Supplementary Material for "On the simultaneous inference of susceptibility distributions and intervention effects from epidemic curves"

Ibrahim Mohammed, Chris Robertson, M. Gabriela M. Gomes

Here we provide extra text and figures referenced in the main text.

S1 Setting initial conditions for simulations

To set the initial conditions for simulations in Section 3 ("Simulation"), we assume a number of infectious individuals at time zero (i.e., $\mathsf{I}(0)$) and, from that, approximate how many are expected to be in the other model compartments. Making the approximations that no-one has recovered from infection by that time ($\mathsf{R}(0) = 0$), and that depletion of susceptibles and behavioural change have not impacted early infection dynamics, system (4) of the main text is approximated by the linear system

$$\begin{aligned} \frac{\mathrm{d}\mathsf{E}}{\mathrm{d}t} &= \beta\ (\rho\ \mathsf{E} + \mathsf{I})\ -\delta\ \mathsf{E}, \\ \frac{\mathrm{d}\mathsf{I}}{\mathrm{d}t} &= \delta\ \mathsf{E} - \gamma\ \mathsf{I}, \end{aligned} \tag{1}$$

and $\mathsf{S}(0) = \mathsf{N} - \mathsf{E}(0) - \mathsf{I}(0)$. Then we use system (1) to determine $\mathsf{E}(0)$. In matrix form, the system becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \mathsf{E} \\ \mathsf{I} \end{pmatrix} = \underbrace{\begin{pmatrix} \rho\beta - \delta & \beta \\ \delta & -\gamma \end{pmatrix}}_{A}\begin{pmatrix} \mathsf{E} \\ \mathsf{I} \end{pmatrix}. \tag{2}$$

The eigenvalues of A are

$$\lambda_{\pm} = \frac{\rho\beta - \delta - \gamma \pm \sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta}}{2}. \tag{3}$$

Substituting the values of the parameters in Table 1 of the main text (namely, $\delta = 1/5.5$ per day, $\gamma = 1/4$ per day, $\rho = 0.5$, and $\mathcal{R}_0 = 3$), and using the relation $\mathcal{R}_0 = \beta\left(\frac{\rho}{\delta} + \frac{1}{\gamma}\right)$ (main text, formula (2)), we see that $\lambda_+ \approx 0.2 > 0$ and $\lambda_- \approx -0.09 < 0$. Hence, solutions of system (1) tend asymptotically to the eigenvector corresponding to eigenvalue λ_+.

The eigenvectors are $(\mathsf{E}_{\pm} \quad \mathsf{I}_{\pm})^T$ such that

$$\mathsf{E}_{\pm} = \frac{\gamma + \frac{1}{2}(\rho\beta - \delta - \gamma) \pm \frac{1}{2}\sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta}}{\delta}\mathsf{I}_{\pm}, \tag{4}$$

and hence solutions of system (1) take the form

$$\begin{aligned} \begin{pmatrix} \mathsf{E}(t) \\ \mathsf{I}(t) \end{pmatrix} = c_+ &\begin{pmatrix} \gamma + \frac{1}{2}(\rho\beta - \delta - \gamma) + \frac{1}{2}\sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta} \\ \delta \end{pmatrix} e^{\lambda_+ t} \\ + c_- &\begin{pmatrix} \gamma + \frac{1}{2}(\rho\beta - \delta - \gamma) - \frac{1}{2}\sqrt{(\rho\beta - \delta + \gamma)^2 + 4\delta\beta} \\ \delta \end{pmatrix} e^{\lambda_- t} \end{aligned} \tag{5}$$

Since $\lambda_- < 0$, all solutions tend to the first term on the right hand side of (5), which satisfies a ratio $\mathsf{E}(t)/\mathsf{I}(t)$ of approximately 2.55. Moreover, the convergence is sufficiently fast for this to be an excellent approximation for all scenarios simulated in this paper. Specifically, all our simulations are conducted for $\mathsf{I}(0) \geq 20$ and Figure S1 shows that the ratio E/I is already very close to 2.5 well before I reaches 20.

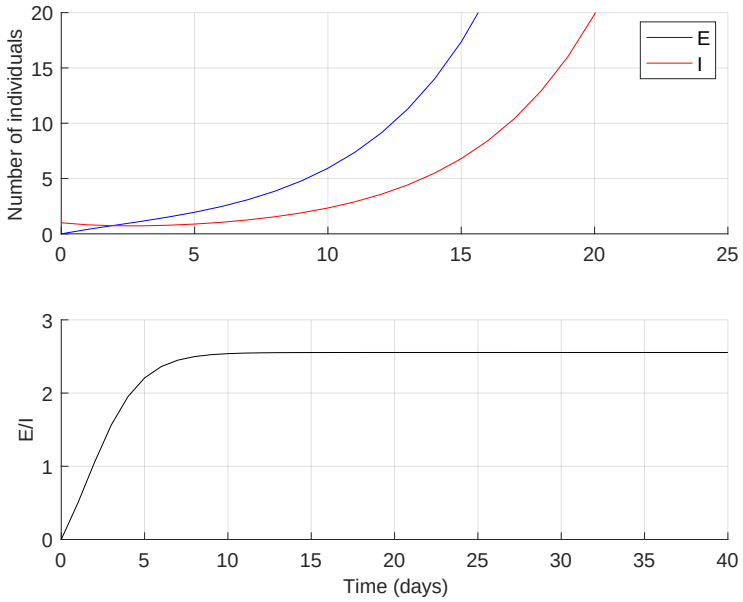

Figure S1: Solution of linear system (1) starting from $(\mathsf{E}, \mathsf{I}) = (0, 1)$. Trajectories of E and I over time (top); ratio E/I (bottom).

S2 Model fittings

Here we show a sample of the fittings associated with results reported in Section 5 ("Baseline analysis") of the main text.

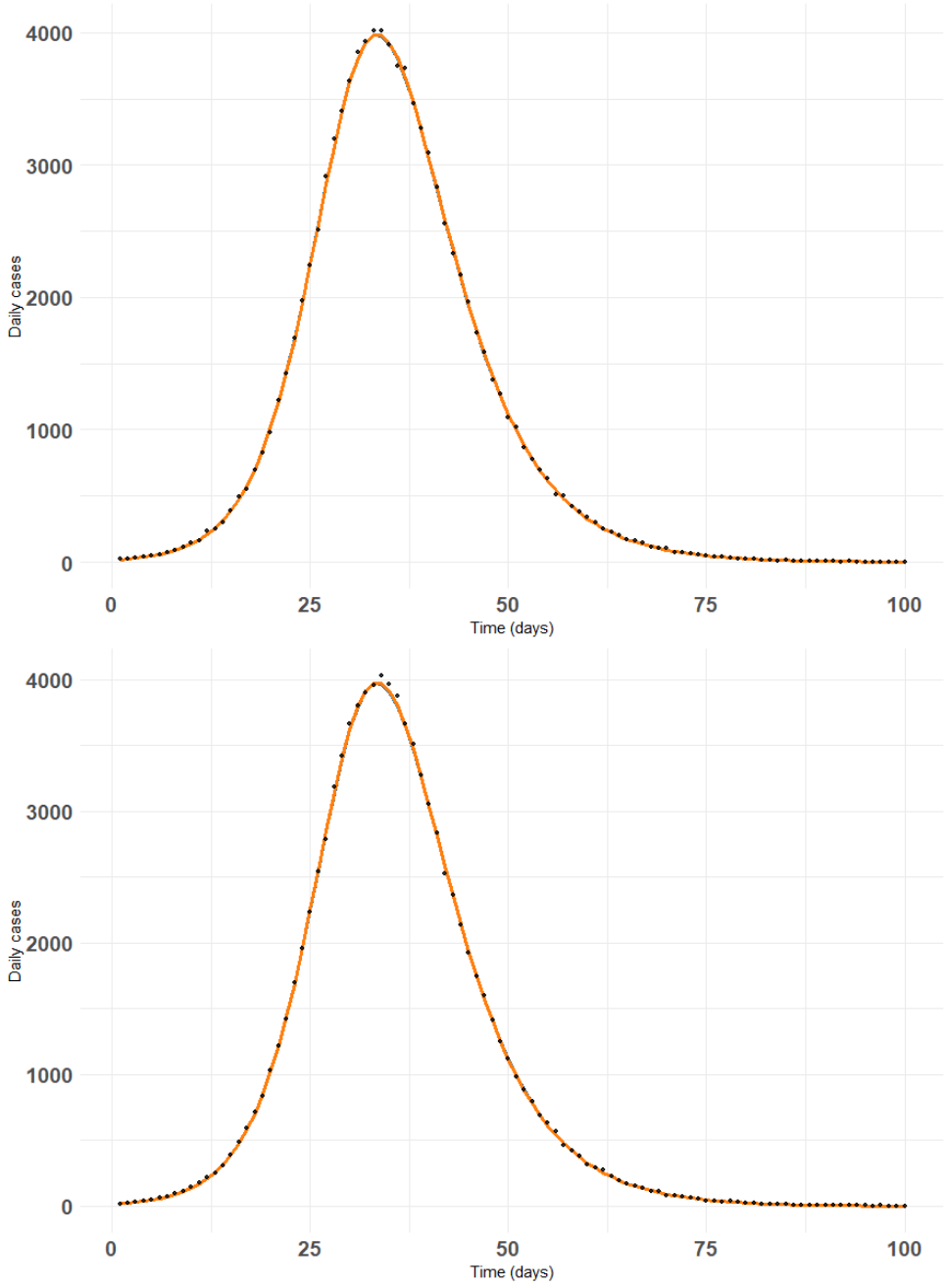

Figure S2: Case I(a): Homogeneous data without NPIs. Black dots represent simulated daily cases. Heterogeneous model fits (blue lines) and homogeneous model fits (orange lines) are practically coincident. Fittings conducted without NPIs (scenario (i), top); fittings allowing NPI effects (scenario (ii), bottom).

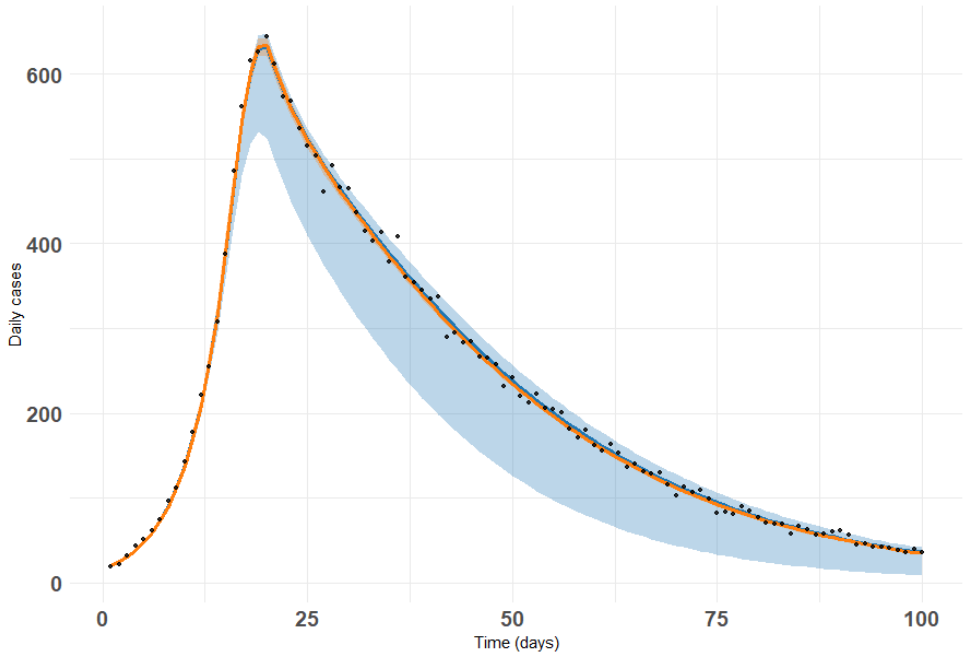

Figure S3: Case I(b): Homogeneous data with NPIs. Black dots represent simulated daily cases. Heterogeneous model fit (blue line) and homogeneous model fit (orange line) are practically coincident. Fittings allowing NPI effects.

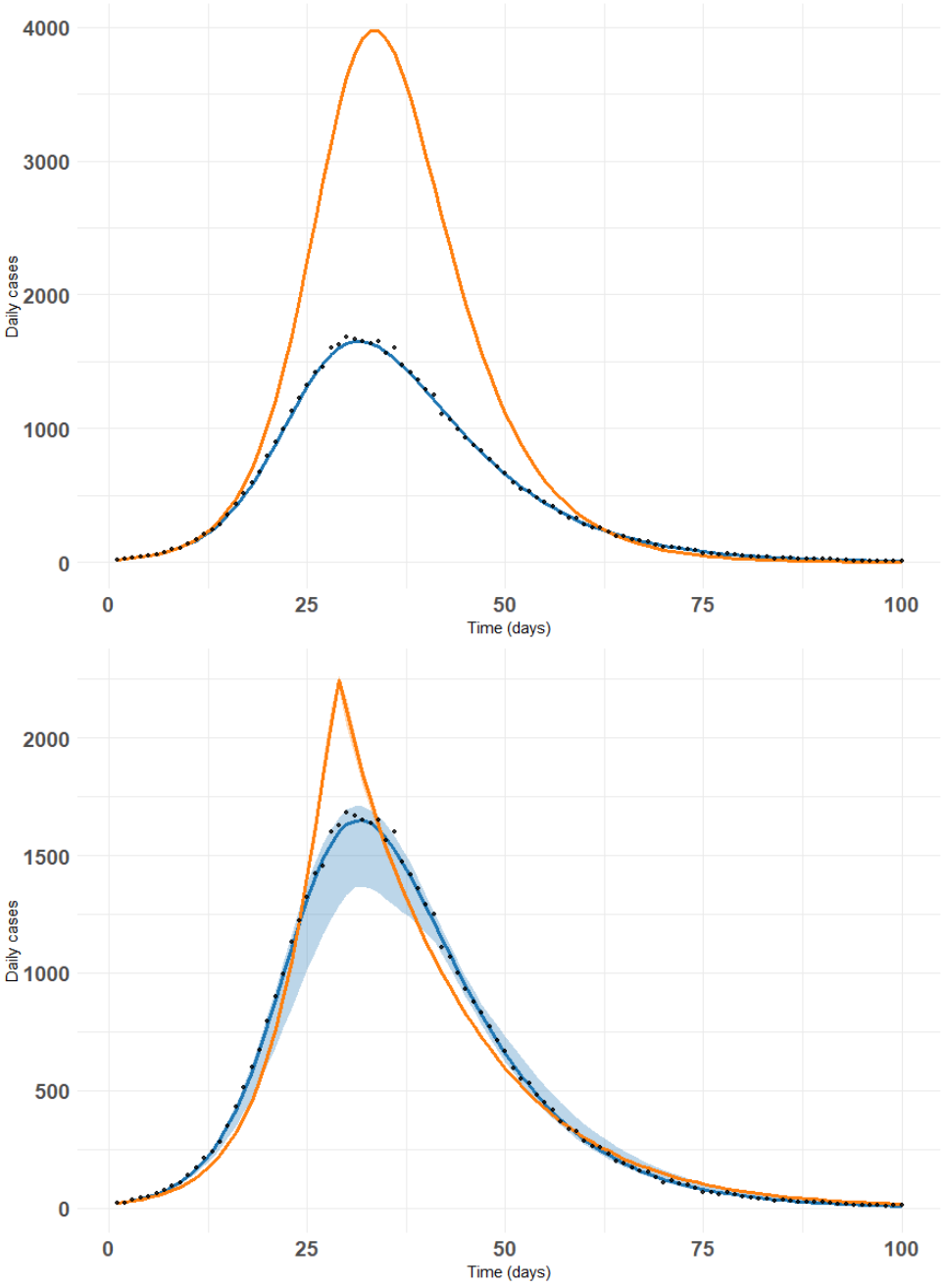

Figure S4: Case II(a): Heterogeneous data without NPIs. Black dots represent simulated daily cases. Heterogeneous model fits (blue lines); homogeneous model fits (orange lines). Fittings conducted without NPIs (scenario (i), top); fittings allowing NPI effects (scenario (ii), bottom).

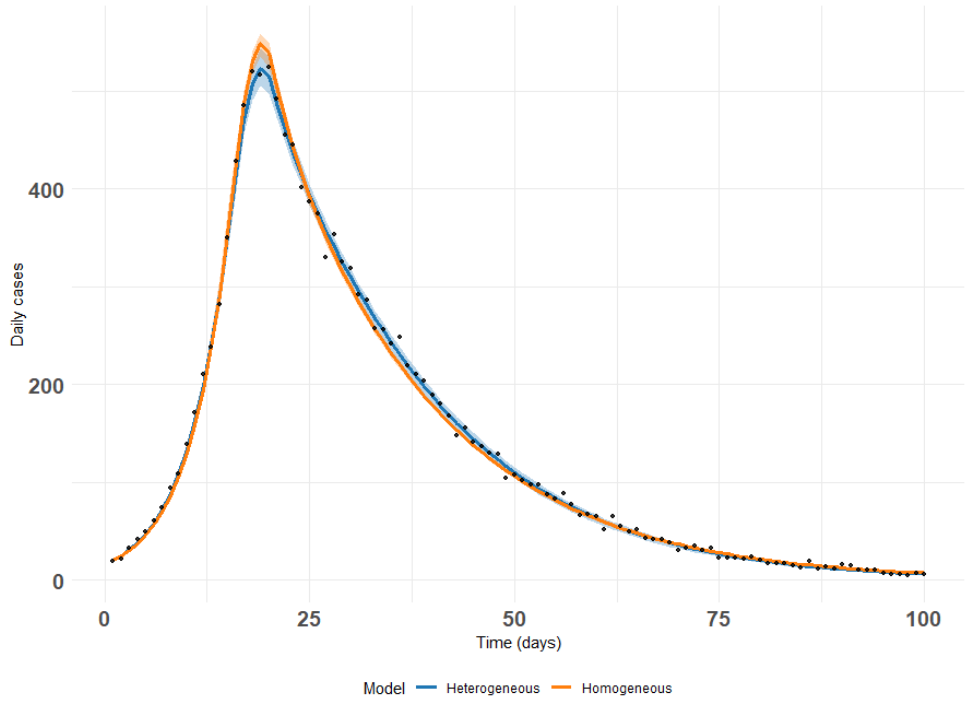

Figure S5: Case II(b): Heterogeneous data with NPIs. Black dots represent simulated daily cases. Heterogeneous model fit (blue line); homogeneous model fit (orange line). Fittings allowing NPI effects.

S3 Sensitivity to initial conditions and intervention strength

Here we investigate the sensitivity of parameter identifiability to initial conditions and intervention strengths. We examine how the size of the initial infected population ($\mathsf{I}(0)$) and the parameter controlling the maximum strength of the intervention (c_1) affect the correlation structure between model parameters.

We conducted a systematic analysis using simulated epidemic data with varying initial conditions and intervention strength. For each initial infected population $\mathsf{I}(0) \in \{20, 40, 80, 160, 320, 400\}$, we generated 200 synthetic epidemic datasets using the heterogeneous SEIR model with fixed parameters ($\mathcal{R}_0 = 3$, $\nu = 1.414$, $t_0 = 15$ days). We examined three intervention strengths $c_1 \in \{0.2, 0.3, 0.4\}$. We estimated parameters for each scenario.

Figures S6, S7, and S8 show the distribution of parameter correlations across initial conditions for both single and two epidemic approaches.

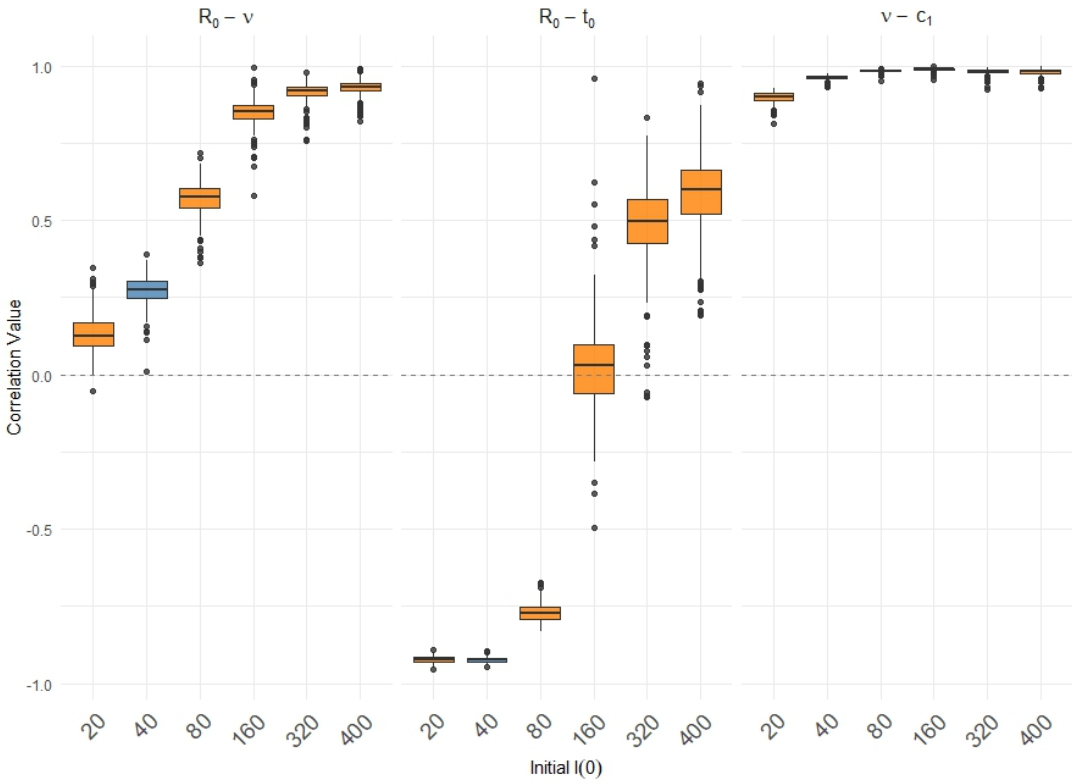

Figure S6: Correlation distributions in terms of initial condition ($\mathsf{I}(0) \in \{20, 40, 80, 160, 320, 400\}$) for intervention parameter $c_1 = 0.4$. Cases with initial conditions used in the baseline analysis ($\mathsf{I}(0) = 40$) are marked in blue.

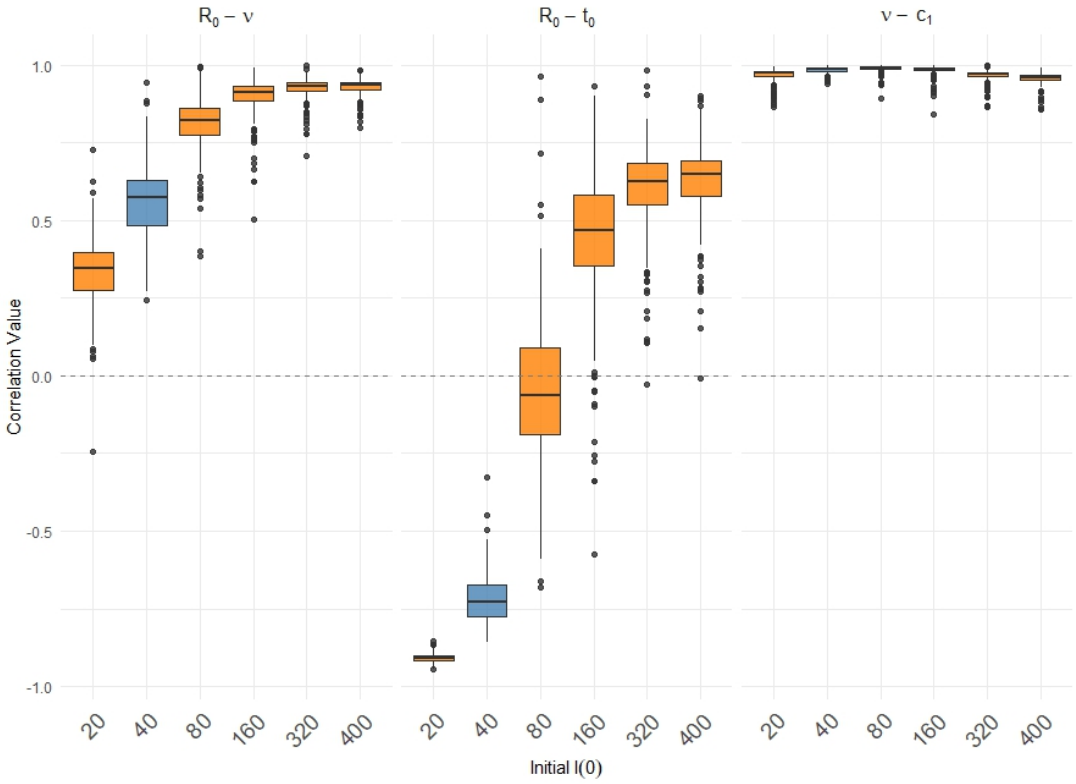

Figure S7: Correlation distributions in terms of initial condition ($\mathsf{I}(0) \in \{40, 80, 160, 320, 400\}$) for intervention parameter $c_1 = 0.3$. Cases with initial conditions used in the baseline analysis ($\mathsf{I}(0) = 40$) are marked in blue.

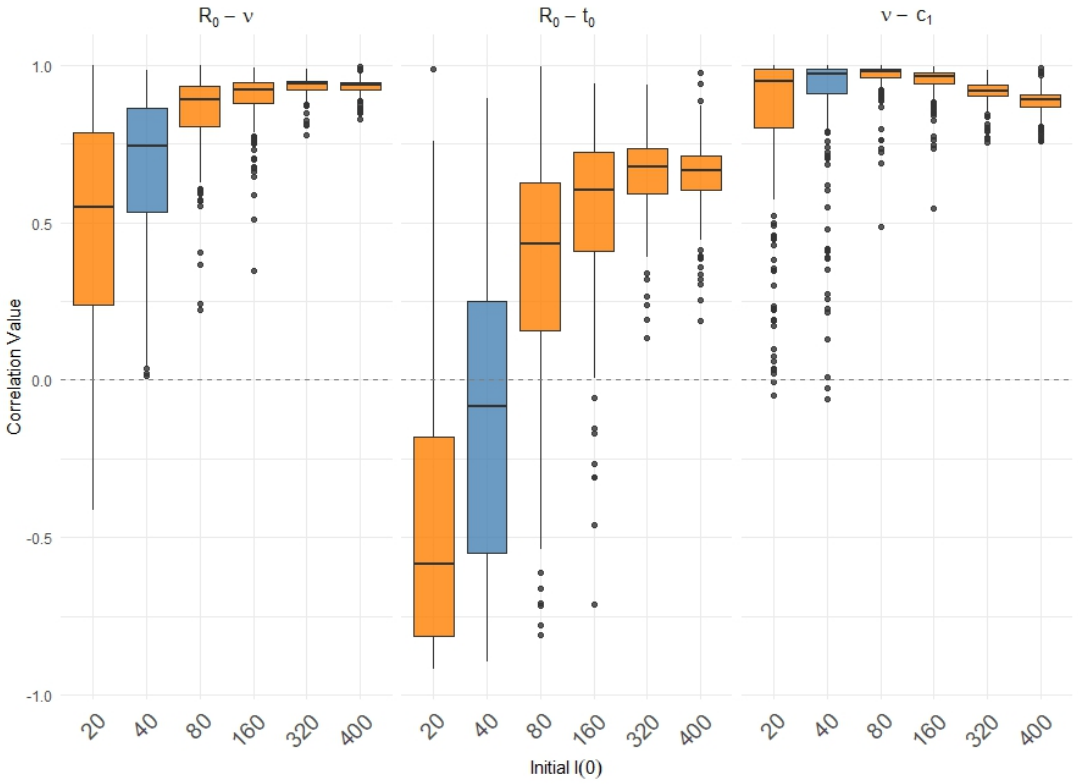

Figure S8: Correlation distributions in terms of initial condition ($\mathsf{I}(0) \in \{20, 40, 80, 160, 320, 400\}$) for intervention parameter $c_1 = 0.2$. Cases with initial conditions used in the baseline analysis ($\mathsf{I}(0) = 40$) are marked in blue.

Correlations between $\mathcal{R}_0$ and ν increase sharply with the initial infected population, highlighting the importance of information provided by the initial growth phase for parameter identification. Correlations between ν and c_1 are high across the entire range of initial conditions. Correlations between $\mathcal{R}_0$ and t_0 are more difficult to synthesise. They are negative for small initial infected conditions and positive when the initial numbers infected are higher. Since both increasing $\mathcal{R}_0$ and t_0 increase growth, we might expect the correlation to be negative. An explanation for how it becomes positive when initial conditions are higher does not appear to be simple and it may involve more complex interactions where other parameters may need to be invoked.

S4 Profile likelihood analysis

Here we provide materials supplementing Section 6 ("Profile likelihood analysis and parameter identifiability") of the main text.

S4.1 Profile likelihood curves

Figures S9–S12 show representative profile likelihood curves for all four key parameters from selected sample dataset, comparing the single-epidemic versus two-epidemic approaches.

The profile likelihood curves reveal substantially narrower confidence intervals in the two-epidemic approach compared to the single-epidemic approach across all parameters. This is particularly evident for the coefficient of variation parameter (ν), where the confidence interval width is reduced by approximately 93%, and for the intervention strength parameter (c_1), where the confidence interval narrows by approximately 85.4%.

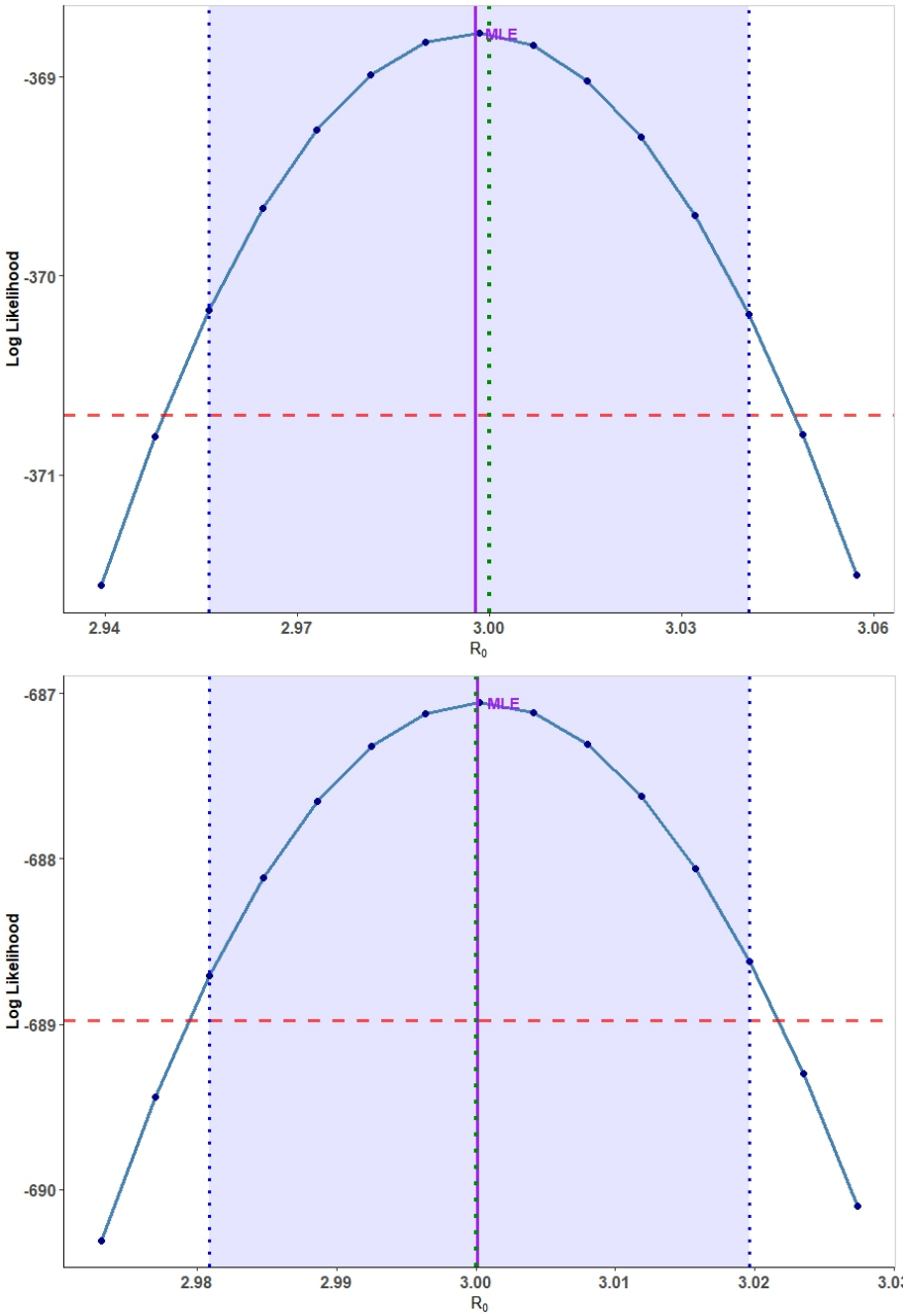

Figure S9: Profile likelihood curves for the basic reproduction number $\mathcal{R}_0$ using single-epidemic (top) and two-epidemic (bottom) models. Horizontal dashed red lines represent the 95% confidence threshold. The green dotted vertical line shows the true value, and the purple solid line the MLE.

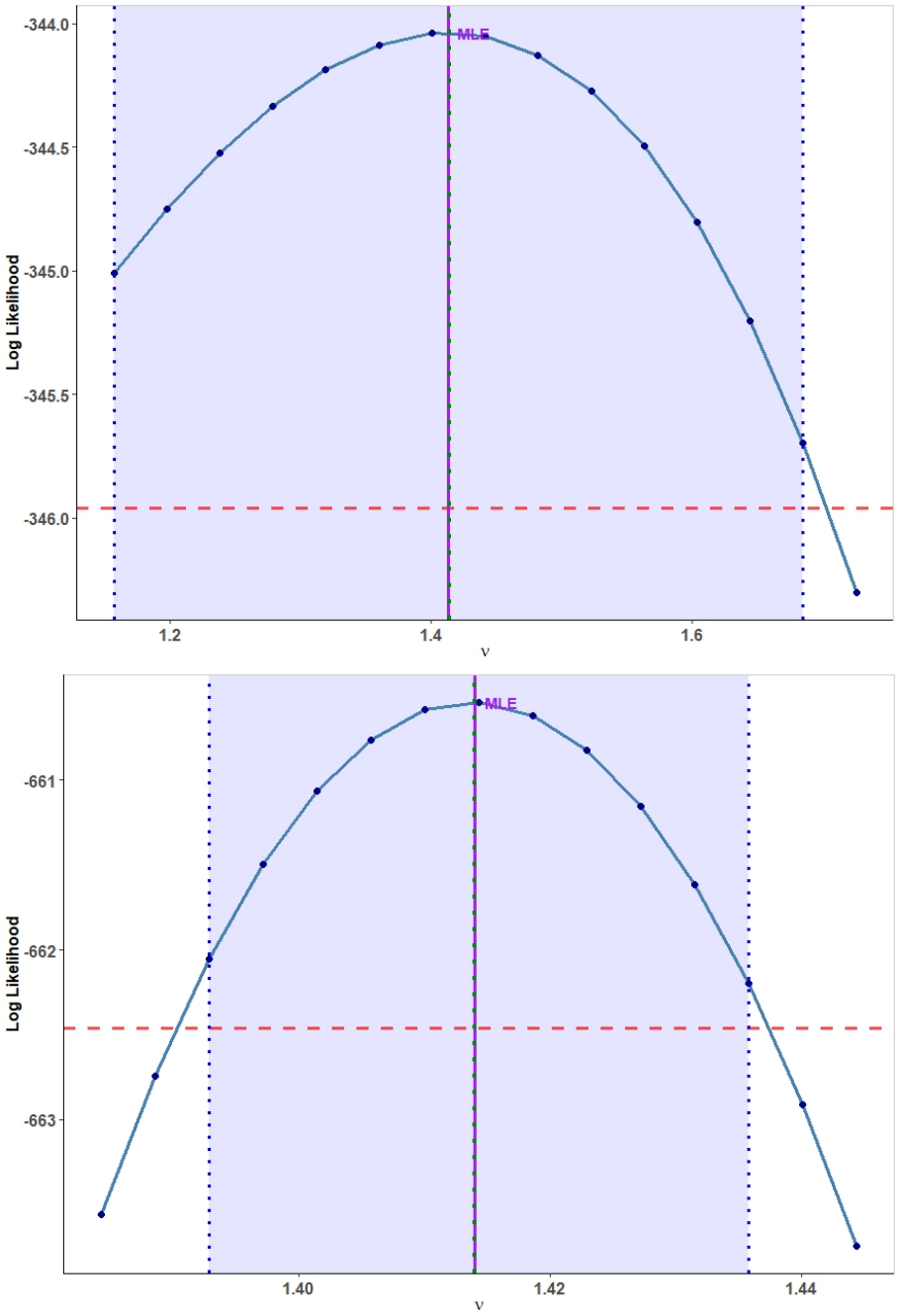

Figure S10: Profile likelihood curves for the coefficient of variation ν using single-epidemic (top) and two-epidemic (bottom) models. Horizontal dashed red lines represent the 95% confidence threshold. The green dotted vertical line shows the true value, and the purple solid line the MLE.

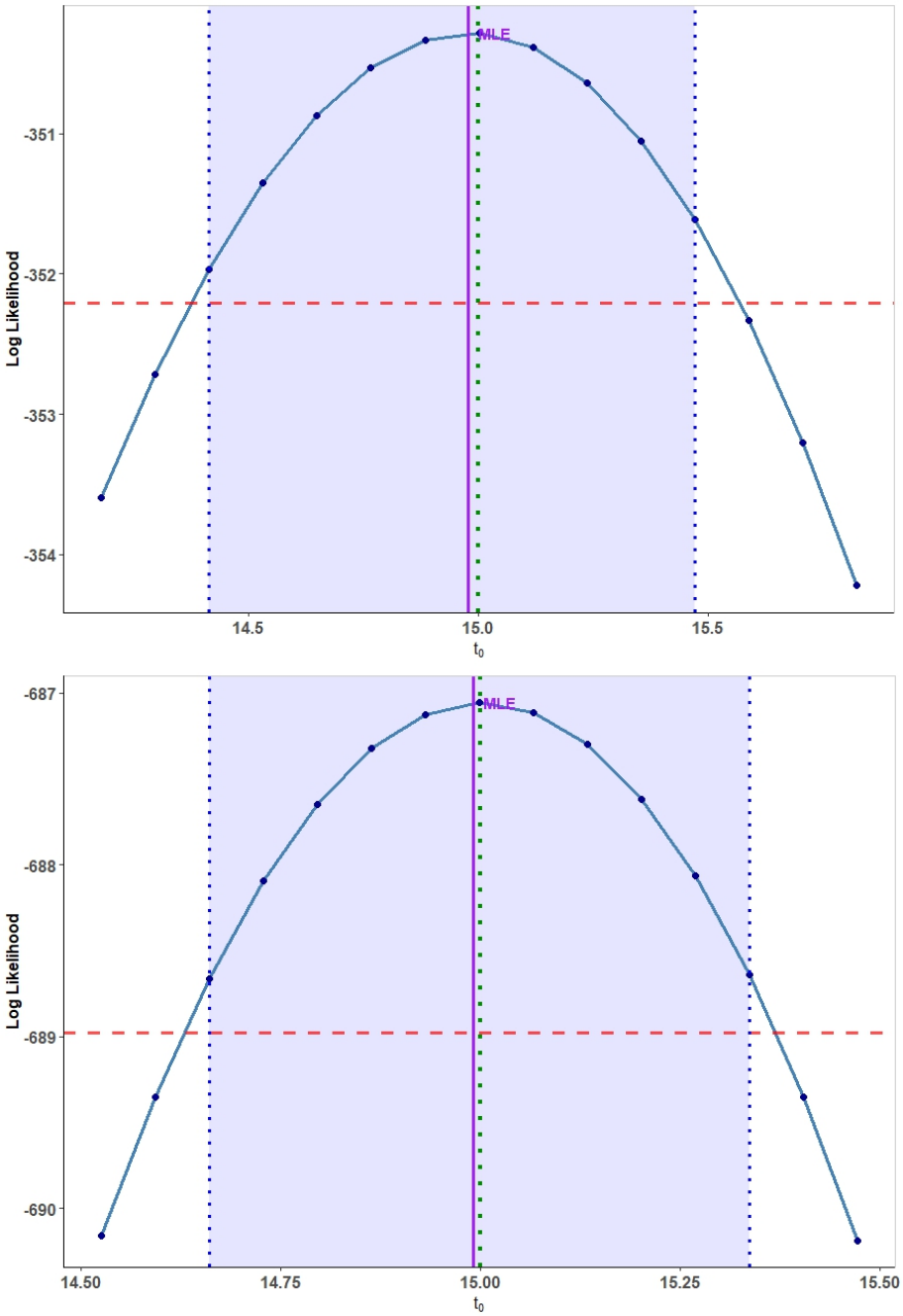

Figure S11: Profile likelihood curves for behavioural change timing t_0 under single-epidemic (top) and two-epidemic (bottom) models. Horizontal dashed red lines represent the 95% confidence threshold. The green dotted vertical line shows the true value, and the purple solid line the MLE.

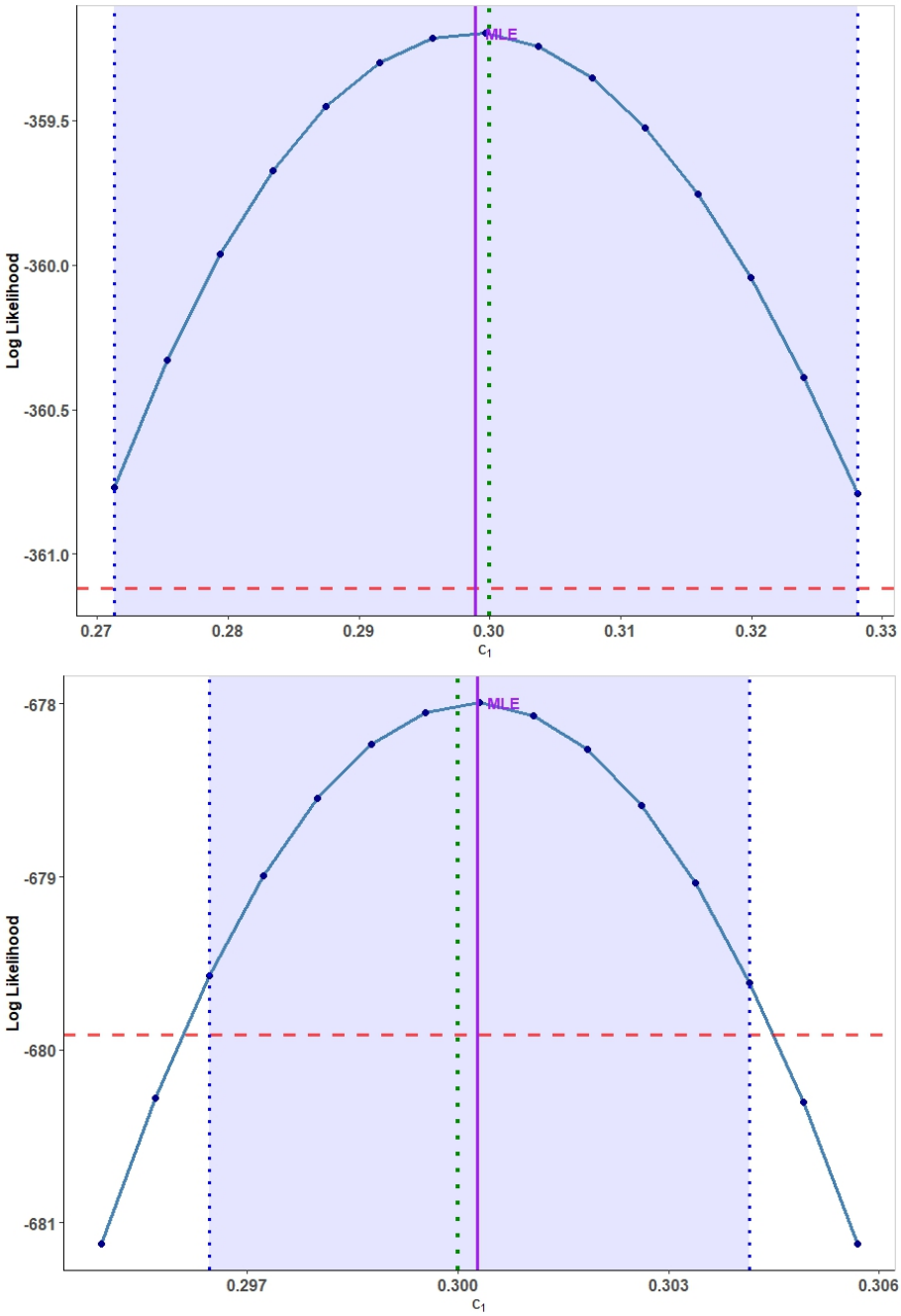

Figure S12: Profile likelihood curves for intervention strength c_1 under single-epidemic (top) and two-epidemic (bottom) models. Horizontal dashed red lines represent the 95% confidence threshold. The green dotted vertical line shows the true value, and the purple solid line the MLE.

S4.2 Accuracy and precision

Figure S13 displays the median and 95% confidence intervals of parameter estimates from all the datasets. The results demonstrate that the two-epidemic approach provides more precise estimates (narrower confidence intervals) and lower relative bias for most parameters.

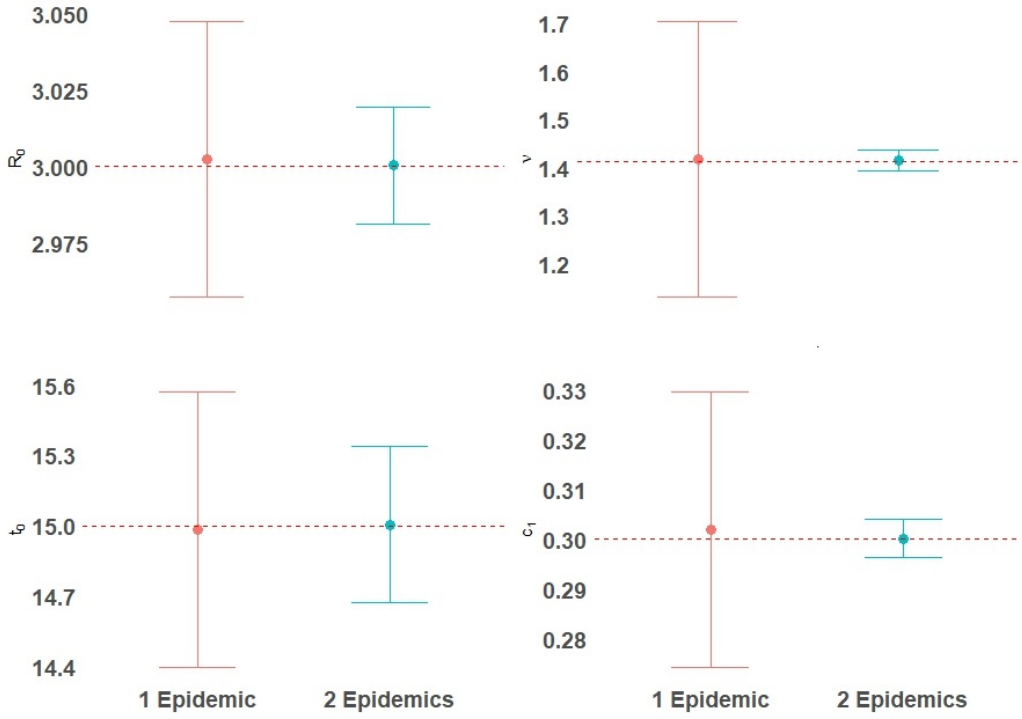

Figure S13: Comparison of parameter estimates between single-epidemic and two-epidemic approaches. Error bars represent the average 95% confidence intervals.

Tables S1 and S2 further examine the statistical performance of single and two-epidemic approaches, including relative bias and coverage probabilities of the confidence intervals across all 200 datasets.

Table S1: Statistical performance metrics for the single-epidemic approach.

Parameter	Single epidemic		
	Relative bias (%)	Relative width (%)	Coverage (%)
$\mathcal{R}_0$	0.086	3.01	90.36
ν	0.303	40.41	89.34
t_0	-0.112	7.83	89.85
c_1	0.624	18.46	90.36

Table S2: Statistical performance metrics for the two-epidemic approach.

Parameter	Two epidemics		
	Relative bias (%)	Relative width (%)	Coverage (%)
$\mathcal{R}_0$	0.014	1.27	93.0
ν	0.038	3.04	93.5
t_0	0.036	4.44	93.0
c_1	0.052	2.56	91.0

S4.3 Numerical stability

Figure S14 displays the distribution of the Hessian matrix condition numbers for the single and two-epidemic approaches.

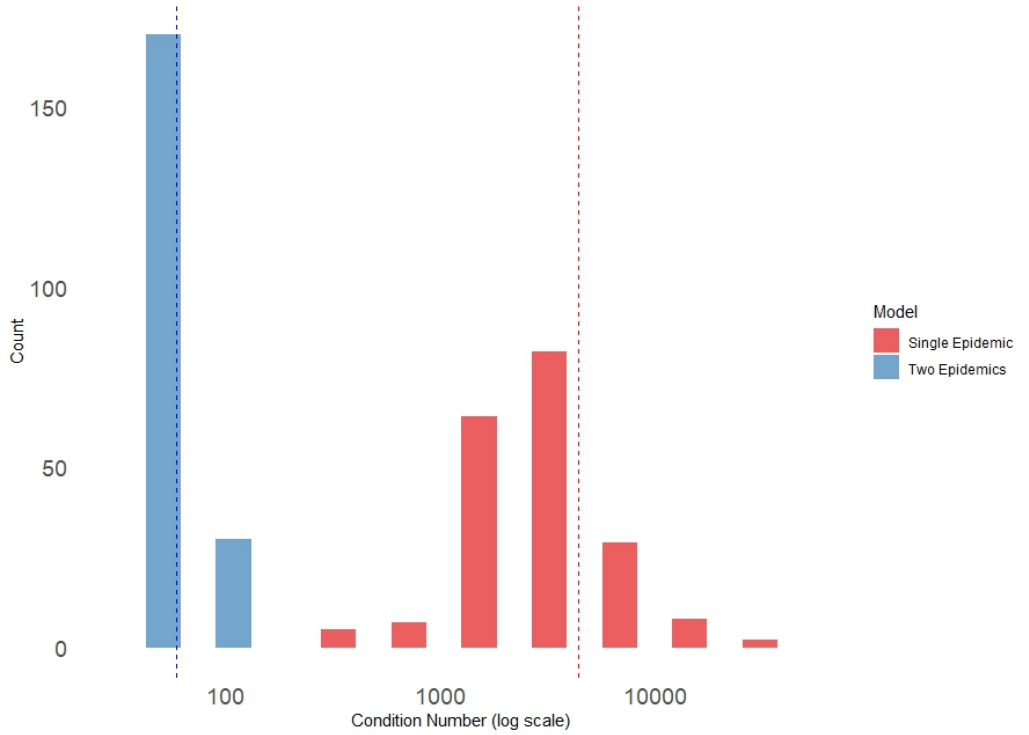

Figure S14: Distribution of Hessian matrix condition numbers (log scale) comparing single and two-epidemic approaches.