each week using the simulation-based method with P = 7 (Fig 3C). Since 295 the true underlying value of  $R_t$  was unknown, the percentage error was computed relative to 296 applying the simulation-based method with a very large value of P = 168 (this is 297 representative of the best possible estimate of  $R_t$  obtainable from the weekly incidence data; 298 using a partitioning value of P = 168 returns inferred values of  $R_t$  estimated with a one-hour 299 timestep). We also explored how  $R_t$  estimates depend on the value of P that is used (Fig 4). 300 Estimates obtained using the Cori method and using the simulation-based method with P = 1301 again matched closely (Fig 4A). We found that a value of P = 7 is large enough for accurate 302 inference of  $R_t$  (Fig 4D).

The analyses of the Wales influenza data from 2019-20 were then repeated for the data from 2022-23, with similar results (Figs 5,6). Notably, the Cori method generally led to a higher estimate of  $R_t$  than the simulation-based method when  $R_t$  was estimated to be greater than one, and a lower estimate of  $R_t$  than the simulation-based method when  $R_t$  was estimated to be less than one (see Discussion).





309

310 Figure 3. Estimation of R<sub>t</sub> for influenza in Wales, 2019-2020. A. Weekly numbers of ILI cases in Wales from

311 28 October 2019 to 2 February 2020, estimated from surveillance data collected in sentinel practices. B.

312 Estimates of  $R_t$  using the Cori method (blue) and the novel simulation-based approach (with P = 7; red). Blue

313 and red lines are the mean estimates, and the shaded regions represent the 95% credible intervals. C. The

314 percentage error in the mean estimate of  $R_t$  each week using the simulation-based method with P = 7,

315 compared to using a larger value of P = 168 (which corresponds to estimating  $R_t$  with a one-hour timestep).



317 Figure 4. Dependence of R<sub>t</sub> estimates using the simulation-based method on the value of P used, for 318 influenza in Wales, 2019-2020. A. Estimates of  $R_t$  obtained when the Cori method (blue) and the novel 319 simulation-based approach with P = 1 (red) are applied to the 2019-20 influenza dataset (Fig 3A). B. Analogous 320 to panel A, but with P = 3 in the simulation-based approach. C. Analogous to panel A, but with P = 5 in the 321 simulation-based approach. D. The average weekly absolute error in mean  $R_t$  estimates obtained using the 322 simulation-based method with different values of P, compared to using a larger value of P = 168 (which 323 corresponds to estimating  $R_t$  with a one-hour timestep). For a given value of P, this measure represents the 324 absolute value of the error in the estimate of  $R_t$  in week t (compared to using P = 168), averaged over all 325 values of *t*.





327 Figure 5. Estimation of R<sub>t</sub> for influenza in Wales, 2022-2023. A. Weekly numbers of ILI cases in Wales from

328 31 October 2022 to 5 February 2023, estimated from surveillance data collected in sentinel practices. B.

329 Estimates of  $R_t$  using the Cori method (blue) and the novel simulation-based approach (with P = 7; red). Blue

330 and red lines are the mean estimates, and the shaded regions represent the 95% credible intervals. C. The

331 percentage error in the mean estimate of  $R_t$  each week using the simulation-based method with P = 7,

332 compared to using a larger value of P = 168 (which corresponds to estimating  $R_t$  with a one-hour timestep).



334 Figure 6. Dependence of  $R_t$  estimates using the simulation-based method on the value of P used, for 335 influenza in Wales, 2022-2023. A. Estimates of  $R_t$  obtained when the Cori method (blue) and the novel 336 simulation-based approach with P = 1 (red) are applied to the 2022-23 influenza dataset (Fig 5A) B. Analogous 337 to panel A, but with P = 3 in the simulation-based approach. C. Analogous to panel A, but with P = 5 in the 338 simulation-based approach. D. The average weekly absolute error in mean  $R_t$  estimates obtained using the 339 simulation-based method with different values of P, compared to using a larger value of P = 168 (which 340 corresponds to estimating  $R_t$  with a one-hour timestep). For a given value of P, this measure represents the 341 absolute value of the error in the estimate of  $R_t$  in week t (compared to using P = 168), averaged over all 342 values of *t*.

343

#### **Discussion**

344 During infectious disease outbreaks, evaluation of time-varying changes in pathogen 345 transmission is essential to inform outbreak responses. Different metrics can be tracked, 346 including incidence of new cases, hospitalisations and deaths, and outbreak growth rates 347 [34,35]. A key metric that has been estimated in real-time during outbreaks of a range of 348 pathogens is  $R_t$ , in part because of its straightforward interpretation [7,9,15]. Not only is 349 there a threshold value of  $R_t = 1$ , below which an outbreak can be inferred as being under 350 control, but the value of  $R_t$  also provides information about the extent to which the level of 351 transmission must change (relative to current transmission) for an outbreak to grow or decline. For example, if  $R_t = 2$ , then more than half of transmissions must be prevented for 352 353 the outbreak to decline. Similarly, if  $R_t = 0.5$ , then up to twice as many transmissions may 354 occur before the outbreak begins to grow. Precise estimation of  $R_t$  is therefore crucial.

355 Here, we have presented a novel simulation-based approach for estimating  $R_t$  in scenarios in 356 which disease incidence time series data are aggregated temporally (Fig 1). While 357 epidemiological data may be collected at a fine temporal resolution, it is common for the data 358 to then be aggregated (e.g., into weekly or monthly counts). Aggregated data may be easier to 359 report and can be more accurate than data presented at a high temporal resolution when there 360 is uncertainty in the precise times at which cases occurred. However, as we have shown, 361 frequently used methods for inferring  $R_t$ , such as the Cori method [4,5], may not generate 362 accurate estimates when applied to temporally aggregated data if transmission occurs more 363 rapidly than the temporal resolution of the aggregated data. This is because the renewal 364 equation model underlying the Cori method involves assuming that an individual appearing 365 in the disease incidence time series data at timestep t cannot have infected other individuals 366 appearing in the same timestep. Our proposed simulation-based approach addresses this, by

exactly matching simulations of a renewal equation model run with a shorter timestep (*P* timesteps for each timestep in the aggregated data) to the temporally aggregated incidence data. The simulation-based approach not only provides accurate estimates of  $R_t$  (Fig 2), but can also be applied easily to real-world datasets (Figs 3-6). While using a very large value of *P* allows the most accurate possible  $R_t$  estimates to be obtained from the aggregated data, even relatively small values of *P* are sufficient for  $R_t$  to be inferred accurately (Figs 4,6).

We found that, while the Cori method did not always provide an accurate estimate of  $R_t$  due 373 374 to the temporal aggregation of the disease incidence data, it was able to identify whether or 375 not  $R_t$  is below one (i.e., the outbreak is under control). While this is useful, as noted above 376 precise estimation of  $R_t$  is important as it provides information about the number of 377 transmissions that must be prevented for an outbreak to be controlled (or the number of 378 transmissions that can occur for an outbreak to remain under control). The difference in  $R_t$ estimates between the two methods can be explained by the assumption of no same-timestep 379 380 cases (i.e., infectors and infectees cannot appear in the disease incidence time series in the 381 same timestep) in the renewal equation. When the Cori method is applied to weekly data, this 382 then leads to overestimation of the serial interval, which is known in turn to lead to overestimation of  $R_t$  if the true value of  $R_t$  is greater than one and underestimation of  $R_t$  if 383 384 the true value of  $R_t$  is less than one [36,37].

A closely related study by Nash *et al.* [24], undertaken at the same time as the analyses presented here, has also considered estimation of  $R_t$  from temporally aggregated disease incidence time series data. In that approach, an expectation-maximisation (EM) algorithm is used to reconstruct daily incidence from any aggregation of disease incidence data using the serial interval (on a daily timescale). The original version of the Cori method is then applied to the estimated daily data. This EM approach has been integrated into the EpiEstim R

391 software package [16]. There are several differences between the approach by Nash *et al.* [24] 392 and the simulation-based method described here. First, the two approaches are 393 methodologically distinct, relying on entirely different underlying methods (EM or model 394 simulation). Second, under the approach by Nash *et al.*, only a single estimated daily disease 395 incidence time series is obtained. In contrast, our method involves matching a range of 396 simulations to the temporally aggregated data, thereby considering different possible 397 disaggregated disease incidence time series that could have led to the weekly aggregated data. 398 Third, our method can be run straightforwardly for a range of values of P, allowing the most 399 accurate possible estimates of  $R_t$  to be inferred from temporally aggregated incidence data 400 (discretisation into a timestep of less than one day is straightforward). Fourth, our approach 401 can be applied easily in scenarios in which the serial interval distribution is discrete rather 402 than continuous (if, for example, the serial interval distribution is constructed directly from 403 observations of dates on which infector-infectee pairs report symptoms). A rigorous 404 comparison of estimated values of  $R_t$  obtained using the simulation-based method and using the EM approach of Nash et al. [24] is a target for future exploration. However, an initial 405 406 investigation involving applying the methods to the Wales influenza datasets considered here 407 suggests that the two approaches can obtain consistent results (Fig S3).

408 Our simulation-based method is conceptually straightforward, simply requiring repeated 409 simulation of a renewal equation model. It is also computationally efficient to run, as 410 simulations are only required to match the real-world data for one aggregated timestep at a 411 time. This contrasts with using ABC rejection sampling to estimate all values of  $R_t$ 412 simultaneously, which would involve matching entire simulated time series to the entire real-413 world dataset. The efficiency of our approach allowed us to require that the simulations used 414 to infer  $R_t$  match the real-world data exactly. Further computational efficiency could be achieved by removing this condition, and instead setting a threshold "distance" within which 415

416 a simulation is determined to match the real-world data, as is common when using ABC 417 [30,38]. However, this necessitates that a distance metric is chosen, and resulting estimates of 418  $R_t$  may be less accurate.

419 As in any modelling study, our framework in its current form involves assumptions. We 420 followed previous publications in which the Cori method has been used [4,5] and assumed that the time series datasets from which we estimated  $R_t$  represent numbers of new 421 422 symptomatic cases in each timestep. In the disease incidence time series data, it is then 423 assumed that each infectee appears after their infector following a time period that reflects a random draw from the serial interval distribution, which is assumed to always take strictly 424 425 positive values. However, in reality, realised serial intervals can be negative (if an infectee develops symptoms before their infector; this is possible, for example, for transmission of 426 427 SARS-CoV-2 [39–42]). Rather than using disease incidence time series, it is possible to apply 428 both the Cori method and the simulation-based method to data describing incidence of 429 infections, replacing the serial interval distribution as an input with the distribution of the 430 generation time (the interval between infection times in infector-infectee pairs). This can be 431 beneficial as realised generation times are always positive. However, since times of infection 432 are often unknown, infection incidence data are typically not observed directly. Consequently, 433 further inference would then be required to estimate incidence of infections, as well as to 434 estimate the generation time distribution [8,43–45].

In our analyses, we assumed that all cases in the disease incidence time series (after the first timestep) arose because of transmission within the population under consideration, and that all cases were recorded. In reality, some infected individuals may become infected outside the local population [5,11,46,47] and under-reporting of cases is likely for many pathogens [48– 51]. Extension of our method to account for these features of real-world outbreaks is a target

for future research. Similarly, our method assumes that a Poisson distributed number of cases
occur in each timestep of the modified renewal equation model. Considering different
possible probability distributions, including accounting for the possibility of superspreading
events on some days [20,21,52], is another possible area for future work.

444 Further testing of the performance of the simulation-based approach in different scenarios

445 would also be worthwhile. For example, in settings in which disease incidence time series are

subject to a "day-of-the-week effect" [53] (e.g., if cases occurring at the weekend are

447 typically reported with a longer reporting delay than those arising during the week),  $R_t$ 

448 inference using the simulation-based method applied to weekly aggregated incidence data

449 may generate more robust estimates than attempting to infer  $R_t$  from less accurate daily

450 incidence data. Our method can also be adapted for scenarios in which the disease incidence

451 time series data are aggregated into timesteps that are not all of equal length. For example,

452 when incidence data are derived from World Health Organization reports that are published

453 irregularly in time, the timestep changes during the outbreak [54], and those irregular

454 timesteps can be used directly in our simulation-based method.

In summary, we have presented a novel method for estimating  $R_t$  from temporally aggregated disease incidence time series. Going forwards, the ideal scenario is for disease incidence time series to be recorded accurately at a fine temporal resolution (e.g., daily). If that occurs, then existing methods for estimating  $R_t$  are generally expected to perform well. However, if disease incidence time series continue to be aggregated temporally for pathogens for which transmission occurs on a short timescale, then methods allowing accurate  $R_t$  inference from temporally aggregated data are of paramount importance.

## 462 **COMPETING INTERESTS**

463 We have no competing interests.

### 464 AUTHORS' CONTRIBUTIONS

- 465 IOG formal analysis, investigation, visualisation, validation, writing original draft,
- 466 writing review and editing.
- 467 WSH methodology, writing review and editing.
- 468 JS methodology, writing review and editing.
- 469 RKN methodology, writing review and editing.
- 470 JP methodology, writing review and editing.
- 471 AC methodology, writing review and editing.
- 472 EMH methodology, supervision, visualisation, writing review and editing.
- 473 RNT conceptualization, methodology, project administration, supervision, visualisation,
- 474 writing original draft, writing review and editing.

# 475 FUNDING

- 476 This research was funded by the EPSRC through the Mathematics for Real-World Systems
- 477 CDT (ZO-G, RNT; grant number EP/S022244/1) and a doctoral prize (WSH; grant number
- 478 EP/W524311/1). The collaboration between JS and RNT was funded by a grant from Public
- 479 Health Wales. EMH and RNT would like to acknowledge the help and support of the
- 480 JUNIPER partnership, funded by MRC (grant number MR/X018598/1), to which they are
- 481 linked.

# 482 ACKNOWLEDGEMENTS

- 483 Thanks to the Communicable Disease Surveillance Centre at Public Health Wales for
- 484 providing us with the data used in this research. Thanks to members of the Zeeman Institute
- 485 for Systems Biology and Infectious Disease Epidemiology Research at the University of

- 486 Warwick, particularly Alex Kaye, and the Wolfson Centre for Mathematical Biology at the
- 487 University of Oxford, for useful discussions about this work.

## 488 DATA AVAILABILITY

- 489 The computing code used to perform the analyses in this article is available in the following
- 490 GitHub repository: www.github.com/billigitt/R\_Estim\_Simulation\_Method. All computer
- 491 code was written in the MATLAB programming environment (compatible with version
- 492 R2022a).
- 493

# 494 **<u>References</u>**

- Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. Mitigation strategies
   for pandemic influenza A: Balancing conflicting policy objectives. PLoS Comput
   Biol. 2011;7: e1001076-e1001076.
- 498 2. Smith RD, Keogh-Brown MR, Barnett T, Tait J. The economy-wide impact of pandemic
  499 influenza on the UK: a computable general equilibrium modelling experiment. BMJ.
  500 2009;339: b4571-b4571.
- 3. Tildesley MJ, Vassall A, Riley S, Jit M, Sandmann F, Hill EM, et al. Optimal health and
  economic impact of non-pharmaceutical intervention measures prior and post
  vaccination in England: a mathematical modelling study. R Soc Open Sci. 2022;9:
  211746.
- 4. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to
  estimate time-varying reproduction numbers during epidemics. Am J Epidemiol.
  2013;178: 1505–12.
- 508 5. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et
  al. Improved inference of time-varying reproduction numbers during infectious
  disease outbreaks. Epidemics. 2019;29: 100356–100356.
- 511 6. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical
  512 estimation of time-dependent epidemic trends. Math Stat Estim App Epidem. 2009.
  513 pp. 103–121.
- 514 7. Nash RK, Nouvellet P, Cori A. Real-time estimation of the epidemic reproduction
  515 number: Scoping review of the applications and challenges. PLoS Digit Health.
  516 2022;1: e0000052.

- 8. Gostic KM, McGough L, Baskerville E, Abbott S, Joshi K, Tedijanto C, et al. Practical
  considerations for measuring the effective reproductive number, Rt. PLoS Comput
  Biol. 2020.
- 520 9. Vegvari C, Abbott S, Ball F, Brooks-Pollock E, Challen R, Collyer BS, et al. Commentary
  521 on the use of the reproduction number R during the COVID-19 pandemic. Stat Meth
  522 Med Res. 2021;1: 1–11.
- Thompson RN, Hollingsworth TD, Isham V, Arribas-Bel D, Ashby B, Britton T, et al.
   Key questions for modelling COVID-19 exit strategies. Proc Roy Soc B. 2020;287:
   20201405–20201405.
- 526 11. Creswell R, Augustin D, Bouros I, Farm HJ, Miao S, Ahern A, et al. Heterogeneity in
  527 the onwards transmission risk between local and imported cases affects practical
  528 estimates of the time-dependent reproduction number. Phil Trans R Soc A.
  529 2022;380: 20210308.
- 530 12. Fraser C. Estimating individual and household reproduction numbers in anemerging epidemic. PLoS One. 2007;2: e758.
- 532 13. Dai C, Zhou D, Gao B, Wang K. A new method for the joint estimation of
  533 instantaneous reproductive number and serial interval during epidemics. PLoS
  534 Comput Biol. 2023;19: e1011021.
- 535 14. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory
  536 syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160:
  537 509–516.
- 538 15. White LF, Moser CB, Thompson RN, Pagano M. Statistical estimation of the
  539 reproductive number from case notification data. Am J Epidem. 2020; kwaa211.
- 540 16. EpiEstim Team. EpiEstim: Estimate time varying reproduction numbers from
  541 epidemic curves. Version 2.2-4. 2021. Available: www.cran.r542 project.org/web/packages/EpiEstim/
- 543 17. EpiEstim App Team. EpiEstim App. 2019. Available:544 www.shiny.dide.imperial.ac.uk/epiestim/
- Li W, Bulekova K, Gregor B, White LF, Kolaczyk ED. Estimation of local time-varying
  reproduction numbers in noisy surveillance data. Phil Trans Roy Soc A. 2022.
- 547 19. Tsang TK, Wu P, Lau EHY, Cowling BJ. Accounting for imported cases in estimating
  548 the time-varying reproductive number of COVID-19 in Hong Kong. J Infect Dis.
  549 2021;224: 783–787.
- Johnson KD, Beiglböck M, Eder M, Grass A, Hermisson J, Pammer G, et al. Disease
  momentum: Estimating the reproduction number in the presence of
  superspreading. Infect Dis Model. 2021;6: 706–728.

- 553 21. Ho F, Parag KV, Adam DC, Lau EHY, Cowling BJ, Tsang TK. Accounting for the
  554 potential of overdispersion in estimation of the time-varying reproduction number.
  555 Epidemiology. 2023;34: 201–205.
- Bhatia S, Wardle J, Nash RK, Nouvellet P, Cori A. Extending EpiEstim to estimate the
  transmission advantage of pathogen variants in real-time: SARS-CoV-2 as a casestudy. Epidemics. 2023;44: 100692.
- Brizzi A, O'Driscoll M, Dorigatti I. Refining reproduction number estimates to
  account for unobserved generations of infection in emerging epidemics. Clin Infect
  Dis. 2022;75: e114-e121.
- 562 24. Nash RK, Bhatt S, Cori A, Nouvellet P. Estimating the epidemic reproduction
  563 number from temporally aggregated incidence data: a statistical modelling
  564 approach and software tool. PLoS Comput Biol. 2023;19: e1011439.
- 565 25. UK Health Security Agency. The COVID-19 dashboard moves to weekly updates.
  566 2022. Available: www.ukhsa.blog.gov.uk/2022/06/28/the-covid-19-dashboard567 moves-to-weekly-updates/
- 568 26. UK Health Security Agency. National Influenza and COVID-19 surveillance report:
  569 Week 29 report (up to week 28 data). 2023. Available:
  570 www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance571 reports-2022-to-2023-season
- 572 27. Cowling BJ, Fang VJ, Riley S, Malik Peiris JS, Leung GM. Estimation of the serial
  573 interval of influenza. Epidemiology. 2009;20: 344–347.
- te Beest DE, Wallinga J, Donker T, Van Boven M. Estimating the generation interval
  of Influenza A (H1N1) in a range of social settings. Epidemiology. 2013;24: 244–
  250.
- 577 29. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the
  578 reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic
  579 review of the literature. BMC Infect Dis. 2014;14: 480.
- 580 30. Minter A, Retkute R. Approximate Bayesian Computation for infectious disease
  581 modelling. Epidemics. 2019;29: 100368.
- 582 31. Public Health Wales. Weekly influenza and acute respiratory infection surveillance
  583 report: Wednesday 22nd February 2023 (covering week 07 2023). 2023. Available:
  584 www.phw.nhs.wales/topics/immunisation-and-vaccines/fluvaccine/weekly585 influenza-and-acute-respiratory-infection-report/october-2022-october-2023-flu586 season-202223/phw-influenza-surveillance-report-for-2023-week-7pdf/
- 587 32. Office for National Statistics. Population and household estimates, Wales: Census
  588 2021. 2021. Available:
- 589www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/popul590ationestimates/bulletins/populationandhouseholdestimateswales/census2021

- Solution 591
  Solution 33. Cauchemez S, Donnelly CA, Reed C, Ghani AC, Fraser C, Kent CK, et al. Household transmission of 2009 pandemic Influenza A (H1N1) virus in the United States. N Engl J Med. 2009;361: 2619–2627.
- 34. Pellis L, Scarabel F, Stage HB, Overton CE, Chappell LH, Fearon E, et al. Challenges in
  control of Covid-19: short doubling time and long delay to effect of interventions.
  Phil Trans Roy Soc B. 2021;376: 20200264–20200264.
- 597 35. Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative
  598 than reproduction numbers? J R Stat Soc Ser A. 2022;1: 1–11.
- 36. Wallinga J, Lipsitch M. How generation intervals shape the relationship between
  growth rates and reproductive numbers. Proc R Soc B Biol Sci. 2007;274: 599–604.
- Knight J, Mishra S. Estimating effective reproduction number using generation time
  versus serial interval, with application to COVID-19 in the Greater Toronto Area,
  Canada. Infect Dis Model. 2020;5: 889–896.
- 38. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian
  computation scheme for parameter inference and model selection in dynamical
  systems. J R Soc Interface. 2009;6: 187–202.
- Bu Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial interval of COVID-19 among
  publicly reported confirmed cases. Emerg Infect Dis. 2020;26: 1341–1343.
- 40. Hart WS, Maini PK, Thompson RN. High infectiousness immediately before COVID19 symptom onset highlights the importance of continued contact tracing. eLife.
  2021;10: e65534.
- 612 41. Geismar C, Nguyen V, Fragaszy E, Shrotri M, Navaratnam AMD, Beale S, et al.
  613 Bayesian reconstruction of SARS-CoV-2 transmissions highlights substantial
  614 proportion of negative serial intervals. Epidemics. 2023;44: 100713.
- 42. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Vespignani A, Dean NE. Rapid review
  and meta-analysis of serial intervals for SARS-CoV-2 Delta and Omicron variants.
  BMC Infect Dis. 2023;23: 429.
- Abbott S, Hellewell J, Thompson RN, Sherratt K, Gibbs HP, Bosse NI, et al. Estimating
  the time-varying reproduction number of SARS-CoV-2 using national and
  subnational case counts. Wellcome Open Res. 2020;5: 112.
- 44. Hart WS, Abbott S, Endo A, Hellewell J, Miller E, Andrews N, et al. Inference of the
  SARS-CoV-2 generation time using UK household data. eLife. 2022;11: e70767.
- 45. Hart WS, Miller E, Andrews NJ, Waight P, Maini PK, Funk S, et al. Generation time of
  the alpha and delta SARS-CoV-2 variants: an epidemiological analysis. Lancet Inf
  Dis. 2022;22: 603–610.
- 46. Daon Y, Thompson RN, Obolski U. Estimating COVID-19 outbreak risk through air
  travel. J Travel Med. 2020;27: taaa093.

628	47.	Didelot X, Helekal D, Kendall M, Ribeca P. Distinguishing imported cases from
629		locally acquired cases within a geographically limited genomic sample of an
630		infectious disease. Bioinformatics. 2023;39: btac761.

- 48. Dalziel BD, Lau MSY, Tiffany A, McClelland A, Zelner J, Bliss JR, et al. Unreported
  cases in the 2014-2016 Ebola epidemic: Spatiotemporal variation, and implications
  for estimating transmission. Althouse B, editor. PLoS Negl Trop Dis. 2018;12:
  e0006161.
- 49. Albani V, Loria J, Massad E, Zubelli J. COVID-19 underreporting and its impact on
  vaccination strategies. BMC Infect Dis. 2021;21: 1111.
- 637 50. Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, et al.
  638 Measuring underreporting and under-ascertainment in infectious disease datasets:
  639 a comparison of methods. BMC Public Health. 2014;14: 147.
- 51. Lovell-Read FA, Funk S, Obolski U, Donnelly CA, Thompson RN. Interventions
  targeting non-symptomatic cases can be important to prevent local outbreaks:
  SARS-CoV-2 as a case study. J R Soc Interface. 2021;18: 20201014.
- 52. Bradbury NV, Hart WS, Lovell-Read FA, Polonsky JA, Thompson RN. Exact
  calculation of end-of-outbreak probabilities using contact tracing data. medRxiv.
  2023.
- 646 53. Guzmán-Rincón LM, Hill EM, Dyson L, Tildesley MJ, Keeling MJ. Bayesian estimation
  647 of real-time epidemic growth rates using Gaussian processes: local dynamics of
  648 SARS-CoV-2 in England. J R Stat Soc C. 2023;00: 1–22.
- 54. Shaman J, Yang W, Kandula S. Inference and forecast of the current West African
  Ebola outbreak in Guinea, Sierra Leone and Liberia. PLoS Curr. 2014;1: 6.