**A simulation-based approach for estimating the time-dependent reproduction number from temporally aggregated disease incidence time series data**

I Ogi-Gittins1,2, WS Hart3, J Song4, RK Nash5, J Polonsky6, A Cori5, EM Hill1,2, RN Thompson1,2,3\*

**Affiliations:**

1Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK

2Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research (SBIDER), University of Warwick, Coventry, CV4 7AL, UK

3Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK

4Communicable Disease Surveillance Centre, Health Protection Division, Public Health Wales, Swansea, SA2 8QA, UK

5MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College, London, W2 1PG, UK

6Geneva Centre of Humanitarian Studies, University of Geneva, Geneva, 1205, Switzerland

\*Correspondence to: [robin.n.thompson@warwick.ac.uk](mailto:robin.n.thompson@warwick.ac.uk)

**Abstract**

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

Keywords: Mathematical modelling, Infectious disease epidemiology, Reproduction number, Parameter inference, Serial interval, Approximate Bayesian Computation

**Introduction**

An important challenge for policy makers during infectious disease outbreaks is to devise public health measures that limit transmission without placing an undue burden on the population [1–3]. Central to this decision making is an ability to monitor changes in pathogen transmissibility in real-time during outbreaks, to determine whether current interventions are sufficient or whether additional restrictions may be required.

A widely used measure of transmissibility is the time-dependent reproduction number () [4–9]. The value of represents the expected number of infections generated by someone infected at time over the course of their entire infectious period. This quantity changes during an outbreak in response to interventions. If the value of is (and remains) below one, then the outbreak will decline. On the other hand, if the value of is (and remains) above one, then the outbreak will grow.

Two distinct versions of exist. First, the “instantaneous” reproduction number [4,5,10–12] represents the expected number of infections generated by someone infected at time over their infectious period if transmission conditions do not change in future (i.e. assuming that the control interventions in place at time are not altered after time ). Second, the “case” reproduction number [11,13] is an analogous quantity but accounting for changes in transmissibility that occur after time (due to e.g. changes in public health policy). Methods exist for estimating each of these versions of [14]. However, here we focus on the instantaneous reproduction number as it is more amenable to analyses conducted in real-time during outbreaks when future changes in pathogen transmissibility are unlikely to be known. Hereafter, we refer to the instantaneous reproduction number as .

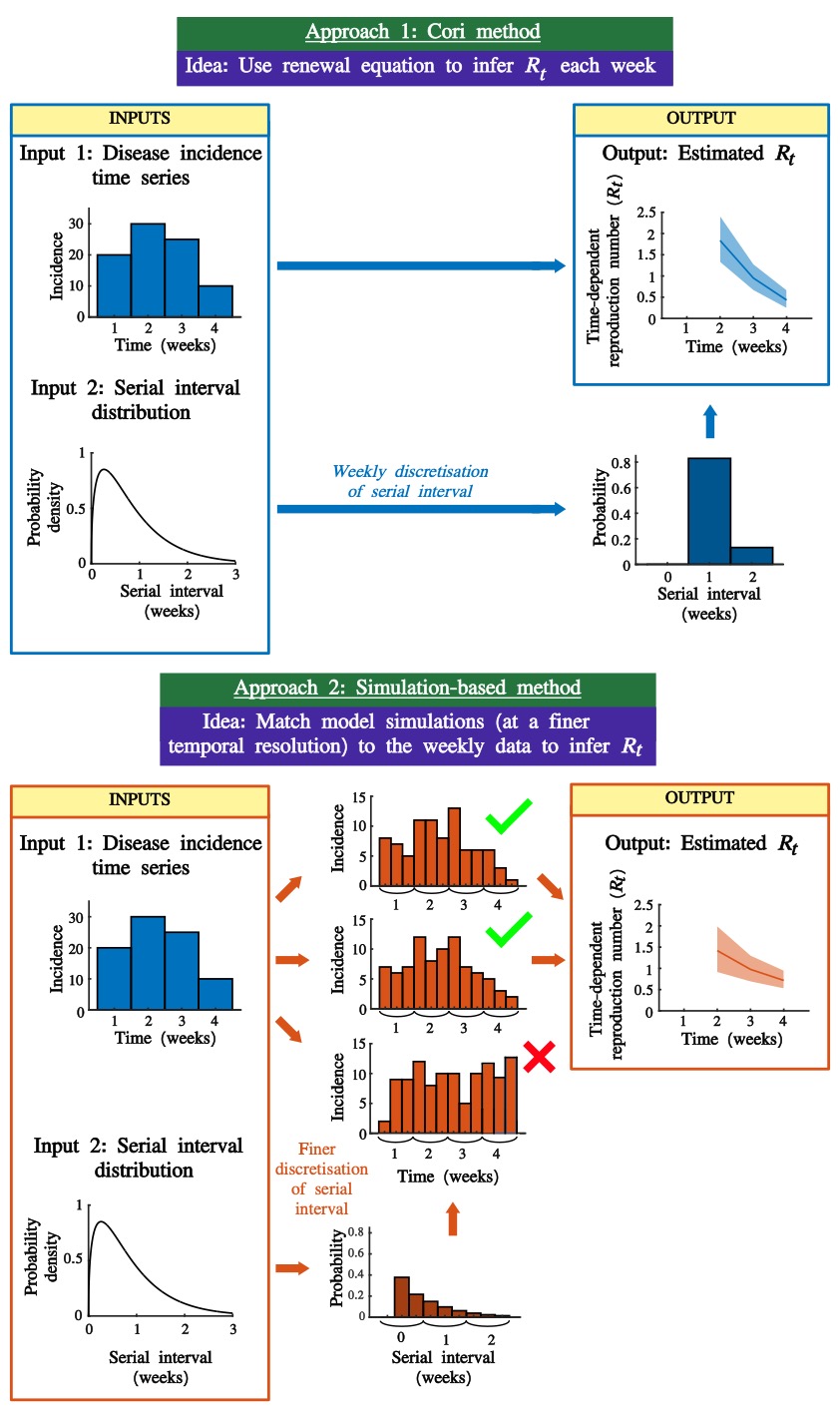
A commonly used method for estimating is the Cori method, implemented in the R software package *EpiEstim* [15] and the online application *EpiEstim App* [16]. This approach is based on a renewal equation model of pathogen transmission (see Methods) and involves estimation of from disease incidence time series and an estimate of the serial interval distribution (the probability distribution characterising the time between symptom onset times in infector-infectee transmission pairs). The Cori method has been extended in a range of ways following its original development, including accounting for imported cases [5,10,17,18], uncertainty in the serial interval distribution [5], superspreading [19,20], multiple pathogen variants [21] and unobserved generations of transmission [22].

However, a challenge that besets estimation of using the Cori method is temporal aggregation of disease incidence time series data [7,23]. For COVID-19, for example, many public health agencies switched from publishing daily numbers of reported cases to weekly summaries after the height of the pandemic [24]. Often, disease incidence is even reported weekly for pathogens including influenza [25] for which realised serial intervals and generation times are typically only a few days [26–28]. This is problematic not only because it is hard to unpick within-week changes in pathogen transmissibility when data are reported weekly, but also because an assumption of the transmission model underlying the Cori method is that all cases arising at timestep are generated by infectors from earlier timesteps. In other words, if the Cori method is applied with a weekly timestep, then it is assumed that an infector and infectee cannot both appear as cases in the disease incidence data in the same week. As the timescale of transmission (as characterised by the serial interval or generation time) of many pathogens is less than one week, this assumption is often incorrect.

To address this issue, in this research article we present a novel simulation-based method for estimating from temporally aggregated disease incidence time series data and the serial interval distribution. Our approach involves repeated simulation of a renewal equation transmission model for different values of with a timestep that is smaller than that of the disease incidence data. Using an iterative version of Approximate Bayesian Computation (ABC), we show how can be estimated in real-time during outbreaks by matching model 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 obtained using the Cori method. We go on to apply our method to real-world outbreak data from the 2019-20 and 2022-23 influenza seasons in Wales in the United Kingdom.

**Methods**

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



**Figure 1. Schematic illustrating the approaches for estimating that we consider.** Approach 1 involves the application of the commonly used Cori method to weekly aggregated disease incidence time series. Approach 2 is the novel simulation-based approach, which involves matching simulations run with a smaller timestep to the weekly aggregated data to estimate . The second approach relaxes the assumption that individuals appearing in the incidence data cannot have infected other individuals appearing in the same week. This is particularly important during outbreaks in which the typical timescale of transmission is shorter than the temporal aggregation of the data (e.g. if data are aggregated weekly, but serial intervals or generation times can be shorter than one week).

Cori method

Following previous descriptions of the Cori method [4,5,10], we assume that the expected number of cases, , in week , is given by

in which is the probability that the (weekly discretised) serial interval takes the value weeks. We use the notation to denote the vector of values of (. If the number of cases in week is drawn from a Poisson distribution, then the probability of observing weekly incidence over the time window [, (which consists of incidence data from weeks) is

The goal of the Cori method is to estimate , assuming that it takes a constant value during the time period from week to week . Assuming that the prior for is a gamma distribution with shape parameter and rate parameter , then the posterior for is

in which the notation represents the probability density of a gamma distribution at value with shape parameter and rate parameter . In all of our analyses, as in previous studies [4,5,10], we set and . The prior for therefore has mean and standard deviation equal to five. The large standard deviation is chosen so that the prior is relatively uninformative, while the high mean ensures that the outbreak is not evaluated as being under control () unless this is very likely to be the case.

Throughout the manuscript, we consider estimating individual values of each week, based on the numbers of new cases observed in that week. In other words, we assume that , in which case the above expression simplifies to

The Cori method can therefore be used to obtain a posterior for for weeks.

Simulation-based inference of

In the renewal equation model underlying the Cori method, the number of cases arising in week depends on the numbers of cases in previous weeks. Implicit in this approach is an assumption that individuals appearing in the incidence data in any week cannot generate new cases in the same week. When disease incidence data are temporally aggregated, so that the timescale of transmission is often shorter than the timestep in the incidence data, this assumption may be incorrect. To relax this assumption, we consider a novel simulation-based approach for estimating . The goal of this method is again to estimate the value of for each week, , but using a renewal equation model with a timestep that is shorter than one week (e.g., a daily timestep).

*Modified renewal equation*

In this approach, we consider partitioning the number of cases in each week into timesteps, where each new timestep is weeks. If, for example, , then we are using a daily timestep in the simulation-based method. We introduce the following notation:

* represents the number of cases in the th timestep within week .
* represents the probability that the serial interval, discretised into timesteps of length weeks (see below and Supplementary Material), takes the value timesteps.
* represents the vector of values of (.

In forward simulations of the corresponding renewal equation model, we assume that the number of cases in the th timestep of week is drawn from a Poisson distribution with mean

*Inference of*

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

This procedure is then repeated each week. Since this approach only involves obtaining matching simulations for a single week at a time, estimates of can be obtained relatively quickly (compared to attempting to match an entire simulation run over multiple weeks to the real-world data, as in standard ABC rejection sampling [29]). For a more detailed description of the simulation-based inference method, including an explanation of how cases are distributed between timesteps within the first week in each simulation, see the Supplementary Material. A schematic explaining the steps involved in the inference procedure is shown in Fig S1.

Outbreak datasets

We consider three outbreak datasets in our analyses. We first test our approach on a simulated dataset. This not only enables us to compare estimates of 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 generates accurate estimates in a setting in which the true value of (i.e., the value used to generate the simulated dataset) is known. 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)*

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 interval. Specifically, a disease incidence time series was generated starting from one initial case (in the first timestep) using a timestep of 10 minutes (). To generate a classic epidemic curve, the simulation was run for 11 weeks with for weeks and for weeks.

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

To demonstrate our approach on real-world data, we considered two disease incidence time series datasets provided by Public Health Wales describing estimated numbers of cases of influenza-like illness (ILI) in Wales each week. The original data comprised the clinical consultation rate per 100,000 individuals in sentinel practices in Wales each week [30]. The total number of weekly cases was then estimated by multiplying each value in the original data by 31.075 (i.e. scaling these values based on the population size of Wales, which is 3,107,500 [31]). 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.

Serial interval

Since we analyse influenza outbreak datasets in this study, we assume throughout that the (continuous) serial interval distribution is a gamma distribution with mean 0.37 weeks (2.6 days) and standard deviation 0.19 weeks (1.3 days) [32]. While this estimate was derived from household data for pandemic influenza, we expect it to be approximately in line with the serial interval for seasonal influenza (i.e., a mean value of less than one week). Denoting the probability density function of the serial interval distribution by , then .

We discretise this distribution into timesteps of length weeks to obtain . To do this, we adapt the method used by Cori *et al.* [4] in which the serial interval distribution is discretised into timesteps of length one. Specifically, we set

as derived in the Supplementary Material. We then choose so that is a valid probability distribution (i.e., the sum of the entries of is one). The rationale for normalising in this way is that same-timestep infections are not possible in the renewal equation model. Our approach involves assigning all probability density near zero in to , which is the shortest possible serial interval in the model.

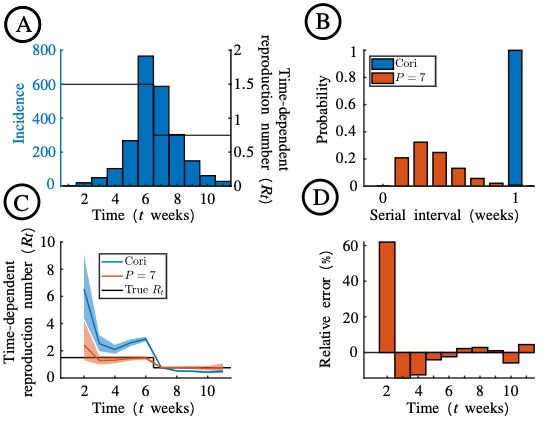
**Results**

Simulated dataset

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

We applied both inference methods to the simulated dataset, finding that the simulation-based approach generates more accurate estimates of than the Cori method in this scenario (Fig 2C). The percentage error in the estimated value of each week using the simulation-based approach (compared to value of used to generate the dataset) is shown in Fig 2D. These error values reflect stochasticity in the number of cases each week in the simulation, which is more substantial when there are fewer cases.

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 , the output is identical to when the Cori method is used to estimate (as would be expected, since the assumptions underlying the two methods are identical in this case; Fig S2A). We also considered how estimates obtained using the simulation-based method change when different values of are chosen (Fig S2B-D), finding that the method can obtain accurate estimates for relatively small values of (using a value of led to similar errors compared to using ).

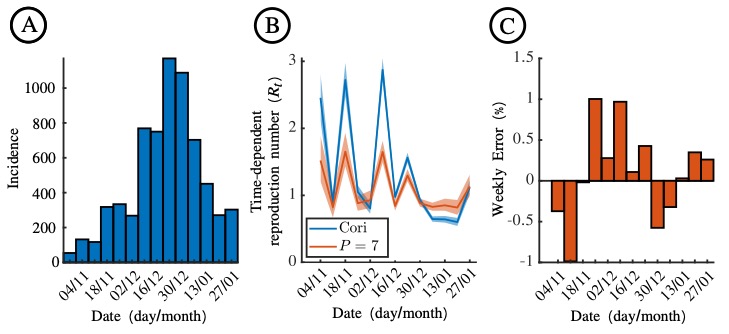


**Figure 2. Estimation of from the simulated disease incidence time series dataset.** A. The simulated outbreak dataset (blue bars), generated with for weeks and for weeks (black line). The outbreak is simulated with starting from one initial case in the first timestep, and new cases are then aggregated into weekly case counts. B. The discretised serial interval, for (blue) and (red). C. Estimates of using the Cori method (blue) and the novel simulation-based approach (with ; red). Blue and red lines are the mean estimate, and the shaded region represents the 90% credible interval. The value of underlying the simulation is shown in black. D. The percentage error in the estimate of each week using the simulation-based method.

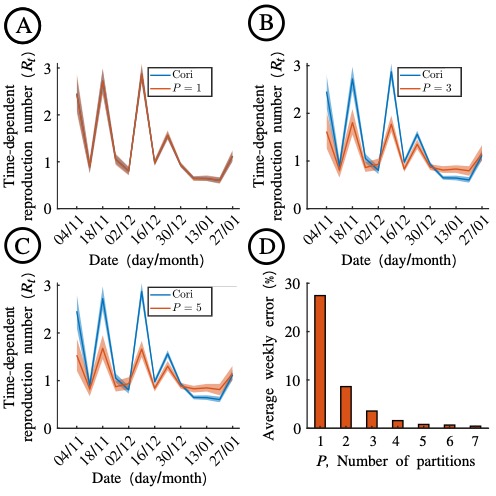
Influenza in Wales, 2019-20 and 2022-23

We then considered the two Wales influenza outbreak datasets, again using both the Cori and simulation-based methods to estimate . First, we considered the weekly case counts from the 2019-20 influenza season (Fig 3A). As with the simulated dataset, the simulation-based approach led to different estimates of than the Cori method. We computed the percentage error in the estimate each week using the simulation-based method with (Fig 3C). Since the true underlying value of was unknown, the percentage error was computed relative to applying the simulation-based method with a very large value of (this is representative of the best possible estimate of obtainable from the weekly incidence data; in this case, this is the inferred value of estimated with a one-hour timestep). We also explored how estimates depend on the value of that is used (Fig 4). Estimates obtained using the Cori method and using the simulation-based method with were identical (Fig 4A). We again found that a value of is large enough for accurate inference of (Fig 4D).

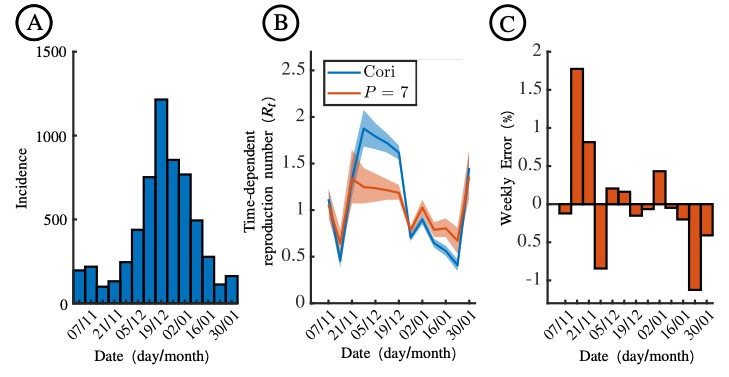
The analyses of data from 2019-20 were then repeated for analogous data for 2022-23, with similar results (Figs 5,6). Notably, in all of our analyses, the Cori method led to a higher estimate of than the simulation-based method when was estimated to be greater than one, and a lower estimate of than the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



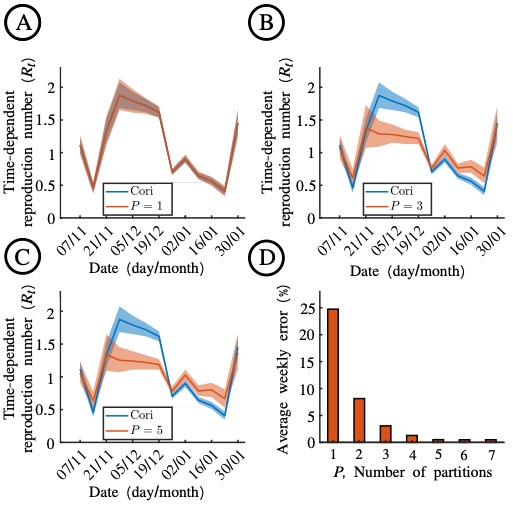
**Figure 3. Estimation of for influenza in Wales, 2019-2020.** A. Weekly numbers of ILI cases in Wales from 28 October 2019 to 2 February 2020, estimated from surveillance data collected in sentinel practices. B. Estimates of using the Cori method (blue) and the novel simulation-based approach (with ; red). Blue and red lines are the mean estimate, and the shaded region represents the 90% credible interval. C. The percentage error in the estimate of each week using the simulation-based method with , compared to using a larger value of (which corresponds to estimating with a one-hour timestep).



**Figure 4. Dependence of estimates using the simulation-based method on the value of used, for influenza in Wales, 2019-2020.** A. Estimates of obtained when the Cori method (blue) and the novel simulation-based approach with (red) are applied to the 2019-20 influenza dataset (Fig 3A). B. Analogous to panel A, but with in the simulation-based approach. C. Analogous to panel A, but with in the simulation-based approach. D. The average weekly error in estimates obtained using the simulation-based method with , compared to using a larger value of (which corresponds to estimating with a one-hour timestep).



**Figure 5. Estimation of for influenza in Wales, 2022-2023.** A. Weekly numbers of ILI cases in Wales from 31 October 2022 to 5 February 2023, estimated from surveillance data collected in sentinel practices. B. Estimates of using the Cori method (blue) and the novel simulation-based approach (with ; red). Blue and red lines are the mean estimate, and the shaded region represents the 90% credible interval. C. The percentage error in the estimate of each week using the simulation-based method with , compared to using a larger value of (which corresponds to estimating with a one-hour timestep).



**Figure 6. Dependence of estimates using the simulation-based method on the value of used, for influenza in Wales, 2022-2023.** A. Estimates of obtained when the Cori method (blue) and the novel simulation-based approach with (red) are applied to the 2022-23 influenza dataset (Fig 5A) B. Analogous to panel A, but with in the simulation-based approach. C. Analogous to panel A, but with in the simulation-based approach. D. The average weekly error in estimates obtained using the simulation-based method with , compared to using a larger value of (which corresponds to estimating with a one-hour timestep).

**Discussion**

During infectious disease outbreaks, evaluation of time-varying changes in pathogen transmission is essential to inform outbreak responses. Different metrics can be tracked, including incidence of new cases, hospitalisations and deaths, and outbreak growth rates [33,34]. An important metric that has been inferred in real-time during outbreaks of a range of pathogens is , in part because of its straightforward interpretation [7,9,14]. Not only is there a threshold value of , below which an outbreak can be inferred as being under control, but the value of also provides information about the extent to which the level of transmission must change (relative to current transmission) for an outbreak to grow or decline. For example, if , then more than half of transmissions must be prevented for the outbreak to decline. Similarly, if , then twice as many transmissions may occur before the outbreak begins to grow. Precise estimates of are therefore very important.

Here, we have presented a novel simulation-based approach for estimating in scenarios in which disease incidence time series data are aggregated temporally (Fig 1). While epidemiological data may be collected at a fine temporal resolution, it is common for the data to then be aggregated (e.g., into weekly or monthly counts). As we have shown, frequently used methods for inferring , such as the Cori method [4,5], may not generate accurate estimates when transmission occurs more rapidly than the temporal resolution of the aggregated data. This is because the renewal equation model underlying the Cori method involves assuming that an individual appearing in the time series data at timestep cannot have infected other individuals 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 ( timesteps for each timestep in the aggregated data) to the temporally aggregated data. The simulation-based approach not only provides accurate estimates of (Fig 2), but can also be applied easily to real-world data (Figs 3-6). While using a very large value of allows the most accurate possible estimates to be obtained from the aggregated data, even relatively small values of are sufficient for to be inferred accurately (Figs 4,6).

We found that, while the Cori method did not always provide an accurate estimate of due to the temporal aggregation of the disease incidence data, it was able to identify whether or not is below one (i.e., the outbreak is under control). While this is useful, as noted above precise estimation of is important as it provides information about the number of transmissions that must be prevented for an outbreak to be controlled. This result can be explained by the assumption of no same-timestep infections (i.e., infectors and infectees cannot appear in the disease incidence time series in the same timestep) in the renewal equation. When the Cori method is applied to weekly data, this then leads to overestimation of the serial interval, which is known in turn to lead to overestimation of if the true value of is greater than one (and vice versa) [35,36].

A closely related study by Nash *et al.* [23], undertaken at the same time as the analyses presented here, has also considered estimation of from temporally aggregated disease incidence time series data. In that approach, an expectation-maximisation algorithm is used to reconstruct daily incidence from weekly disease incidence data, and the Cori method is then applied to the estimated daily data. There are several differences between that approach and the simulation-based method described here. First, under the approach by Nash, only a single estimated daily disease incidence time series is obtained. In contrast, our method involves matching a range of simulations to the temporally aggregated data, thereby considering different possible daily disease incidence time series that could have led to the weekly aggregated data. Second, our method can be run straightforwardly for a range of values of , allowing the most accurate possible estimates of to be inferred from weekly incidence data (discretisation into a timestep of less than one day is possible). Third, our approach can be applied easily in scenarios in which the serial interval distribution is discrete rather than continuous (if, for example, the serial interval distribution is constructed directly from observations of dates on which infector-infectee pairs report symptoms). Fourth, our method can be adapted for scenarios in which the disease incidence time series data are aggregated into timesteps that are not all the same length. For example, when incidence data are derived from World Health Organization reports that are published irregularly in time, the timestep changes during the outbreak [37], and those irregular timesteps can be used directly in our simulation-based method. Finally, we note that our approach is conceptually straightforward, simply requiring repeated simulation of a renewal equation model.

Our simulation-based method is computationally efficient to run, as simulations are only required to match the real-world data for one aggregated timestep at a time. This is in contrast to using ABC rejection sampling to estimate all values of simultaneously, which would involve matching entire simulated time series to the entire real-world dataset. The efficiency of our approach allowed us to require that the simulations used to infer match the real-world data exactly. Further computational efficiency could be achieved by removing this condition, and instead setting a threshold “distance” within which a simulation is determined to match the real-world data, as is common when using ABC [29,38]. However, this necessitates that a distance metric is chosen, and resulting estimates of may be less accurate.

As in any modelling study, our framework in its current form involves a range of different assumptions. We followed previous publications in which the Cori method has been used [5] and assumed that the time series datasets from which we estimated represent numbers of new symptomatic cases each day. In the disease incidence time series data, it is then 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 positive values. However, in reality, realised serial intervals can be negative (if an infectee develops symptoms before their infector; this is not uncommon, for example, for transmission of SARS-CoV-2 [39–41]). Rather than using disease incidence time series, it is possible to apply both the Cori method and the simulation-based method to data describing incidence of infections, replacing the input serial interval distribution with the distribution of the generation time (the interval between infection times of infector-infectee pairs). This can be beneficial as realised generation times are always positive. However, since times of infection are often unknown, infection incidence data are typically not observed directly, so further inference is required to estimate numbers of infections and the generation time [8,42–44].

In our analyses, we assumed that all cases in the disease incidence time series (after the first timestep) arose as a result 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 [5,10,45,46], and under-reporting of cases is likely for many pathogens [47–49]. 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 [19,20], is another possible area for future work.

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

**COMPETING INTERESTS**

We have no competing interests.

**AUTHORS’ CONTRIBUTIONS**

IOG – formal analysis, investigation, validation, writing – original draft, writing – review and editing.

WSH – methodology, writing – review and editing.

JS – methodology, writing – review and editing.

RKN – methodology, writing – review and editing.

JP – methodology, writing – review and editing.

AC – methodology, writing – review and editing.

EMH – methodology, supervision, writing – review and editing.

RNT – conceptualization, methodology, project administration, supervision, writing – original draft, writing – review and editing.

**FUNDING**

This research was funded by the EPSRC through the Mathematics for Real-World Systems CDT (ZO-G, RNT; grant number EP/S022244/1) and a doctoral prize (WSH; grant number EP/W524311/1). The collaboration between JS and RNT was funded by a grant from Public Health Wales.

**ACKNOWLEDGEMENTS**

Thanks to members of the Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research at the University of Warwick and the Wolfson Centre for Mathematical Biology at the University of Oxford for useful discussions about this research.

**DATA AVAILABILITY**

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

**References**

1. 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.

2. Smith RD, Keogh-Brown MR, Barnett T, Tait J. The economy-wide impact of pandemic influenza on the UK: a computable general equilibrium modelling experiment. BMJ. 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.

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.

6. Nishiura H, Chowell G. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. Math Stat Estim App Epidem. 2009. pp. 103–121.

7. Nash RK, Nouvellet P, Cori A. Real-time estimation of the epidemic reproduction number: Scoping review of the applications and challenges. PLOS Digit Health. 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.

9. Vegvari C, Abbott S, Ball F, Brooks-Pollock E, Challen R, Collyer BS, et al. Commentary on the use of the reproduction number R during the COVID-19 pandemic. Stat Meth Med Res. 2021;1: 1–11.

10. Creswell R, Augustin D, Bouros I, Farm HJ, Miao S, Ahern A, et al. Heterogeneity in the onwards transmission risk between local and imported cases affects practical estimates of the time-dependent reproduction number. Phil Trans R Soc A. 2022;380: 20210308.

11. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One. 2007;2: e758.

12. Dai C, Zhou D, Gao B, Wang K. A new method for the joint estimation of instantaneous reproductive number and serial interval during epidemics. PLOS Comput Biol. 2023;19: e1011021.

13. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol. 2004;160: 509–516.

14. White LF, Moser CB, Thompson RN, Pagano M. Statistical estimation of the reproductive number from case notification data. Am J Epidem. 2020; kwaa211.

15. EpiEstim Team. EpiEstim: Estimate time varying reproduction numbers from epidemic curves. Version 2.2-4. 2021. Available: www.cran.r-project.org/web/packages/EpiEstim/

16. EpiEstim App Team. EpiEstim App. 2019. Available: www.shiny.dide.imperial.ac.uk/epiestim/

17. 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.

18. Tsang TK, Wu P, Lau EHY, Cowling BJ. Accounting for imported cases in estimating the time-varying reproductive number of COVID-19 in Hong Kong. J Infect Dis. 2021;224: 783–787.

19. 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.

20. Ho F, Parag KV, Adam DC, Lau EHY, Cowling BJ, Tsang TK. Accounting for the potential of overdispersion in estimation of the time-varying reproduction number. Epidemiology. 2023;34: 201–205.

21. 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 case-study. Epidemics. 2023;44: 100692.

22. 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.

23. Nash RK, Cori A, Nouvellet P. Estimating the epidemic reproduction number from temporally aggregated incidence data: a statistical modelling approach and software tool. medRxiv. 2023.

24. UK Health Security Agency. The COVID-19 dashboard moves to weekly updates. 2022. Available: www.ukhsa.blog.gov.uk/2022/06/28/the-covid-19-dashboard-moves-to-weekly-updates/

25. UK Health Security Agency. National Influenza and COVID-19 surveillance report: Week 29 report (up to week 28 data). 2023. Available: www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports-2023-to-2024-season

26. Cowling BJ, Fang VJ, Riley S, Malik Peiris JS, Leung GM. Estimation of the serial interval of influenza. Epidemiology. 2009;20: 344–347.

27. 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.

28. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect Dis. 2014;14: 480.

29. Minter A, Retkute R. Approximate Bayesian Computation for infectious disease modelling. Epidemics. 2019;29: 100368.

30. Public Health Wales. Weekly influenza and acute respiratory infection surveillance report: Wednesday 22nd February 2023 (covering week 07 2023). 2023. Available: www.phw.nhs.wales/topics/immunisation-and-vaccines/fluvaccine/weekly-influenza-and-acute-respiratory-infection-report/october-2022-october-2023-flu-season-202223/phw-influenza-surveillance-report-for-2023-week-7pdf/

31. Office for National Statistics. Population and household estimates, Wales: Census 2021. 2021. Available: www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationandhouseholdestimateswales/census2021

32. 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.

33. 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.

34. Parag KV, Thompson RN, Donnelly CA. Are epidemic growth rates more informative than reproduction numbers? J R Stat Soc Ser A. 2022;1: 1–11.

35. 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.

36. 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.

37. 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.

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.

39. Du 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 COVID-19 symptom onset highlights the importance of continued contact tracing. eLife. 2021;10: e65534.

41. 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.

42. 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.

43. 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.

44. 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.

45. Daon Y, Thompson RN, Obolski U. Estimating COVID-19 outbreak risk through air travel. J Travel Med. 2020;27: taaa093.

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

47. 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.

48. Albani V, Loria J, Massad E, Zubelli J. COVID-19 underreporting and its impact on vaccination strategies. BMC Infect Dis. 2021;21: 1111.

49. Gibbons CL, Mangen M-JJ, Plass D, Havelaar AH, Brooke RJ, Kramarz P, et al. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. BMC Public Health. 2014;14: 147.