Journal of Complex Networks, 2025, **4**, cnaf018
https://doi.org/10.1093/comnet/cnaf018
Advance access publication 15 July 2025
Article

OXFORD

Inferring contact network characteristics from epidemic data via compact mean-field models

Andrés Guzmán[1], Federico Malizia[1], Gyeong Ho Park[2], Boseung Choi[2,3,4,*], Diana Cole[5], and István Z. Kiss[1,6,*]

[1]Network Science Institute, Northeastern University London, 58 St Katharine's Way, London E1W 1LP, United Kingdom
[2]Department of Big Data Science, Korea University Sejong Campus, 2511 Sejong-ro, Sejong 30019, South Korea
[3] Biomedical Mathematics Group, Institute for Basic Science, 55 Expo-ro, Yuseong-gu, Daejeon 34126, South Korea
[4]College of Public Health, The Ohio State University, 1735 Neil Ave, Columbus OH 43210, United States
[5]School of Engineering, Mathematics and Physics, University of Kent, Cornwallis Buildings, Canterbury CT2 7NZ, United Kingdom
[6]Department of Mathematics, Northeastern University, 360 Huntington Ave, Boston MA 02115, United States

[*]Corresponding authors. Boseung Choi, Department of Big Data Science, Korea University Sejong Campus, 2511 Sejong-ro, Sejong 30019, South Korea. E-mail: cbskust@korea.ac.kr; István Z. Kiss, Network Science Institute, Northeastern University London, 58 St Katharine's Way, London E1W 1LP, United Kingdom. E-mail: istvan.kiss@nulondon.ac.uk.

ABSTRACT

Modelling epidemics using contact networks provides a significant improvement over classical compartmental models by explicitly incorporating the network of contacts. However, while network-based models describe disease spread on a given contact structure, their potential for inferring the underlying network from epidemic data remains largely unexplored. In this work, we consider the edge-based compartmental model, a compact and analytically tractable framework, and we integrate it within dynamical survival analysis to infer key network properties along with parameters of the epidemic itself. Despite correlations between structural and epidemic parameters, our framework demonstrates robustness in accurately inferring contact network properties from synthetic epidemic simulations. Additionally, we apply the framework to real-world outbreaks—the 2001 UK foot-and-mouth disease outbreak and the COVID-19 epidemic in Seoul—to estimate both disease parameters and network characteristics. Our results show that our framework achieves good fits to real-world epidemic data and reliable short-term forecasts. These findings highlight the potential of network-based inference approaches to uncover hidden contact structures, providing insights that can inform the design of targeted interventions and public health strategies.

KEYWORDS: epidemics; inference; contact networks.

1. INTRODUCTION

The spread of infectious diseases is inherently tied to the structure of human interactions. Network theory provides a powerful framework for understanding how diseases propagate by capturing the complex web of contacts between individuals [1–5]. Studies have highlighted how structural properties such as heterogeneity [6–10], communities [11–14], clustering [15–18], and degree correlations [19–22] play a significant role in shaping epidemic dynamics.

Received: 17 February 2025. **Editorial decision:** 10 June 2025. **Accepted:** 17 June 2025

Epidemic models have traditionally been used to describe and predict disease spread based on assumptions about the underlying contact structure [2, 23]. Ultimately, their applicability to real-world processes depends on the availability and quality of data [24]. These models range in complexity, from classical mass-action approaches, where populations are assumed to mix homogeneously [25], to sophisticated network-based frameworks that explicitly incorporate individual-level connectivity patterns [26–28]. Simpler models offer easier tractability but may overlook key structural features, while more complex models provide richer descriptions but require more detailed input data [24, 29–31]. Finding a balance between these aspects is essential for effective epidemic modelling [32].

A particularly elegant and efficient modelling framework is the edge-based compartmental model (EBCM), which provides a compact yet powerful representation of epidemic processes on networks [33–35]. Unlike standard compartmental models where incorporating heterogeneity significantly increases model complexity, EBCM encodes network structure and characteristics through probability-generating functions, allowing epidemic dynamics to be described with only a few parameters and a reduced number of equations.

In many real-world scenarios, direct measurements of contact networks are unavailable or incomplete. Although collecting data from contact networks is feasible in certain cases, such as sexually transmitted infections [36–39], it remains challenging for respiratory diseases [40–42]. While epidemic models are often used to simulate outbreaks given a known network structure, inferring the structure of the contact network from observed epidemic data represents an equally important challenge [43–47]. Since spreading dynamics inherently reflect network properties, they can be used to extract valuable information about the underlying structural information. Various methods have been proposed to reconstruct networks from data, including likelihood-based optimization approaches [48–52] and Bayesian inference techniques [53–57]. However, these methods often require detailed temporal data or strong prior assumptions, making them difficult to apply in real-world epidemic surveillance [58].

An alternative approach, Dynamical Survival Analysis (DSA), has been introduced to estimate epidemic parameters using infection and recovery time distributions [59]. Originally developed for mass-action models [60, 61], DSA was recently extended to network-based models [62], enabling parameter estimation while incorporating some aspects of network structure. However, existing applications remain limited in their ability to fully capture the heterogeneity of contact networks.

In this paper, we integrate the DSA approach with the EBCM [33] to develop a Bayesian framework for inferring both disease and network parameters from epidemic data. This extends previous works [63], shifting from identifiability analysis to active inference in both synthetic and real-world scenarios. The manuscript is structured as follows: Section 2.1 introduces the EBCM framework, and Section 2.2 details the Bayesian inference procedure. Section 3 presents validation on synthetic and real data, specifically the first wave of COVID-19 in Seoul and the 2001 foot-and-mouth disease epidemic in the UK. Finally, Section 5 discusses the implications of our findings.

2. METHODS

In this section, we outline the methodologies that form the foundation of our inference framework. First, we introduce the EBCM, which provides a compact representation of SIR processes on networks. This model serves as the backbone for describing the epidemic dynamics in structured populations.

Next, we present the complete inference process, detailing how these methods are integrated to estimate both epidemic and network parameters from observed outbreak data. Specifically, we employ Dynamic Survival Analysis (DSA) to construct the likelihood function, leveraging its ability to handle censored and aggregated epidemic data. Moreover, we describe the Robust Adaptive Metropolis (RAM) algorithm, a Markov Chain Monte Carlo (MCMC) technique designed for efficient exploration of the parameter space. RAM adapts to the local structure of the posterior distribution, improving convergence and robustness in high-dimensional settings. Together, these

methods form a comprehensive framework for inferring epidemic dynamics and network structures from real-world outbreak data.

2.1 Edge-based compartmental model

We consider a Susceptible-Infected-Recovered (SIR) epidemic process, where individuals can be in one of three states: susceptible (S), infected (I), or recovered (R). Infection occurs at rate β along a link between a susceptible and an infected node, while infected nodes recover independently of the network at rate γ. In this study, we employ the EBCM [33], which provides a compact and analytically tractable representation of epidemic dynamics on contact networks. The EBCM assumes that disease transmission occurs on a network generated by the configuration model (CM) [64, 65], which is characterized by a degree distribution $P(k)$. The key idea behind EBCM is to track the probability that a randomly chosen node remains susceptible rather than explicitly tracking individual infection events.

A central variable in the model is $\theta(t)$, defined as the probability that a randomly selected neighbor of a test node u has not transmitted the disease to u by time t. From now on, we omit the obvious time dependence. Given that a node u has degree k, the probability that it remains susceptible is $s_u(k,\theta) = \theta^k$. Thus, the overall fraction of susceptible nodes in the population is given by:

$$S(t) = \sum_k P(k)\theta^k = \Psi(\theta), \tag{1}$$

where $\Psi(\theta)$ represents the probability generating function (PGF).

If a fraction ρ of the population is initially infected at $t = 0$, we modify this expression as $S(t) = \hat{\Psi}(\theta) = \sum_k P(k)S(k,0)\theta^k$, where $S(k,0)$ is the probability that a node with degree k is initially susceptible. Since initially infected nodes are selected at random, it follows that $S(k,0) = 1-\rho$. To fully characterize the system, we decompose θ into three probabilities, namely $\theta = \psi_S + \psi_I + \psi_R$, where ψ_S, ψ_I, and ψ_R denote the probabilities that a randomly selected neighbor of node u is, respectively, in the susceptible state at time t; infected but has not yet transmitted the disease to u by time t; or recovered without having transmitted the infection to u during their infectious period. Note that $\dot{\theta} = -\beta\psi_I$, where β represents the rate at which an infected partner transmits the disease to the test node. Furthermore, we can express $\psi_I = \theta - \psi_S - \psi_R$, which leads to $\dot{\theta} = -\beta(\theta - \psi_S - \psi_R)$. Additionally, we express ψ_R and ψ_S as $\psi_R = \psi_R(0) + \gamma(1-\theta)/\beta$ and $\psi_S = \psi_S(0)\hat{\Psi}'(\theta)/\langle k\rangle$, where $\hat{\Psi}'(\theta)$ denotes the derivative of the probability generating function (PGF) with respect to θ. Additionally, the average degree can be defined as $\langle k\rangle = \sum_k kP(k)S(k,0)$ which is also equivalent to the derivative of the PGF evaluated at $\theta = 1$. Finally, $\psi_R(0)$ and $\psi_S(0)$ are the probabilities of the test node being initially connected to a recovered or susceptible node respectively. Further details on the derivation of these expressions can be found in the supplementary material or in the original papers [1, 33]. By expressing ψ_R and ψ_S as functions of θ, we can redefine $\dot{\theta}$ as a differential equation that depends only on θ, β, γ, and the initial condition. With these considerations, the model is fully described by the following system of equations:

$$\begin{aligned} \tfrac{d\theta}{dt} &= -\beta\theta + \beta\psi_S(0)\tfrac{\hat{\Psi}'(\theta)}{\langle k\rangle} + \gamma(1-\theta) + \beta\psi_R(0), \\ \tfrac{dR}{dt} &= \gamma(1-S-R), \quad S = \hat{\Psi}(\theta). \end{aligned} \tag{2}$$

Typically, we assume $\psi_R(0) = 0$ and $\psi_S(0) = 1-\rho$. Solving Equations (2) provides the evolution of $S(t)$, $I(t)$, and $R(t)$. Moreover, the basic reproductive number (R_0) of the EBCM is defined as

$$R_0 = \frac{\beta}{\beta+\gamma}\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}, \tag{3}$$

Table 1. Details of the probability generating functions used throughout the paper[a,b]

	Poisson	Negative binomial
Parameter(s)	μ	(r, μ)
$\Psi(x)$	$e^{\mu(x-1)}$	$\left(\frac{r}{r+\mu(1-\theta)}\right)^r$
$\Psi'(x)$	$\mu e^{\mu(x-1)}$	$\mu\left(\frac{r}{r+\mu(1-\theta)}\right)^{r+1}$

[a]The parameter μ for both distributions corresponds to the average degree given by $\sum_k kP(k)$

[b]Table comparing the probability generating functions (PGFs) and their derivatives for Poisson and Negative Binomial degree distributions. It includes the associated parameters and analytical forms of $\Psi(x)$ and $\Psi'(x)$ used in the study.

where $\langle k^2\rangle - \langle k\rangle = \sum_k k(k-1)P(k)S(k,0)$, corresponding to the derivative of the PGF evaluated at $\theta = 1$. For simplicity, from now on, we apply a change of variable $\mu \equiv \langle k\rangle$.

In this study, we consider two different degree distributions, which are summarized in Table 1 along with their parameters and probability generating functions. The Poisson distribution is characterized by a single parameter μ, which defines both its mean and variance, resulting in a relatively homogeneous degree distribution. In contrast, the Negative Binomial distribution, parametrized by μ, which represents the average degree, and r, which, together with μ, determine the variance of the distribution. In particular, smaller values of r lead to greater overdispersion. This flexibility makes the Negative Binomial distribution well-suited for modeling both homogeneous and heterogeneous network structures.

2.2 Statistical inference framework

Accurate parameter estimation in epidemic modelling usually relies on optimizing a likelihood function that reflects both the underlying transmission dynamics and the nature of the available data. A common approach involves fitting model-generated epidemic curves to observed data by minimizing discrepancies between them. However, this method is highly sensitive to noise, biases, and incomplete datasets, which can compromise inference accuracy. To address these challenges, we employ the *Dynamic Survival Analysis* (DSA) framework [59, 66–68], which provides a more robust approach by directly incorporating individual transition times between epidemic states into the likelihood function.

DSA was developed to overcome the limitations of traditional inference methods in infectious disease epidemiology by integrating dynamical systems theory with survival analysis techniques. Unlike conventional approaches that rely on aggregate epidemic curves, DSA leverages the mean-field ordinary differential equations (ODEs) governing population-level dynamics to model the probability distributions of transition times, such as the time of infection or recovery. This formulation allows DSA to construct likelihood functions for individual-level trajectories, making it particularly effective in handling censored, truncated, or incomplete data. In this framework, the susceptible fraction of the population, $S(t)$, is reinterpreted as a survival function, satisfying $S(0) = 1$. More generally, when a fraction ρ of individuals is initially infected, we introduce three rescaled survival functions, which are defined as

$$\tilde{S}(t) = \frac{S(t)}{1-\rho} = \Psi(\theta), \qquad \tilde{I}(t) = \frac{I(t)}{1-\rho}, \qquad \text{and} \qquad \tilde{R}(t) = \frac{R(t)}{1-\rho}. \tag{4}$$

By substituting Equation (4) in the system of equations for the EBCM, as given by Equations (2), we have

$$\dot{\tilde{S}}(t) = \frac{d\tilde{S}}{d\theta}\frac{d\theta}{dt} = \Psi'(\theta)\dot{\theta} = \Psi'(\theta)\Big[-\beta\theta + \beta(1-\rho)\frac{\hat{\Psi}'(\theta)}{\langle k\rangle} + \gamma\,(1-\theta)\Big]. \tag{5}$$

$$\dot{\tilde{R}}(t) = \gamma\,\tilde{I}(t) \qquad \text{and} \qquad \tilde{I}(t) = 1/(1-\rho) - \tilde{S}(t) - \tilde{R}(t),$$

where, at $t = 0$ we have $\tilde{S}(0) = 1, \tilde{I}(0) = \rho/(1-\rho)$ and $\tilde{R}(0) = 0$.

DSA interprets the susceptible curve as an improper survival function representing the time of infection for a randomly chosen initially susceptible individual. That is, $\tilde{S}(t) = \mathsf{P}(T_I > t)$, where the random variable T_I denotes the infection time. The density function of T_I is given by $-\dot{\tilde{S}}(t)$, which is improper since $\lim_{t\to\infty}\tilde{S}(t) = \mathsf{P}(T_I = \infty) > 0$. We define $\mathsf{P}(T_I = \infty) = 1 - \tau$, where τ represents the final epidemic size. To obtain a proper survival function, we condition it on a final observation time $T \in (0, \infty)$ and the final epidemic size τ at time T. The resulting probability density function $f_\tau(t)$ on the interval $[0, T]$ is then given by:

$$f_\tau(t) = -\frac{\dot{\tilde{S}}(t)}{\tau}. \tag{6}$$

Note that DSA does not require knowledge of recovery times. However, if these times are available, they can be incorporated to enhance the quality of inference. Let T_R represent the time of recovery of an infected individual. Given the infection time T_I, the infectious period $T_R - T_I$ follows an exponential distribution with rate γ. Using Equation (6) and the density of the infectious period, we can define the density of the recovery time T_R as:

$$g(t) = \int_0^t f_\tau(u)\gamma\, e^{-\gamma\,(t-u)}\mathrm{d}u. \tag{7}$$

Equation (7) represents the convolution of the density of the infection time $f_\tau(t)$ and an exponential distribution with rate γ, corresponding to the density of the infectious period. In practice, solving the system of ODEs (5) with respect to the observed recovery times is computationally more convenient.

Finally, the normalized density of the recovery time is given by:

$$\tilde{g}(t) = \frac{g(t)}{\int_0^T g(t)\mathrm{d}t}. \tag{8}$$

One of the advantages of the DSA method is the ability to build various likelihood functions based on the observed data. Let N be the size of the population and M be the initial number of infected individuals at the beginning time $t = 0$, and usually $N >> M$. We have K individuals out of N, who are infected by time T and $t_1, t_2, \ldots, t_K$ represent the time of infection for each infected individual. If L individuals have recovered by time T out of a total of $K + M$ infected, let $r_1, r_2, ., r_L$ denote the time of recovery. If we can observe the time of infection and recovery exactly for each individual, then we can define the infectious period, $w_i = r_i - t_i$ or $w_i = r_i$ for initially infected, $i = 1, 2, \ldots, L$ respectively. Also, we could have $\tilde{L}$ infected individuals who have not recovered by time T, $\epsilon_j = T - t_j$ or $\epsilon_j = T, j = 1, 2, \ldots, \tilde{L}$, denote the censored infectious period respectively.

Given random samples of time of infection, $t_1, t_2, \ldots, t_K$,the log likelihood function is given by:

$$\ell_1 = K\log(\tau) + (N - (M+K))\log(1-\tau) + \sum_{i=1}^{K}\log(f_\tau(t_i)). \tag{9}$$

The log likelihood function for the infectious period, $w_1, w_2, \ldots, w_L$ is given by

$$\ell_2 = L\log(\gamma) - \gamma \sum_{j=1}^{L} w_j. \tag{10}$$

The log likelihood function for the time of recovery, $r_1, r_2, \ldots, r_L$ is given by

$$\ell_3 = \sum_{j=1}^{L} \log(\tilde{g}(r_j)). \tag{11}$$

Finally, we can define the log likelihood for the censored infectious period, $\epsilon_1, \epsilon_2, \ldots, \epsilon_{\tilde{L}}$;

$$\ell_4 = -\gamma \sum_{k=1}^{\tilde{L}} \epsilon_k. \tag{12}$$

In some cases, we only know the number of infected individuals, K, by a given time, T, but lack information about the total population size N. This situation is common in real epidemic scenarios, where data about the segment of the population at risk of infection is often unavailable. For such cases, the likelihood ℓ_1 can be reformulated as: $\ell_1 = \sum_{i=1}^{K} f_\tau(t_i)$, excluding terms related to the total population size. With this formulation, we can analyze the dynamics of the proportion of infected, susceptible, and recovered individuals. Furthermore, it is possible to estimate an effective population size using the following equation

$$N_{\mathit{eff}} = \frac{K}{1 - S_T}. \tag{13}$$

Depending on the data and particular context, we can use any of the four likelihoods described above or a combination of them. For example, suppose we just observe K infections before a cut-off time T, then we could just use ℓ_1 (9). If, besides the infection times, we also observe L recovery times but not the specific nodes who underwent these changes, then we could use the combination of $\ell_1 + \ell_3$, see Equations (9) and (11). However, if the data is available at the node level and we have pairs of infection and recovery times for each node, we could use the combination $\ell_1 + \ell_2$, see Equations (9) and (10), since we can calculate the infectious period for each node. The likelihood functions described above do not explicitly incorporate the network parameters, making it impossible to derive closed-form solutions for their maximum likelihood estimates. To address this limitation, we adopt a Bayesian approach, which allows us to sample from the posterior distribution and derive point estimates from it. To construct the sample, we employ a Markov Chain Monte Carlo (MCMC) method, specifically the Robust Adaptive Metropolis (RAM) algorithm. The RAM algorithm is more efficient than the standard Metropolis-Hastings algorithm [69], as it dynamically adjusts the variance-covariance matrix of the proposal distribution to maintain an optimal acceptance rate during the Metropolis steps. We assign vaguely informative prior distributions (little information is given about the parameters to be estimated) to the model parameters: a Gamma distribution, Gamma(a, b), for the β and γ parameters, which is defined as

$$\text{Gamma}(x, a, b) \sim \frac{b^a x^{a-1} e^{-bx}}{\Gamma(a)}, \tag{14}$$

where $\Gamma(a)$ represents the gamma function. Additionally, for the parameter ρ in the SIR model, we assume a Beta distribution, Beta(a, b), where, in this context, $a = 1$ and $b = 1$ are chosen from a

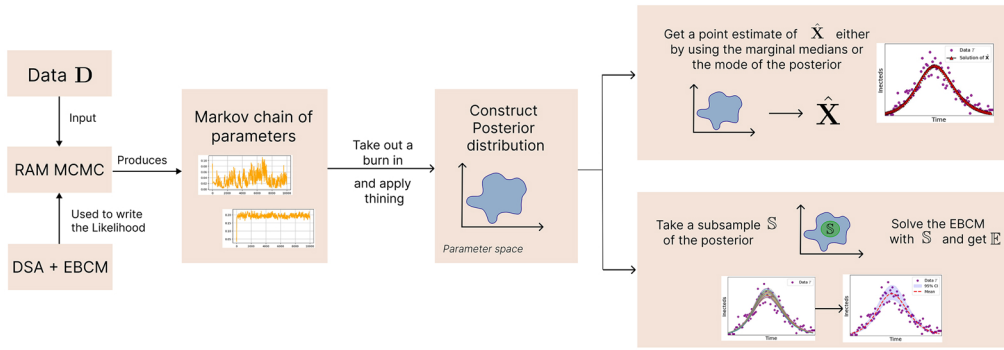

Figure 1. Graphical representation of the inference process. The procedure starts with epidemic data **D** as input for the Robust Adaptive Metropolis (RAM) algorithm, which employs the likelihood function derived from the DSA framework (see main text). The RAM algorithm generates a Markov chain of parameter samples, forming the posterior distribution. From this distribution, we obtain parameter estimates using two approaches: credible intervals and point estimates, as detailed in the main text.

uniform distribution. Additionally, we assign a non-informative Gamma prior to the parameters of the degree distribution based on the support of these parameters.

Given a dataset **D**, our goal is to fit the data with the EBCM, which is characterized by the set of parameters $\mathbf{X} = (\beta, \gamma, \rho, \Delta)$. Here, we use the vector Δ to represent the parameters of the probability generating function of the network's degree distribution. Our objective is to infer the parameter set **X** such that the output from the EBCM matches the observed data. To achieve this, we utilize the RAM algorithm to generate one or more chains for the values of **X**, which allows us to construct a sample of the posterior distribution. From this sample, we remove the burn-in and apply a thinning procedure to reduce autocorrelation. The inference scheme is illustrated in the left panels of Fig. 1.

Once a posterior distribution is obtained, we can find a point estimate for each parameter. This can be done by either calculating the *marginal mean*, *marginal median* or taking the *joint mode* of the posterior distribution, i.e. the point of highest probability in the full posterior. As shown in the top right panel in Fig. 1. Either of these can be used in conjunction with the EBCM to produce one single epidemic curve that can be compared to the data. However, to generate a credible interval around this single epidemic curve, we take a subsample $\mathbb{S}$ from the full posterior and use each element to solve the EBCM, thereby generating a set of solutions, as shown in the bottom right panel of Fig. 1. These can used to generate the desired credible interval over a specified period of time.

In subsequent sections, we show the application of this workflow to estimate network and epidemic dynamics parameters for two different scenarios. First, by the use of synthetic data corresponding to a controlled case where ground truth is available. Second, we consider two different datasets corresponding to real-world epidemics.

3. INFERENCE FROM SYNTHETIC DATA

We begin by analyzing synthetic data to verify the method's accuracy and effectiveness in a controlled environment. First, we consider 100 realizations of Gillespie simulations [70] with $\beta = 0.2$, $\gamma = 1$, on networks with 10^4 nodes exhibiting Poisson ($POI(\mu = 10)$) and Negative Binomial ($NB(\mu = 10, r = 1)$) degree distributions. Since in a real epidemic process, complete data are rarely available, we consider four possible scenarios which reflect varying data availability, which are:

- $\ell_1 + \ell_2$: infection times $(t_1, t_2, \ldots, t_K)$ and infection periods $(w_1, w_2, \ldots, w_L)$

- $\ell_1 + \ell_3$: list of infection times $(t_1, t_2, \ldots, t_K)$ and a decoupled list of recovery times $(r_1, r_2, \ldots, r_L)$.
- ℓ_1: infection times only $(t_1, t_2, \ldots, t_K)$.
- ℓ_3: recovery times only $(r_1, r_2 \ldots r_L)$.

Likelihood ℓ_4 is not used for the synthetic data, as the cut-off time is set after the end of the epidemic. However, this likelihood is used in Section 4.2.

Following the procedure outlined in Section 2.2, we fit our models to all stochastic realizations. Initially, we consider the case where the data are fitted using the correct model. Specifically, data from simulations on networks with a Poisson degree distribution were fitted using the Poisson (Poi) model in the EBCM; we refer to this scenario as Poisson-Poi. Similarly, a match with the Negative Binomial is denoted by Negbin-NB. We also consider model mismatch, where data from networks with a Poisson degree distribution were fitted using the Negative Binomial (NB) model (Poisson-NB), and vice-versa, that is Negbin-Poi.

For all four possible combinations (Negbin-NB, Poisson-Poi, Negbin-Poi, Poisson-NB), we used the four different scenarios discussed above. Then, for each stochastic realization, we construct a sample of the posterior distribution and calculated the marginal median of β, γ, and μ, and computed the basic reproduction number as defined in Equation (3). The point estimates for $\hat{\beta}$, $\hat{\gamma}$, $\hat{\mu}$, $\hat{R}_0$ are obtained as the mean values of the distributions of the marginal medians. In Fig. 2, we show the density of the marginal medians and their mean, i.e. point estimates, for these four parameters and for both model match and mismatch. Additionally, we present these results for the four different data availability scenarios presented above.

As expected, the correct matches, Negbin-NB, first column, and Poisson-Poi, second column, yield accurate estimations for all parameters and for all different likelihoods. In the presence of data and model mismatch, we observe that fitting Poisson data with the Negative Binomial model (fourth column) often yields reasonable results. This can be attributed to the flexibility of the Negative Binomial distribution, i.e. one more free parameter compared to Poisson, which allows it to capture both homogeneous and heterogeneous degree distributions. In contrast, fitting data generated with the Negative Binomial with the Poisson model underperforms. This is especially evident in the estimates for β and μ in the first and third rows, where the EBCM with a Poisson degree distribution tends to overestimate β and underestimate μ.

It is noteworthy that the mismatched Negbin–Poi model produced better estimates for β and μ in the $\ell_1 + \ell_3$ and ℓ_3 scenarios. However, even in these cases, γ was substantially overestimated. These qualitative observations are further supported by the larger bias and mean squared error (MSE) of the estimates, as detailed in Supplementary Material (SM). For the estimation of γ, scenario $\ell_1 + \ell_2$ uses the infectious periods as data. Since the infectious period directly reflects γ (the inverse of the mean infectious period), this scenario yields highly accurate estimates, even for mismatched models and data. By contrast, estimates of γ exhibit a larger bias in mismatched scenarios $\ell_1 + \ell_3$ and ℓ_3, where the recovery times are used instead, but they do not align with the infection times.

Interestingly, in the mismatch cases, in particular, see the panel on the third row and third column in Fig. 2, we note that the estimates of μ based on scenario $\ell_1 + \ell_2$ perform poorly when compared to estimates based on scenarios ℓ_1 and ℓ_3. This, at first, is surprising since $\ell_1 + \ell_2$ provides the most complete data. However, due to the model mismatch manifesting itself mainly in the infection part, this is likely to lead to a better estimation of γ, which in turn has a negative effect on the estimation of β and μ. In contrast, in scenarios ℓ_1 and ℓ_3, the estimation of μ improves at the expense of obtaining less accurate estimates of γ. These observations are further supported by the MSE values presented in SM, where this analysis is discussed in detail.

Even though accurate and precise estimations can be obtained using the marginal medians of the posterior distributions, it is important to recognize that the posterior distribution is multivariate. This means that the marginals are not always fully representative of the entire distribution, especially when there is a correlation between parameters. When used for inference, network-based mean-field models have been noted to exhibit correlations between the infectivity β and parameters of

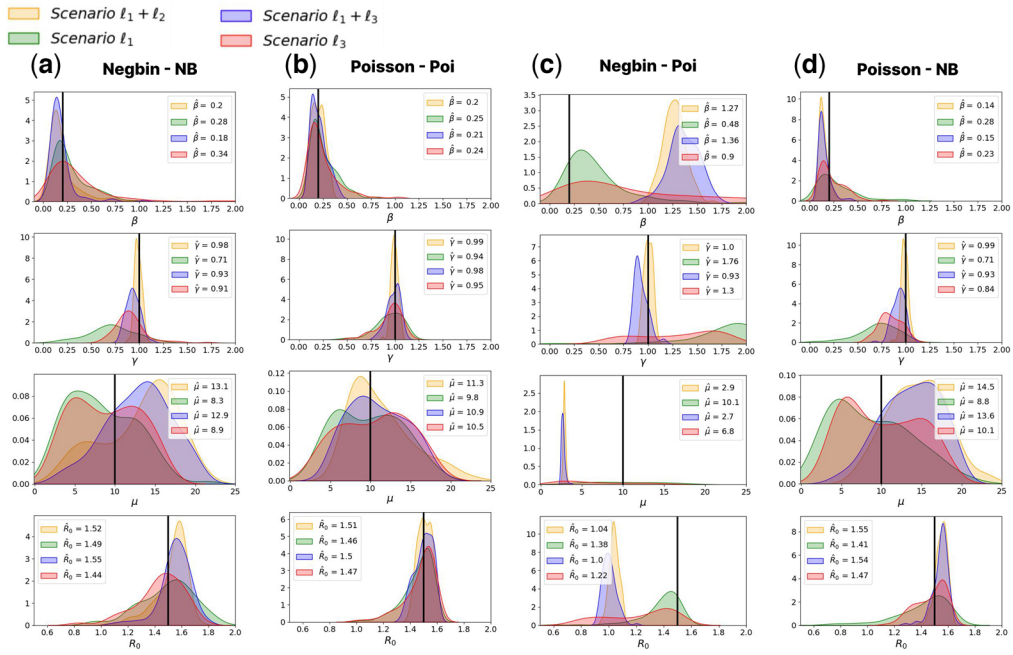

Figure 2. Distributions of the marginal medians for model parameters and the basic reproduction number inferred from the DSA framework. Each row represents the distribution of a different inferred parameter. The synthetic data were generated using either a Poisson or a Negative Binomial degree distribution, and inference was performed using both distributions for comparison. Each column corresponds to different inference cases: (a) and (b) show results where the data were fitted with the same degree distribution used for generation—(a) for Negative Binomial and (b) for Poisson. (c) and (d) show cases where the data were fitted using the opposite degree distribution—(c) for Negative Binomial data fitted with a Poisson model and (d) for Poisson data fitted with a Negative Binomial model. This analysis highlights the impact of assuming different degree distributions on parameter inference. The results are obtained by fitting 200 distinct realizations of Gillespie simulations with parameters $\beta = 0.2, \gamma = 1$, and an initial proportion of infected individuals set to 10^{-4}. Networks with a Negative Binomial (Negbin) degree distribution were generated using parameters $\mu = 10, r = 1$, while networks with a Poisson degree distribution were generated with $\mu = 10$.

the degree distribution, such as the average degree μ [62]. For this reason, we now investigate the projections of the posterior distribution. We do this for Poisson-Poi case, and we consider the scenario where the data is complete, i.e. $\ell_1 + \ell_2$. In Fig. 3a, we show the projection of the medians of the marginals in the (β, μ) space. We observe a strong inverse correlation between the infection rate β and the average degree μ.

Although we observe a considerable variance in both the average degree and the infection rate, it's important to highlight that most of the estimates of parameters are around the master parameter used in the simulation. Furthermore, no correlation is observed when exploring the (β, γ) parameter space. As expected, the correlations between parameters can be mitigated by fixing either the infection rate or the average degree. This approach leads to a posterior sample without long tails, resulting in more accurate estimates. Further discussions on the process of fixing parameters are provided in SM.

Up to this point, we have focused on parameter estimation with their validity based on comparison to their true values. However, another important consideration is to assess how the epidemic curve, such as new infecteds or prevalence, generated using the EBCM with the point estimates

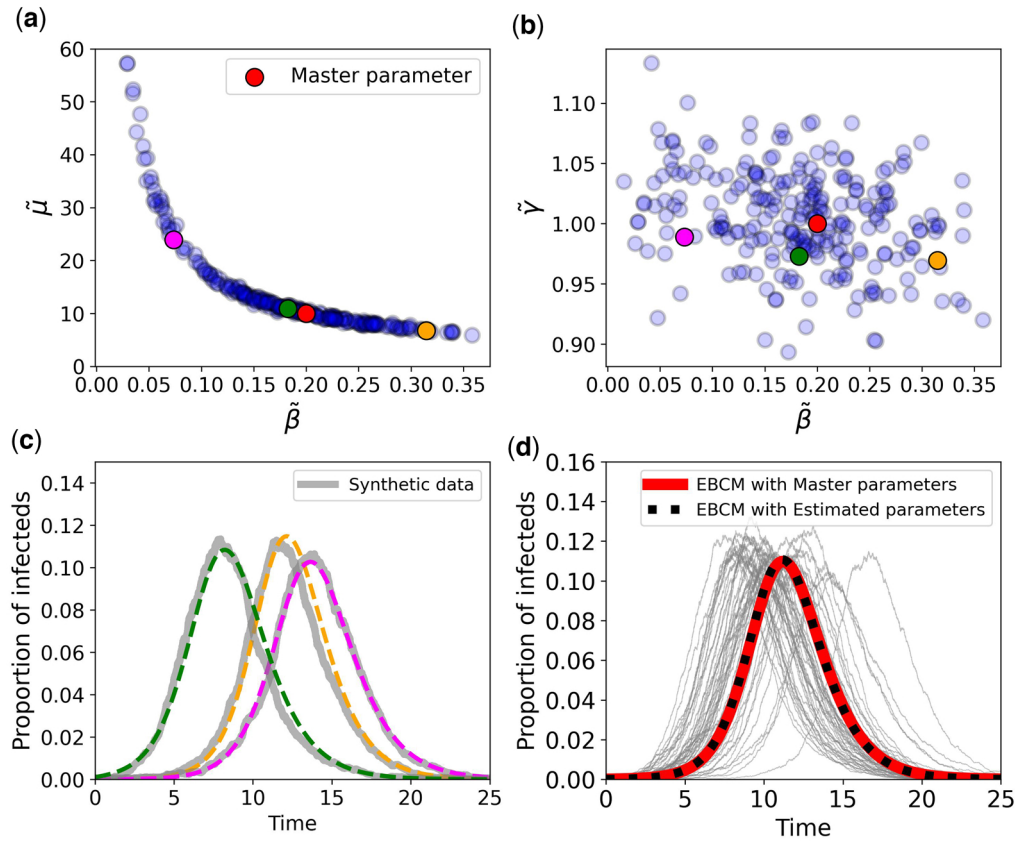

Figure 3. DSA parameters estimation for each synthetic dataset. (a) Projection of the 250 median estimates from the marginal posterior distributions of each dataset in the β–μ space. (b) Projection of the median estimates in the γ –β space. (c) Epidemic curves from three selected datasets (gray lines) alongside the corresponding curves obtained by solving the EBCM using the median parameter estimates [pink, green, and orange points in panels (a) and (b)]. Despite variations in parameter estimates, the model accurately reproduces the observed epidemic dynamics of the data. Panel (d) shows 50 of the 250 datasets (gray lines) with the solution of the EBCM obtained using the master parameters (dotted black line) and the solution from the final point estimates from the marginal posteriors (red line). Results are based on 250 synthetic epidemic datasets generated on a network with a Poisson degree distribution. The true parameters used were $\beta = 0.2$, $\gamma = 1$, $\mu = 10$, and $\rho = 1 \times 10^{-4}$. For inference, the cut-off time was set to $T = 25$.

compares to the original outbreak data. To address this, we consider three different stochastic realizations of the epidemic, see grey curves in Fig. 3c. We fit each of these separately with the corresponding point estimates reported as the pink, green, and orange points in Fig. 3a and b. The EBCM with these point estimates leads to excellent agreement with the original stochastic realizations, see Fig. 3c and d.

4. INFERENCE FROM REAL-WORLD DATA

In this section, we further validate our methodology by analyzing real epidemic data from two outbreaks. Based on the general procedure presented in Section 2.2, we apply it to two real epidemic datasets. The first one is the 2001 Foot-and-Mouth Disease (FMD) epidemic in the United Kingdom, which involved a highly contagious virus affecting farm animals. The second dataset

captures the first wave of COVID-19 in Seoul, South Korea, documenting the onset of symptoms and confirmation of infection for approximately 500 individuals over the first 82 days following the appearance of the initial confirmed case.

For both datasets, we fit the EBCM using a probability-generating function corresponding to a Negative Binomial degree distribution, chosen for its flexibility in modelling both homogeneous and heterogeneous contact patterns. We adopt non-informative prior distributions, assigning β, γ, μ, and r a GAMMA(a, b) prior, where a is randomly selected from the parameter space and b is fixed at 10^{-4}. The initial number of infected individuals follows a BETA$(1, 1)$ prior, representing a uniform distribution and reflecting complete uncertainty about the starting conditions. A key distinction from our analysis of synthetic data is that the total population size is unknown. As discussed in Section 2.2, one approach is to estimate an effective population size using $N_{\mathit{eff}} = K/(1 - S[T])$ [66]. To obtain a denser posterior distribution, we performed multiple runs of the RAM algorithm, varying the initial conditions of the chain for each run. Consequently, we find point estimates for each parameter and evaluate the accuracy of the EBCM predictions by comparing them to real-world epidemic data. To quantify the discrepancy, we use the mean squared error (MSE), defined as

$$MSE = \frac{1}{T}\sum_{d=0}^{T}(\hat{J}(d) - J(d))^2, \tag{15}$$

where d denotes the day index, T is the cut-off time, $\hat{J}(d)$ represents the incidence predicted by the EBCM, and $J(d)$ corresponds to the epidemic incidence observed in the real-data.

Here, we compare the predictions of the EBCM with those obtained using the standard mass-action (MA) SIR model [59]. The MA model is incorporated into the inference framework described in Section 2.2 and serves as a benchmark for epidemic curve predictions. The governing equations for the SIR MA model are given by:

$$\dot{s}_t = -\sigma s_t \imath_t, \quad \dot{\imath}_t = \sigma s_t \imath_t - \gamma \imath_t, \quad \dot{r}_t = \gamma \imath_t, \tag{16}$$

where γ is the recovery rate, and σ is the infection rate. It is important to note that σ differs from the infection rate β in the EBCM: while β represents the per-contact transmission rate, σ describes the infection rate per infectious individual.

Unlike the EBCM, which explicitly accounts for the network structure, the MA model assumes homogeneous random mixing, meaning it does not incorporate any connectivity patterns. As a result, while the EBCM enables inference of both the epidemic dynamics and the underlying contact structure, the MA model can only be used to predict the epidemic curve. Consequently, our comparison is limited to the epidemic trajectory rather than network-related properties. For parameter inference, we follow the same numerical procedure as outlined earlier.

4.1 Foot-and-mouth disease data

In this section, we analyze the Foot-and-Mouth Disease (FMD) dataset, which provides daily incidence data, $J(d)$, representing the number of newly infected individuals per day over a 200-day period. We focus on the first 82 days, corresponding to the initial wave of the outbreak. In the DSA framework, performing inference requires information about the temporal distribution of infections. To address this, we assume that infection times within each day are uniformly distributed and, based on this assumption, consider only the ℓ_1 likelihood.

We generate five independent Markov chains, each running for 100,000 iterations, following the methodology outlined in Section 2.2. To construct the posterior distribution, we discard the first 50,000 iterations as burn-in and thin the remaining samples by selecting every 50th iteration to mitigate autocorrelation. From the resulting posterior, we estimate the five parameters $(\beta, \gamma, \mu, r, \rho)$

Table 2. Estimated parameter values obtained using three inference methods: *marginal mean*, *marginal median*, and *joint mode*, for the the Foot-and-Mouth disease outbreak[a,b]

Parameter	Marginal mean	Marginal median	Joint mode
β	0.043	0.036	0.041
γ	0.30	0.28	0.32
μ	13.51	9.70	7.34
r	9.71	5.25	1.72
Variance	32.32	27.58	38.55
R_0	1.89	1.32	1.31
MSE incidence	3.44	1.03	0.76

[a]The table also reports the Mean Squared Error (MSE) between the predicted incidence and the observed data.
[b]Table showing estimated parameter values for the Foot-and-Mouth Disease outbreak using three inference methods: *marginal mean*, *marginal median*, and *joint mode*. Parameters include β, γ, μ, r, variance, R_0, and the Mean Squared Error (MSE) of predicted incidence.

using three different methods: *marginal mean*, *marginal median*, and *joint mode*, as shown in 1. The *joint mode* is computed using the mean-shift algorithm, which approximates the kernel density based on the posterior sample.

In Fig. 4a, we compare the EBCM solutions with real epidemic data using different parameter estimation methods. As shown, the parameters obtained via the *marginal mean* fail to capture the system's behavior. While the EBCM integrated with parameters estimated through the *marginal median* provides a better fit, it still does not fully capture the original data. However, the EBCM solution obtained using parameters estimated via the *joint mode* exhibits excellent agreement with the epidemic data. Additionally, we sample a subset of parameters, **S**, from the posterior distribution to construct a 95% credible interval for the daily new infections, as mentioned in the bottom right of 1. Notably, the prediction based on the *joint mode* lies at the centre of this interval. This is further illustrated in Fig. 4b and c, where we present projections of the full posterior distribution for the (β, μ) and (r, μ) parameter spaces, respectively. These figures highlight that the *marginal mean* and *marginal median* estimates lie outside the region of highest posterior density. On the other hand, point estimates obtained using the *joint mode* are closer to the densest regions of the posterior projections when compared to the estimates from the marginal distributions. The discrepancy is likely due to the presence of long tails in the posterior distributions, which obscure important correlations between parameters and lead to suboptimal point estimates. In Table 2, we provide the point estimates for each method, along with the corresponding values of the *MSE* calculated from the infeed incidence.

Additionally, in Fig. 4d, we show the degree distribution with values of the mean and variance that are in line with expectations since highly connected markets were disproportionately affected at the beginning of the outbreak leading to a marked reduction in network heterogeneity [71]. While the model successfully captures the epidemic curve, the inferred contact network structure cannot be directly validated against ground truth data, as no empirical contact network is available for comparison.

Furthermore, we compare the predictions made by the EBCM, with the ones made by using the MA model. In this case, the parameters exhibit a lower correlation, resulting in a posterior distribution without long tails. Consequently, the *marginal median*, *marginal mean*, and *joint mode* produce similar estimates. In the table at the top-left corner in Fig. 5, we show the estimates of β, γ, and R_0. Using these estimates, we can compare the MA and EBCM models in their ability to describe the original data. In Fig. 5, we plot the incidence along with the MSE calculated for each case.

The EBCM, which accounts for an explicit contact structure, provides a closer fit to the original data compared to the classical MA model. This is clearly evidenced by the lower MSE values achieved by the EBCM, suggesting that the FMD outbreak is better described by a model with an explicit contact structure rather than a homogeneously mixed model.

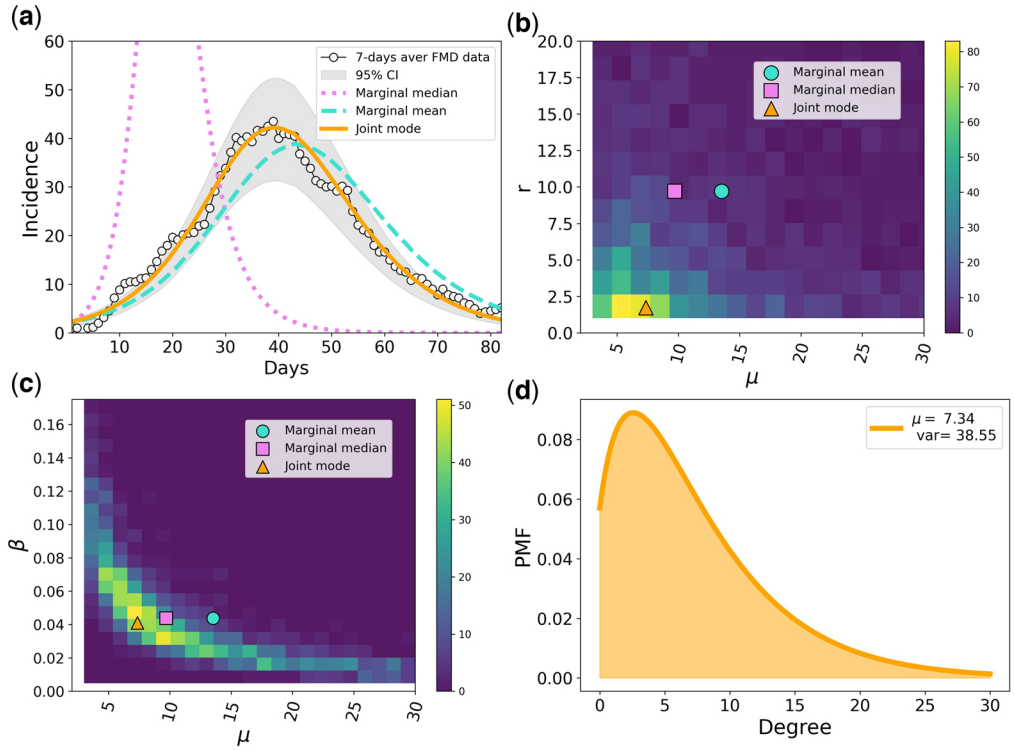

Figure 4. Summary of inference results for the 2001 Foot-and-Mouth Disease outbreak in the UK. (a) Evolution of the incidence based on parameter estimates from the three methods, alongside the 95% credible interval from the posterior distribution. The observed FMD incidence, smoothed using a 7-day moving average, is also shown for comparison. (b) and (c) Heatmaps of the posterior distribution projections, showing parameter density across the (β, μ) and (r, μ) spaces, respectively. Each projection includes point estimates obtained using the *marginal mean*, *marginal median*, and *joint mode*. (d) Inferred degree distribution obtained using the *joint mode* estimation method.

4.2 COVID-19 data from Seoul, South Korea

In this section, we extend our analysis to the first wave of the COVID-19 pandemic in Seoul, South Korea, covering an 84-day period from January 26 to April 18, 2020 [72]. This dataset provides information on the time of symptom onset for each individual, as well as the date of their positive test result. For our analysis, we assume that the time of infection coincides with the onset of symptoms. Given the strict isolation measures in place, we treat the time of a positive test result as the effective recovery time, as individuals were promptly isolated upon testing positive. Using this dataset, we can aggregate the information to track the evolution of incidence (the number of new infections per day, as in Section 4.1), prevalence (the total number of currently infected individuals at any given time), and daily recoveries (the number of individuals recovering each day).

Notably, the Seoul COVID-19 dataset also includes contact pattern data, recording the number of contacts each infected individual had between symptom onset and recovery. This allows us to use the contact data as ground truth for evaluating the network structure characteristics inferred by the EBCM. By comparing the inferred network properties with empirical contact data, we assess the capability of the model to recover meaningful structural information from epidemic observations. To perform our analysis, we consider the Negative Binomial degree distributions and follow the same procedure as in Section 4.1. Additionally, we consider the likelihood $\ell_1 + \ell_2$, which is able to accommodate the most complete type of data.

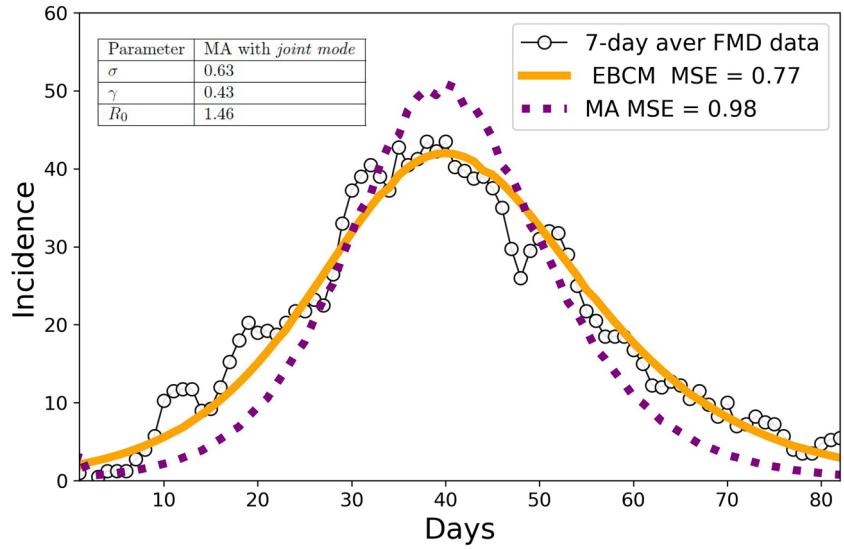

Figure 5. Comparison of predicted incidence of infections using the EBCM and mass-action model for Foot-and-Mouth Disease (FMD). The estimated incidence curves are presented for the DSA framework using the Edge-Based Compartmental Model (EBCM, shown as the thick continuous line) and the Mass-Action (MA) model (depicted square markers). Point estimates were derived using the joint mode for both models. The table in the top-left corner displays the results of the MA model, while the EBCM estimates can be found in Table 2. The thin black line with circle markers represents the 7-day moving average of observed FMD cases.

In Fig. 6, we show the predictions based on the EBCM for prevalence, incidence, and daily recoveries.

As with the FMD disease dataset (Section 4.1), the set of parameters estimated using the *joint mode* provides a temporal evolution of the system that aligns closely with the original data. In this case, the *marginal median* also performs well, yielding results that are closer to those obtained with the *joint mode*. In contrast, the *marginal mean* produces the least accurate results, failing to capture the behavior of the system. These findings are illustrated in Fig. 6a–c, which show the evolution of incidence, prevalence, and daily recoveries for each method. Additional evidence for these results is provided by the Mean Squared Error (MSE) values in Table 3.

In Fig. 6d, we compare the inferred degree distribution with the one obtained from the contact data. While the inferred distribution captures key characteristics of the empirical data, the point estimates derived from the *joint mode* yield a lower average degree and variance than those observed in the real network. Notably, the Negative Binomial distribution struggles to fully reproduce the long tail present in the empirical contact data, suggesting that higher-degree individuals may be under-represented in the inferred network structure.

Furthermore, we can use the contact data to fix the average degree. In this case, we fix $\mu = 7.61$, which is the average number of contacts entered in the survey. We can apply the inference procedure to find β, γ, and the parameter r of the Negative Binomial distribution. Table 4 presents the point estimates obtained from the three methods: *marginal mean*, *marginal median*, and *joint mode*. In this scenario, the three methods produce very similar estimates. Furthermore, in Fig. 7a, we observe the incidence predicted by the three methods, which are nearly identical. This indicates that fixing the average degree results in a much better-behaved posterior distribution. It is important to note that the contact data collected should be interpreted with caution. This data is highly dependent on an individual's perceptions and, as such, cannot be treated as a definitive ground truth for contact patterns.

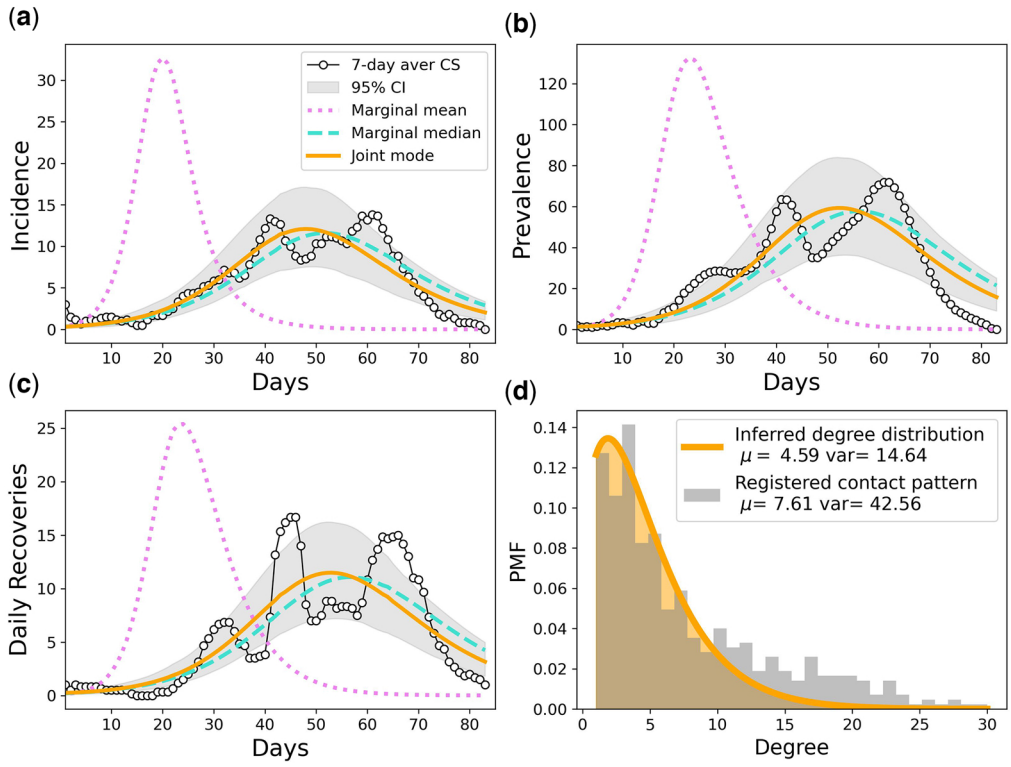

Figure 6. Inference results for the first wave of COVID-19 in Seoul. Panels (a), (b), and (c) depict the time evolution of prevalence, incidence, and daily recoveries, respectively, estimated using three inference methods: *marginal mean* (dotted line), *marginal median* (dashed line), and *joint mode* (solid line). These estimates are compared to the 7-day moving average of observed data from the first wave of COVID-19 in Seoul (CS), represented by the thin line with circle markers. Panel (d) presents the inferred probability mass function derived from epidemic data, alongside the recorded contact data.

Table 3. Estimated parameter values obtained using three inference methods: *marginal mean*, *marginal median*, and *joint mode*, for the 7-day moving average of the first wave of COVID-19 in Seoul (CS)[a,b]

Parameter	Marginal mean	Marginal median	Joint mode
β	0.042	0.035	0.052
γ	0.192	0.192	0.193
μ	12.80	8.96	4.59
Variance	16.82	12.30	14.64
R_0	2.35	1.43	1.44
MSE incidence	1.14	0.48	0.41
MSE prevalence	4.65	2.65	2.37
MSE daily recovered	1.07	0.65	0.6

[a]The table also reports the Mean Squared Error (MSE) between the predicted epidemic trajectory and the observed data

[b]Table presenting estimated parameter values for the first wave of COVID-19 in Seoul using three inference methods: *marginal mean*, *marginal median*, and *joint mode*. Parameters include β, γ, μ, variance, and R_0. The table also reports the Mean Squared Error (MSE) for predicted incidence, prevalence, and daily recoveries compared to observed data.

Table 4. Estimated parameter values obtained using three inference methods: *marginal mean*, *marginal median*, and *joint mode*, for the 7-day moving average of the first wave of COVID-19 in Seoul (CS) for a fixed average degree $\mu = 7.61$, obtained from the COVID-19 Survey data[a,b]

Parameter	Marginal mean	Marginal median	Joint mode
β	0.044	0.044	0.043
γ	0.196	0.196	0.193
Variance	8.77	9.33	12.10
R_0	1.42	1.44	1.47
MSE incidence	0.45	0.42	0.41
MSE prevalence	2.44	2.30	2.25
MSE new recovered	0.63	0.61	0.60

[a]The table also reports the Mean Squared Error (MSE) between the predicted epidemic trajectory and the observed data.
[b]Table showing estimated values of β, γ, variance, and R_0 using three inference methods—*marginal mean*, *marginal median*, and *joint mode*—for the first wave of COVID-19 in Seoul with fixed average degree $\mu = 7.61$. Also includes Mean Squared Error (MSE) values for predicted incidence, prevalence, and new recoveries compared to observed data.

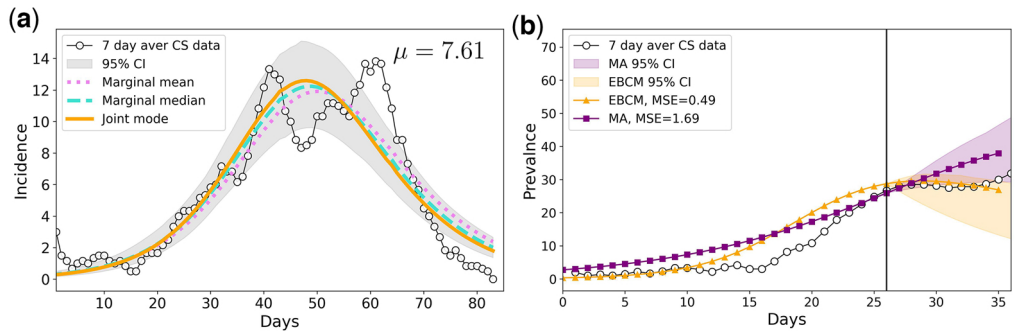

Figure 7. Additional analysis of the first wave of COVID-19 in Seoul: Panel (a) shows the evolution of the incidence obtained from the point estimates using *marginal mean*, *marginal median*, and *joint mode* for the case where the average degree is known and fixed at $\mu = 7.61$. Panel (b) shows the results obtained using partial data. Specifically, the first 26 days of data were used to forecast the following 10 days. In the figure, we show the prediction obtained using the EBCM (line with triangle markers), MA (line with square markers), and the 7 days moving average of the COVID-19 in Seoul (CS) (line with circle markers).

Finally, to further demonstrate the potential of this framework, we performed parameter estimation, including the mean degree—using only partial data, specifically up to a cut-off time T_c before the 84th day. In this case, we used the first 26 days of data to infer parameters and then forecasted the number of new infections for the following 10 days. As a result, the likelihood is now $\ell_1 + \ell_2 + \ell_4$, where ℓ_4 accounts for the infection times of individual who did not recover before T_c. The 95% credible interval was obtained by sampling from the full posterior distribution, while point estimates were determined using the *joint mode*. For comparison, we performed the same analysis using the mass-action (MA) model. The results are shown in Fig. 7b. The forecasted incidence was compared to the observed data using the mean squared error (MSE). The results indicate that the EBCM provides the most accurate forecast.

5. DISCUSSION

Understanding the contact patterns of individuals during an epidemic remains a fundamental challenge in infectious disease modelling. Inferring the underlying network structure of an epidemic process is particularly difficult, even when the goal is not to reconstruct the entire network but rather to estimate key characteristics, such as the average degree or variance. A major obstacle in

this inference process is the practical identifiability between connectivity and infectivity, a challenge previously highlighted in the literature [54, 62].

In this study, we introduced a framework for inferring network properties from epidemic data by integrating the Dynamic Survival Analysis (DSA) framework with the EBCM. The EBCM provides a compact yet effective representation of the epidemic process, where the degree distribution is incorporated as a model parameter. By combining this with the flexibility of DSA, our approach enables the inference of crucial properties of the contact network that drive epidemic spreading.

We validated this framework using synthetic epidemic data generated via Gillespie simulations, considering networks with Poisson and Negative Binomial degree distributions. By applying the DSA-EBCM approach, we sampled from the joint posterior distribution of both disease and network parameters. Despite the inherent correlation between the infection rate, the recovery rate, and network properties such as the average degree, the posterior distributions consistently concentrated around the true parameter values. This allowed for an accurate reconstruction of both the epidemic dynamics and key network characteristics, despite relying solely on epidemic time-series data in which the network structure is only implicitly present.

Beyond synthetic data, we tested our methodology on two real-world outbreaks: the 2001 Foot-and-Mouth Disease (FMD) epidemic in the UK and the first wave of COVID-19 in Seoul, South Korea. In both cases, the framework successfully produced robust posterior distributions despite correlations between parameters. For the FMD dataset, we observed a multidimensional posterior distribution with long tails and a strong inverse correlation between β and μ. Nevertheless, as in the synthetic cases, a high-density region in parameter space provided the best description of the original data. Notably, we found that the *joint mode*, which represents the point of highest density in the sample of the posterior distribution, although more computationally demanding and less exact, yielded better point estimates compared to the *marginal median* or *marginal mean*, as it minimized the mean squared error.

For the Seoul COVID-19 dataset, we observed similar results, with the *joint mode* again emerging as the most accurate estimator of the epidemic process. Additionally, our inferred average degree closely matched independent contact data collected during the outbreak, reinforcing the validity of the approach. The framework also demonstrated its predictive capabilities by generating a short-term forecast with a 95% credible interval for the epidemic's progression over a 10-day period.

This study demonstrates that meaningful insights about underlying contact structures can be extracted solely from epidemic data without requiring explicit network observations. Future work could explore the extent to which this method can distinguish between homogeneous and heterogeneous network structures based only on outbreak dynamics. Furthermore, the approach could be extended to other spreading processes, such as information diffusion, where higher-order interactions may play a significant role. Comparing inferred network properties across different geographic regions—such as cities, counties, or states—could also provide insights into the diverse mechanisms that shape disease transmission.

Overall, our findings highlight the potential of integrating the EBCM and other network-based mean-field models with DSA to infer hidden contact structures from limited epidemic data. This approach provides a powerful tool for reconstructing essential network characteristics, improving epidemic forecasting, and enhancing our understanding of infectious disease spread in real-world settings.

ACKNOWLEDGEMENTS

A.G. acknowledges the PhD studentship support from Northeastern University London. A.G. and I.Z.K. acknowledge useful discussion with Mauricio Santillana. B.C. and G.H.P. acknowledge the Basic Science Research Program through the NRF funded by the Ministry of Education (RS-202300245056). B.C. acknowledges a grant of the project The Government-wide R&D to Advance Infectious Disease Prevention and Control (HG23C1629).

SUPPLEMENTARY DATA

Supplementary data is available at *COMNET Journal* online.

FUNDING

A.G. acknowledges the PhD studentship support from Northeastern University London.

DATA AVAILABILITY

Code and synthetic datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Kiss IZ, Miller JC, Simon PL *et al*. *Mathematics of Epidemics on Networks*. Vol. **598**, Cham: Springer, 2017, 31.
2. Pastor-Satorras R, Castellano C, Van Mieghem P *et al*. Epidemic processes in complex networks. *Rev Mod Phys* 2015;**87**:925–79.
3. Newman MEJ. The structure and function of complex networks. *SIAM review* 2003;**45**:167–256.
4. Keeling MJ, Eames KTD. Networks and epidemic models. *Journal of the royal society interface* 2005;**2**:295–307.
5. Danon L, Ford AP, House, T *et al*. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases* 2011;**2011**:284909.
6. Moreno Y, Pastor-Satorras R, Vespignani A. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 2002;**26**:521–9.
7. Chakrabarti D, Wang Y, Wang C *et al*. Epidemic thresholds in real networks. *Transactions on Information and System Security (TISSEC)* 2008;**10**:1:1–26.
8. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett* 2001;**86**:3200–3.
9. May RM, Lloyd AL. Infection dynamics on scale-free networks. *Phys Rev E* 2001;**64**:066112.
10. Pastor-Satorras R, Vespignani A *et al*. *Epidemics and immunization in scale-free networks. Handbook of Graphs and Networks*. Berlin: Wiley-VCH, 2003.
11. Stegehuis C, Hofstad R, Leeuwaarden J. Epidemic spreading on complex networks with community structures. *Sci Rep* 2016;**6**:29748.07
12. Li C, Jiang G. P, Song Y *et al*. Modeling and analysis of epidemic spreading on community networks with heterogeneity. *J Parallel Distrib Comput* 2018;**119**:136–45.
13. Nadini M, Sun K, Ubaldi E *et al*. Epidemic spreading in modular time-varying networks. *Sci Rep* 2018;**8**:2352.
14. Griffin RH, Nunn CL. Community structure and the spread of infectious disease in primate social networks. *Evol Ecol* 2012;**26**:779–800.
15. Eguíluz VM, Klemm K. Epidemic threshold in structured scale-free networks. *Phys Rev Lett* 2002;**89**:108701.Aug
16. Smieszek T, Fiebig L, Scholz RW. Models of epidemics: when contact repetition and clustering should be included. *Theor Biol Med Model* 2009;**6**:11–5.
17. Miller JC. Percolation and epidemics in random clustered networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2009;**80**:020901.
18. Coupechoux E, Lelarge M. How clustering affects epidemics in random networks. *Adv Appl Probab* 2014;**46**:985–1008.
19. Boguá M, Pastor-Satorras R, Vespignani A. Epidemic spreading in complex networks with degree correlations. *Stat Mech Complex Netw* 2003;127–47.
20. Wang Y, Ma J, Cao J *et al*. Edge-based epidemic spreading in degree-correlated complex networks. *J Theor Biol* 2018;**454**:164–81.
21. Boguñá M, Pastor-Satorras R, Vespignani A. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys Rev Lett* 2003;**90**:028701.
22. Chen X-H, Cai S-M, Wang W *et al*. Predicting epidemic threshold of correlated networks: a comparison of methods. *Phys A Stat Mech Appl* 2018;**505**:500–11.
23. Latora V, Nicosia V, Russo G. *Complex Networks: Principles, Methods and Applications*. Cambridge, United Kingdom: Cambridge University Press, 2017.
24. Malizia F, Gallo L, Frasca M *et al*. Individual-and pair-based models of epidemic spreading: master equations and analysis of their forecasting capabilities. *Phys Rev Res* 2022;**4**:023145.
25. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond Ser A Contain Papers Math Phys Character* 1927;**115**:700–21.
26. Volz E, Meyers LA. Epidemic thresholds in dynamic contact networks. *J R Soc Interface* 2009;**6**:233–41.

27. Ball F, Neal P. Network epidemic models with two levels of mixing. *Math Biosci* 2008;**212**:69–87.
28. Zhang Z, Wang H, Wang C *et al*. Modeling epidemics spreading on social contact networks. *IEEE Trans Emerg Top Comput* 2015;**3**:410–9.
29. Mollison D. *Epidemic Models: Their Structure and Relation to Data. Number 5*. Cambridge, United Kingdom: Cambridge University Press, 1995.
30. Gibson GJ, Streftaris G, Thong D. Comparison and assessment of epidemic models. *Statist Sci* 2018;**33**:19–33.
31. Daunizeau J, Moran R, Mattout J, Friston K. On the reliability of model-based predictions in the context of the current covid epidemic event: impact of outbreak peak phase and data paucity. *MedRxiv*, 2020, preprint: not peer reviewed.
32. Keeling MJ, Eames KT. Networks and epidemic models. *J R Soc Interface* 2005;**2**:295–307.
33. Miller JC, Slim AC, Volz EM. Edge-based compartmental modelling for infectious disease spread. *J R Soc Interface* 2012;**9**:890–906.
34. Miller JC, Volz EM. Model hierarchies in edge-based compartmental modeling for infectious disease spread. *J Math Biol* 2013;**67**:869–99.
35. Volz EM, Miller JC, Galvani A *et al*. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput Biol* 2011;**7**:e1002042.
36. Eng TR, Butler WT *et al*. The hidden epidemic: confronting sexually transmitted diseases. Washington, DC: National Academy Press, 1997.
37. Eames KT, Keeling MJ. Contact tracing and disease control. *Proc R Soc Lond B* 2003;**270**:2565–71.
38. Rothenberg RB, McElroy PD, Wilce MA *et al*. Contact tracing: comparing the approaches for sexually transmitted diseases and tuberculosis. *Int J Tuberc Lung Dis* 2003;**7**:S342–S348.
39. Garnett GP, Anderson RM. Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. *Sex Transm Dis* 1993;**20**:181–91. pages
40. Buckee C. Improving epidemic surveillance and response: big data is dead, long live big data. *Lancet Digit Health* 2020;**2**:e218–e220.
41. Zubaydi F, Sagahyroon A, Aloul F *et al*. Using mobiles to monitor respiratory diseases. *Informatics* 2020;**7**:56–MDPI.
42. Kwok KO, Tang A, Wei VW *et al*. Epidemic models of contact tracing: systematic review of transmission studies of severe acute respiratory syndrome and middle east respiratory syndrome. *Comput Struct Biotechnol J* 2019;**17**:186–94.
43. Vanli OA, Tsekeni DE. Inference of human contact networks based on epidemic data. *IISE Trans Healthcare Syst Eng* 2025;**15**:15–31.
44. Viboud C, Sun K, Gaffey R *et al*., RAPIDD Ebola Forecasting Challenge Group. The Rapidd Ebola forecasting challenge: synthesis and lessons learnt. *Epidemics* 2018;**22**:13–21.
45. Cori A, Kucharski A. Inference of epidemic dynamics in the covid-19 era and beyond. *Epidemics* 2024;**48**:100784.
46. Vasiliauskaite V, Antulov-Fantulin N, Helbing D. On some fundamental challenges in monitoring epidemics. *Philos Trans R Soc A* 2022;**380**:20210117.
47. Swallow B, Birrell P, Blake J *et al*. Challenges in estimation, uncertainty quantification and elicitation for pandemic modelling. *Epidemics* 2022;**38**:100547.
48. Shandilya SG, Timme M. Inferring network topology from complex dynamics. *New J Phys* 2011;**13**:013004.
49. Nitzan M, Casadiego J, Timme M. Revealing physical interaction networks from statistics of collective dynamics. *Sci Adv* 2017;**3**:e1600396.
50. Netrapalli P, Sanghavi S. Learning the graph of epidemic cascades. *Sigmetrics Perform Eval Rev* 2012;**40**:211–22.
51. Malizia F, Corso A, Gambuzza LV *et al*. Reconstructing higher-order interactions in coupled dynamical systems. *Nat Commun* 2024;**15**:5184.
52. Rodriguez MG, Leskovec J, Balduzzi D *et al*. Uncovering the structure and temporal dynamics of information propagation. *Netw Sci* 2014;**2**:26–65.
53. Groendyke C, Welch D, Hunter DR. Bayesian inference for contact networks given epidemic data. *Scand J Stat* 2011;**38**:600–16. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9469.2010.00721.x
54. Britton T, O'Neill PD. Bayesian inference for stochastic epidemics in populations with random social structure. *Scand J Stat* 2002;**29**:375–90. https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9469.00296
55. Lewis F, Hughes GJ, Rambaut A *et al*. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med* 2008;**5**:e50.
56. Demiris N, O'Neill PD. bayesian inference for stochastic multitype epidemics in structured populations via random graphs. *J R Stat Soc Ser B Stat Methodol* 2005;**67**:731–45.
57. Prasse B, Van Mieghem P. Exact network reconstruction from complete SIS nodal state infection information seems infeasible. *IEEE Trans Netw Sci Eng* 2019;**6**:748–59.
58. Gallo L, Frasca M, Latora V *et al*. Lack of practical identifiability may hamper reliable predictions in covid-19 epidemic models. *Sci Adv* 2022;**8**:eabg5234.

59. KhudaBukhsh WR, Choi B, Kenah E *et al*. Survival dynamical systems: individual-level survival analysis from population-level epidemic models. *Interface Focus* 2020;**10**:20190048.
60. Di Lauro F, KhudaBukhsh WR, Kiss IZ *et al*. Dynamic survival analysis for non-Markovian epidemic models. *Journal of The Royal Society Interface* 2022;**19**:20220124.
61. Rempała GA, KhudaBukhsh WR. Dynamical survival analysis for epidemic modeling. In: Sriraman B (ed.), *Handbook of Visual, Experimental and Computational Mathematics: Bridges through Data*. Cham, Switzerland: Springer International Publishing, 2023;1–17.
62. Kiss IZ, Berthouze L, KhudaBukhsh WR. Towards inferring network properties from epidemic data. *Bull Math Biol* 2023;**86**:6.
63. Kiss IZ, Simon PL. On parameter identifiability in network-based epidemic models. *Bull Math Biol* 2023;**85**:18.
64. Molloy M, Reed B. A critical point for random graphs with a given degree sequence. *Random Struct Algorithms* 1995;**6**:161–80.
65. Newman ME, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 2001;**64**:026118.
66. Di Lauro F, KhudaBukhsh WR, Kiss IZ *et al*. Dynamic survival analysis for non-Markovian epidemic models. *J R Soc Interface* 2022;**19**:20220124.
67. Vossler H, Akilimali P, Pan Y *et al*. Analysis of individual-level data from 2018–2020 Ebola outbreak in democratic republic of the Congo. *Sci Rep* 2022;**12**:5534.
68. KhudaBukhsh WR, Bastian CD, Wascher M *et al*. Projecting COVID-19 cases and hospital burden in Ohio. *J Theor Biol* 2023;**561**:111404.
69. Vihola M. Robust adaptive metropolis algorithm with coerced acceptance rate. *Stat Comput* 2012;**22**:997–1008.
70. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 1976;**22**:403–34.
71. Davies G. The foot and mouth disease (fmd) epidemic in the united kingdom 2001. *Comp Immunol Microbiol Infect Dis* 2002;**25**:331–43.
72. Ha JH, Lee JY, Choi SY, Park SK. Covid-19 waves and their characteristics in the seoul metropolitan area (jan 20, 2020–aug 31, 2022). *Public Health Weekly Report* 2023;**16**:111–36.