



Research article

Revisiting the modeling of the 1918 influenza pandemic's three mortality waves in England and Wales: Issues with the integration method and model formulation

Li Wen, Linxi He and Daihai He*

Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR, China

* **Correspondence:** Email: daihai.he@polyu.edu.hk.

Abstract: Mathematical modeling serves as an indispensable and effective tool in epidemiology to quantify the characteristics of an epidemic, predict transmission trends, and evaluate interventions. However, within the plug-and-play framework, model fitting results may be simultaneously affected by numerical integration algorithms and mechanistic model structure. Using data from the 1918–1919 influenza pandemic across 334 administrative units in England and Wales, this study investigated whether numerical integration algorithms or mechanistic model formulation plays a more important role in the mathematical modeling. We compared three numerical schemes (multinomial Euler, deterministic Euler, and fourth-order Runge-Kutta) and further tested whether different functional forms of the temperature-response term (bounded versus exponential form) and the behavioral-response term affect model performance. Our results showed that numerical algorithms and temperature formulation have only limited effects on the overall inference. By comparison, the behavioral-response formulation has a much stronger impact on model performance. These findings underscore that while changing numerical algorithms affects likelihood values, the structural specification of key mechanistic terms plays a far more pivotal role in determining model performance.

Keywords: epidemiological models; numerical integration algorithms; temperature response; behavioral response; 1918–1919 influenza pandemic; optimization

1. Introduction

Mathematical modeling, as an indispensable and effective tool, plays a central role in epidemiology by quantifying various characteristics of an epidemic, predicting transmission trends, and evaluating possible interventions, especially during pandemics with broad population-level impacts. In the study of infectious disease dynamics, continuous-time stochastic models based on partially observed Markov processes (POMPs) are widely used. The formulation and estimation of

key model parameters—most notably the time-varying transmission rate function—play a pivotal role in determining the accuracy and interpretability of epidemic models. The functional form chosen for this parameter may substantially influence model behavior, goodness-of-fit, and computational stability. In addition, mismatches between the continuous-time structure of the model and the discrete nature of observed data may give rise to technical challenges in model fitting. These issues also interact closely with the choice of numerical integration methods, thereby underscoring the need for careful consideration of both functional specification and computational strategy.

Reference [1] proposed a “plug-and-play” inference framework for disease dynamics, allowing likelihood-based inference for complex epidemic models without requiring closed-form transition densities. This approach provides considerable flexibility for mechanistic epidemic models and was shown to be effective in both large and small population systems. Subsequent studies further developed the methods and tools for POMP-based inference. For example, the R package *pomp*, developed by [2], provides a general framework for constructing, simulating, and fitting nonlinear and non-Gaussian POMP models. Reference [3] also developed iterated filtering methods for maximum-likelihood inference in dynamic latent variable models. These methods allow researchers to estimate parameters for complex epidemic models when the likelihood is difficult to compute analytically but simulation from the model is feasible.

However, even within the same plug-and-play framework, model fitting results may still be influenced by two factors. On the one hand, continuous-time epidemic models require numerical integration during simulation and likelihood evaluation. Therefore, different numerical integration algorithms, such as multinomial Euler, deterministic Euler, and fourth-order Runge-Kutta (RK4), may lead to different likelihood values and parameter estimates. On the other hand, the functional forms of key mechanistic terms in the model may also affect the model’s ability to explain epidemic fluctuations, especially the temperature-response term and the behavioral-response term.

This study aims to investigate whether numerical integration algorithms or mechanistic model formulation plays a more important role in mathematical modeling. We use the 1918–1919 influenza pandemic studied by [4] as an example. In that paper, He et al. assumed that the transmission rate was jointly affected by temperature forcing, school-term forcing, and behavioral response to mortality risk. Based on this framework, this study compares three numerical integration algorithms under the same mechanistic model framework, and further tests whether different functional forms of the temperature-response term and the behavioral-response term affect model performance.

The results show that numerical algorithms and temperature formulation have only limited effects on the overall inference. By comparison, the behavioral-response formulation has a much stronger impact on model performance.

2. Methodology

2.1. Data

Following He et al. [4], we used reported weekly influenza deaths during the 1918–1919 pandemic in 334 administrative units of England and Wales. The mortality data were obtained from the dataset compiled by Johnson [5].

We also used central England daily temperature data obtained from the UK Met Office (<https://www.metoffice.gov.uk>), which were directly incorporated into the model as a time-varying

covariate. The epidemic process was simulated at a daily time step, and daily predicted mortality was accumulated over each seven-day observation interval to correspond to the observed weekly mortality used for likelihood evaluation. In this example, temperature data from London were applied as a proxy for other British cities. Although this simplification does not capture fine-scale spatial variation in temperature, it is likely to be an acceptable approximation given the relatively limited geographical extent of the United Kingdom and the strong spatial correlation of temperature patterns across regions.

2.2. Model framework and comparison design

We adopted the compartmental model framework of [4], which is a susceptible-infectious-recovered (SIR) model with additional mortality and risk-perception variables. The model contains six state variables: S , I , R , D , M , and P . Here, S , I , and R represent the numbers of susceptible, infectious, and recovered individuals, respectively. The variable D represents individuals who are no longer infectious but are on track to die from influenza, while M represents cumulative influenza deaths. The variable P represents the perceived risk of influenza within the population. It increases when deaths occur and decays naturally in the absence of new deaths.

The mean-field limit of the stochastic model can be written as

$$\frac{dS}{dt} = -\frac{\beta(t, P)SI}{N}, \quad (2.1)$$

$$\frac{dI}{dt} = \frac{\beta(t, P)SI}{N} - \gamma I, \quad (2.2)$$

$$\frac{dR}{dt} = (1 - \phi)\gamma I, \quad (2.3)$$

$$\frac{dD}{dt} = \phi\gamma I - gD, \quad (2.4)$$

$$\frac{dM}{dt} = gD, \quad (2.5)$$

$$\frac{dP}{dt} = gD - \lambda P. \quad (2.6)$$

Here, $N = S + I + R + D + M$ is the initial population size. The parameter ϕ denotes the case fatality proportion. The parameters γ , g , and λ denote the rates at which individuals leave the I , D , and P compartments, respectively. Therefore, γ^{-1} is the mean infectious period, g^{-1} is the mean time from loss of infectiousness to death, and λ^{-1} is the mean duration of the impact of deaths on public perception.

Following the structure of [4], the transmission rate was constructed from three mechanistic components: temperature forcing, school-term forcing, and behavioral response to mortality risk. In our base model, we retained the school-term and behavioral-response components, but modified the temperature-response term to a bounded form to constrain the transmission rate within a biologically reasonable range. The transmission rate was therefore specified as

$$\beta(t, P) = \beta_0 \frac{2}{1 + \exp(\xi T(t))} [1 + \alpha H(t)] \left(1 - \frac{P(t)}{N}\right)^\kappa. \quad (2.7)$$

In this expression, β_0 is the baseline transmission rate. $T(t)$ denotes temperature, and ξ controls the strength of the temperature effect on transmission. The term $1 + \alpha H(t)$ represents school-term forcing,

where α is the amplitude of the school-term effect and $H(t)$ indicates a binary indicator of the school calendar. Transmission is assumed to be lower during school vacations and higher when schools are in session. The final factor, $\left(1 - \frac{P(t)}{N}\right)^\kappa$, represents human behavioral responses to mortality risk. The parameter κ controls the strength of this behavioral response. The parameters b_1 and b_2 were used to describe the start and end dates of the summer school vacation, respectively.

Together, the estimated parameters included β_0 , α , ξ , λ , ϕ , κ , b_1 , and b_2 . Based on this model, we designed the following comparisons to evaluate the effects of numerical implementation and mechanistic model formulation.

(1) Numerical integration algorithms. First, we compared three numerical integration algorithms for simulating the latent epidemic process defined by the compartmental equations: multinomial Euler, deterministic Euler, and RK4. These methods were used to update the state variables S , I , R , D , M , and P over time, while the same transmission-rate formulation and parameter structure were kept unchanged.

The multinomial Euler algorithm simulates compartmental transitions stochastically and incorporates demographic or process noise. The deterministic Euler algorithm updates the state variables using a first-order deterministic approximation to the ordinary differential equations. RK4 is also deterministic, but uses a higher-order integration scheme to approximate the continuous-time epidemic dynamics.

(2) Temperature-response term. Second, we examined whether the model fit was sensitive to the functional form of the temperature-response term. In the base model, the temperature effect was represented by the bounded form

$$\frac{2}{1 + \exp(\xi T(t))}.$$

This term was replaced by the exponential form used in the original He et al. framework:

$$\exp(-\xi T(t)).$$

Both functions describe the influence of temperature on influenza transmission. For $\xi > 0$, both terms decrease as temperature increases, implying that higher temperature reduces transmission and lower temperature increases transmission. However, the bounded form limits the magnitude of the temperature effect, while the exponential form allows the temperature effect to increase without an upper bound under extremely low temperatures.

(3) Behavioral-response term. Third, we examined the importance of the behavioral-response formulation. In the base model, behavioral response was represented by the power-form term

$$\left(1 - \frac{P(t)}{N}\right)^\kappa.$$

This term represents a behavioral response to the relative level of perceived mortality risk, with κ controlling the strength of the response.

To test whether an alternative behavioral formulation could reproduce the observed mortality waves, this term was replaced by a Holling type-II-like form:

$$\frac{\kappa}{\kappa + P(t)}.$$

This alternative formulation represents a smoother saturating response, where κ can be interpreted as the level of perceived risk at which the behavioral term reduces transmission by one half.

(4) Joint modification of temperature and behavioral terms. Finally, we simultaneously modified the temperature-response and behavioral-response components. Specifically, the original transmission-rate formulation was changed from

$$\beta(t) = \beta_0 \frac{2}{1 + \exp(\xi T(t))} \left(1 - \frac{P(t)}{N}\right)^\kappa [1 + \alpha H(t)]$$

to

$$\beta(t) = \beta_0 \exp(-\xi T(t)) \frac{\kappa}{\kappa + P(t)} [1 + \alpha H(t)].$$

This joint comparison was used to determine whether changes in model performance were mainly driven by the temperature-response formulation or by the behavioral-response formulation.

2.3. Model fitting and parameter estimation

Following [4], the model was formulated as a partially observed Markov process (POMP), in which the latent epidemic process was linked to the observed weekly influenza mortality through a stochastic observation model. The latent epidemic dynamics were propagated using the Euler integration algorithm with a daily integration time step,

$$\Delta t = \frac{1}{365} \text{ year.} \quad (2.8)$$

The model-generated weekly mortality was obtained by accumulating daily deaths over each observation interval,

$$\widehat{C}_t = \int_{1 \text{ week}} gD(t) dt. \quad (2.9)$$

The reported weekly mortality was assumed to follow a negative binomial distribution,

$$C_t \sim \text{NB}(\widehat{C}_t, \widehat{C}_t(1 + \tau \widehat{C}_t)), \quad (2.10)$$

where C_t denotes the observed weekly mortality, $\widehat{C}_t$ is the model-predicted weekly mortality, and τ is the over-dispersion parameter.

The log-likelihood function was defined as

$$\log L(\Theta) = \sum_{t=1}^n \log [f_{\text{NB}}(C_t | C_{1:(t-1)}, \Theta)], \quad (2.11)$$

where $f_{\text{NB}}(\cdot)$ denotes the negative binomial probability mass function, Θ denotes the vector of estimated parameters, and n is the total number of weekly observations.

Model fitting was implemented using the *pomp* package in R under the plug-and-play framework. Likelihood evaluation was performed using sequential Monte Carlo (particle filtering), and maximum likelihood estimation was carried out using the iterated filtering (IF2) algorithm. Unless otherwise stated, the optimization settings were identical for all model comparisons, with $N_p = 20,000$ particles,

cooling factor = 0.95, cooling constant (CC) = 2, and initial temperature parameter $T_0 = 46$. The IF2 algorithm was initialized from manually specified biologically plausible parameter values. Parameter estimates obtained from each IF2 run were subsequently used as the initial values for the next IF2 run, and this procedure was repeated until the log-likelihood stabilized, with further iterations resulting in negligible improvement (<0.5 log-likelihood units).

The estimated parameters included β_0 , α , ξ , λ , ϕ , κ , b_1 , and b_2 . The initial susceptible proportion was constrained to be at least 70% of the total population, while the initial infectious proportion was constrained to be at least 0.01%. The initial proportions of the susceptible, infectious, and recovered populations satisfied

$$S_0 + I_0 + R_0 = 1. \quad (2.12)$$

The initial values of the disease-severity compartment (D_0), behavioral-response compartment (P_0), and cumulative deaths were fixed at zero. Parameters α and ϕ were constrained to the interval $(0, 1)$, whereas all remaining parameters were constrained to be positive. The summer vacation parameters b_1 and b_2 were constrained such that the estimated school summer vacation period remained within the interval from June 23, 1918, to October 15, 1918. No additional parameter restrictions were imposed.

To ensure fair comparisons, all models were fitted using identical optimization settings and estimation procedures. Consequently, differences in model performance can be attributed solely to the numerical integration algorithm or the mechanistic model formulation.

For model comparison, the corrected Akaike information criterion (AICc) was calculated as

$$\text{AICc} = -2 \log L + \frac{2n_p n}{n - n_p - 1}, \quad (2.13)$$

where n_p denotes the number of estimated parameters.

The relative change of each parameter estimate was calculated as

$$\Delta x = \frac{x_1 - x_0}{x_0}, \quad (2.14)$$

where x_0 denotes the estimate from the reference model and x_1 denotes the estimate from the comparison model. No estimate is close to zero.

2.4. Model evaluation metrics

For each administrative unit and each model specification, we fitted the model and computed the corresponding log-likelihood and parameter estimates. Model performance was evaluated using changes in log-likelihood and relative changes in parameter estimates. Because all compared models contained the same number of estimated parameters, log-likelihood values could be directly compared without information-criterion adjustment.

For each comparison, the change in log-likelihood was defined as

$$\Delta LL = LL_{\text{candidate}} - LL_{\text{reference}}, \quad (2.15)$$

where $LL_{\text{candidate}}$ is the log-likelihood of the alternative numerical algorithm or modified model and $LL_{\text{reference}}$ is the log-likelihood of the corresponding reference model. A positive ΔLL indicates that

the candidate model achieved a higher log-likelihood, whereas a negative ΔLL indicates worse model performance.

For the estimated parameters, relative changes were calculated to assess the stability of parameter inference under different algorithms and functional formulations. For each comparison, we summarized ΔLL and the relative parameter changes across the 334 administrative units using the minimum, first quartile, median, mean, third quartile, and maximum.

3. Results

In all comparisons, ΔLL denotes the change in log-likelihood between the candidate model and the corresponding reference model. The remaining rows in each table report the relative changes in estimated parameters. The parameters b_1 and b_2 denote the estimated start and end dates of the summer school vacation.

3.1. Effects of numerical integration algorithms

We first evaluated the effect of numerical integration algorithms while keeping the transmission-rate formulation in the reference model unchanged (Eq (7)). As shown in Table 1, replacing the multinomial Euler algorithm with deterministic Euler consistently improved the model fit across the 334 administrative units. The median increase in log-likelihood was 4.115, with a mean increase of 4.119. However, the improvement was generally modest, as reflected by the first quartile of zero and the relatively small interquartile range. Although a few administrative units exhibited larger improvements (maximum $\Delta LL = 13.348$), most showed only limited gains in likelihood.

Table 1. Impacts of using deterministic Euler in place of multinomial Euler.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	0	0	4.115	4.119	7.283	13.348
$\Delta\beta_0$	-0.143	0	0.024	0.059	0.103	0.357
$\Delta\alpha$	-0.209	-0.029	0	-0.013	0	0.155
$\Delta\xi$	-0.180	0	0	0.036	0.070	0.380
$\Delta\lambda$	-0.269	0	0	0.009	0.035	0.370
$\Delta\phi$	-0.316	-0.008	0	0.002	0.016	0.277
$\Delta\kappa$	-0.271	-0.086	0	-0.033	0	0.380
Δb_1	-0.129	0	0	0.007	0.009	0.148
Δb_2	-0.098	0	0	0.032	0.054	0.553

The estimated parameters were largely unaffected by this numerical change. The median relative change was zero for most parameters, indicating that the parameter estimates remained nearly identical across the majority of the administrative units. Small systematic changes were observed for the transmission coefficient β_0 , whereas the reporting probability ϕ , school vacation parameters (b_1 and b_2), and the seasonal forcing parameter α remained highly stable. Larger relative deviations occasionally occurred for the temperature-response parameter ξ and the behavioral-response parameter κ , although these were confined to a small number of administrative units and did not alter the overall pattern of parameter estimation.

A similar pattern was observed when RK4 replaced multinomial Euler (Table 2). RK4 produced a slightly greater improvement in likelihood than deterministic Euler, with a median ΔLL of 4.723 and a mean of 4.398. Nevertheless, the distribution of likelihood gains remained similar, with many administrative units exhibiting only marginal improvements. Parameter estimates also remained highly consistent between the two implementations. Median relative changes were zero for nearly all parameters, indicating that the higher-order numerical approximation had little influence on epidemiological inference.

Table 2. Impacts of using RK4 in place of multinomial Euler.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	0	0	4.723	4.398	7.657	13.492
$\Delta\beta_0$	-0.173	0	0.032	0.080	0.144	0.485
$\Delta\alpha$	-0.405	-0.053	0	-0.017	0.011	0.540
$\Delta\xi$	-0.233	0	0	0.101	0.154	1.655
$\Delta\lambda$	-0.371	-0.025	0	0.021	0.063	0.606
$\Delta\phi$	-0.532	-0.030	0	-0.001	0.016	0.369
$\Delta\kappa$	-0.446	-0.113	0	0.031	0	3.895
Δb_1	-0.093	0	0	0.016	0.024	0.303
Δb_2	-0.112	0	0	0.038	0.062	0.586

The direct comparison between RK4 and deterministic Euler is summarized in Table 3. In contrast to the previous comparisons, the difference between these two deterministic integration algorithms was minimal. The median increase in log-likelihood was only 0.001 and the mean increase was 0.279, indicating that both methods produced nearly identical likelihood surfaces for most administrative units. Parameter estimates were likewise highly consistent, with relative changes generally close to zero for all estimated parameters. These results indicate that increasing the numerical accuracy from first-order Euler to fourth-order Runge-Kutta had only negligible influence on parameter estimation.

Table 3. Impacts of using RK4 in place of deterministic Euler.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	-0.940	0	0.001	0.279	0.266	9.556
$\Delta\beta_0$	-0.103	0	0	0.019	0.032	0.320
$\Delta\alpha$	-0.282	-0.021	0	-0.007	0.008	0.393
$\Delta\xi$	-0.148	0	0	0.058	0.054	1.586
$\Delta\lambda$	-0.201	-0.014	0	0.010	0.021	0.404
$\Delta\phi$	-0.316	-0.016	0	-0.004	0	0.248
$\Delta\kappa$	-0.326	-0.023	0	0.060	0.009	3.895
Δb_1	-0.049	0	0	0.008	0.009	0.220
Δb_2	-0.172	-0.002	0	0.006	0.008	0.586

Overall, Tables 1–3 demonstrate that deterministic Euler and RK4 consistently achieved slightly higher likelihood values than the multinomial Euler implementation. However, these likelihood improvements were generally modest and were accompanied by only minor changes in parameter estimates. Across the 334 administrative units, the inferred transmission dynamics and

epidemiological conclusions remained highly robust to the choice of numerical integration algorithm. Although deterministic implementations yielded slightly better likelihood optimization, the multinomial Euler algorithm remains biologically appealing because it explicitly incorporates demographic (process) stochasticity.

3.2. Effects of modifying the temperature-response term

The impacts of this modification are summarized in Table 4. The median change in log-likelihood was 0.042, while the mean change was -0.043 . The interquartile range was narrow (from -0.162 to 0.190), indicating that the modified temperature-response function had little influence on model fit across the 334 administrative units.

Table 4. Impacts of replacing the bounded temperature-response term with the exponential form.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	-2.427	-0.162	0.042	-0.043	0.190	1.160
$\Delta\beta_0$	-0.330	-0.027	0.041	0.030	0.096	0.837
$\Delta\alpha$	-0.448	-0.035	-0.006	0.029	0.045	0.928
$\Delta\xi$	-0.681	-0.366	-0.280	-0.251	-0.176	2.612
$\Delta\lambda$	-0.214	-0.020	0.015	0.070	0.083	1.079
$\Delta\phi$	-0.838	-0.060	-0.035	-0.044	-0.010	0.156
$\Delta\kappa$	-0.119	-0.048	-0.010	7.311	8.767	96.442
Δb_1	-0.459	-0.010	0.002	-0.004	0.017	0.091
Δb_2	-0.541	-0.009	0.009	-0.007	0.028	0.895

As also shown in Table 4, among all estimated parameters, the largest systematic change occurred in the temperature-response parameter ξ , whose median relative change was -0.280 . This result is expected because the functional form of the temperature effect was directly modified. By contrast, the median relative changes of the remaining parameters were close to zero, suggesting that the alternative temperature formulation primarily affected the estimation of ξ while most parameters showed limited changes in median relative differences. Only several parameters exhibited occasional large deviations in a small number of administrative units. Overall, modifying the temperature-response function had only a limited influence on both model fit and parameter estimation.

3.3. Effects of modifying the behavioral-response term

The impacts of this modification are summarized in Table 5. Compared with the reference model, replacing the behavioral-response term substantially reduced model performance. The median change in log-likelihood was -0.479 , whereas the mean reduction reached -7.379 . The first quartile was -9.508 , and the minimum reached -75.891 , indicating that a substantial subset of administrative units experienced marked deterioration in likelihood.

The behavioral modification also produced noticeably larger changes in several parameter estimates than those observed for either the numerical integration algorithms or the temperature-response modification. In particular, the behavioral-response parameter κ exhibited a median relative change of -0.119 and a first quartile of -0.853 , while relatively large changes were

Table 5. Impacts of replacing the behavioral-response term with the Holling type-II-like formulation.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	-75.891	-9.508	-0.479	-7.379	0	0.173
$\Delta\beta_0$	-0.757	-0.229	0	-0.045	0.034	2.798
$\Delta\alpha$	-0.634	-0.067	0	0.018	0.039	2.008
$\Delta\xi$	-0.921	-0.133	0	0.007	0.060	3.262
$\Delta\lambda$	-0.896	-0.218	-0.012	-0.080	0.007	1.777
$\Delta\phi$	-0.678	-0.011	0	1.663	0.162	47.401
$\Delta\kappa$	-0.995	-0.853	-0.119	-0.309	0	2.899
Δb_1	-0.233	-0.001	0	0.072	0.040	2.297
Δb_2	-0.369	-0.004	0.002	0.166	0.194	1.679

also observed for β_0 , λ , and several other parameters. These results indicate that modifying the behavioral-response mechanism influenced both model fit and parameter estimation.

To further examine the consequences of the behavioral modification, we compared simulated mortality trajectories with the observed mortality data. As shown in Figure 1, the modified model failed to reproduce the third mortality wave in some administrative units, with the simulated mortality remaining substantially below the observed third-wave peak. Nevertheless, Figure 2 demonstrates that the same modified formulation remained capable of reproducing the third mortality wave in other administrative units. Thus, misspecifying the behavioral-response mechanism did not uniformly degrade model performance across all locations but caused severe fitting failures in a subset of cities.

We also examined the relationship between population size and the corresponding change in log-likelihood (Figure 3). Larger reductions in log-likelihood tended to occur more frequently among smaller administrative units, suggesting that smaller populations were generally more sensitive to misspecification of the behavioral-response mechanism.

3.4. Joint modification of temperature and behavioral terms

The impacts of the joint modification are summarized in Table 6. Relative to the reference model, the median change in log-likelihood was -0.562, the mean change was -6.901, the first quartile was -8.989, and the minimum reached -96.814. These values indicate substantial deterioration in model fit for many administrative units.

Compared with modifying the temperature-response term alone (Table 4), the joint modification produced much larger reductions in log-likelihood. However, the distributions of both the likelihood changes and the relative parameter changes were highly similar to those obtained when only the behavioral-response term was modified (Table 5). For example, the median and mean changes in log-likelihood (-0.562 and -6.901) were close to those obtained after modifying only the behavioral-response term (-0.479 and -7.379, respectively).

These findings indicate that deterioration in model performance was primarily driven by the behavioral-response formulation, not the temperature-response formulation. Although simultaneously modifying both components slightly altered several parameter estimates, it did not substantially worsen performance relative to replacing the behavioral-response term alone.

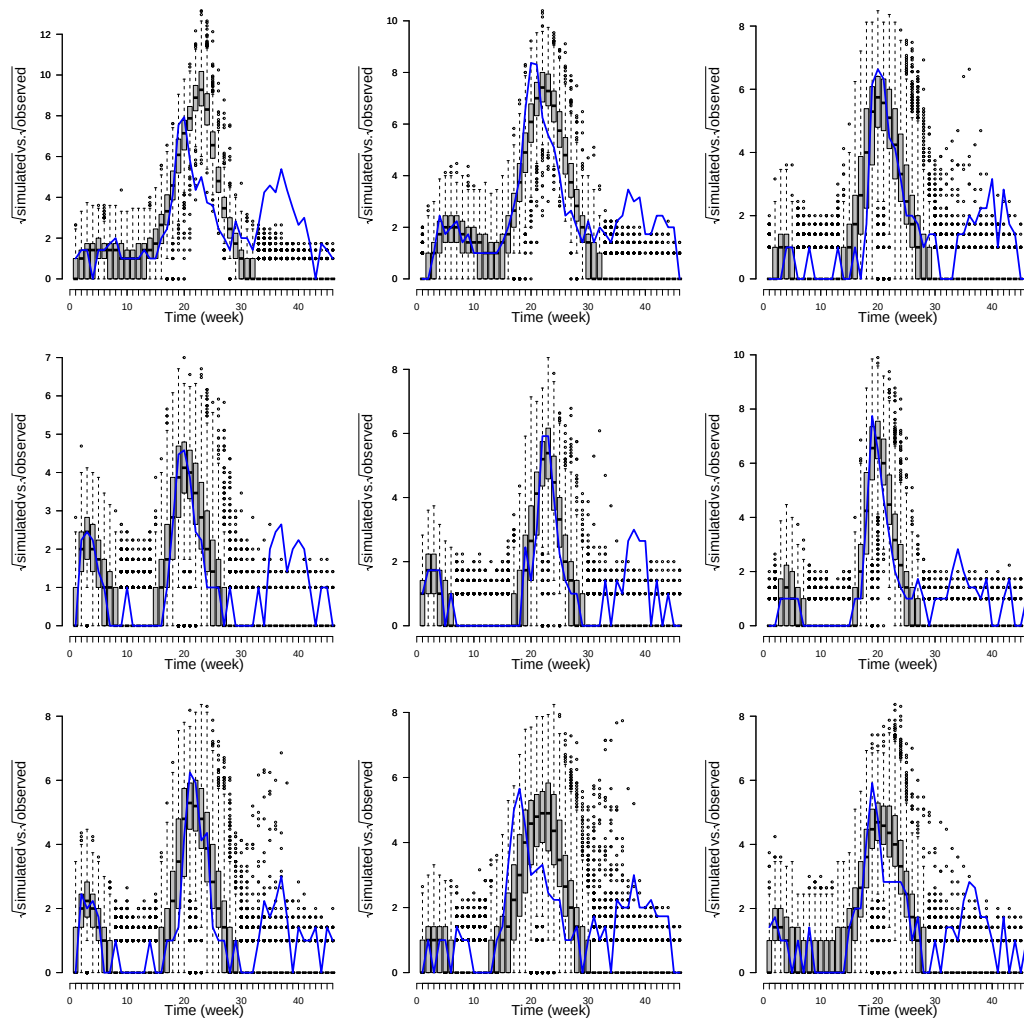


Figure 1. Cities where the model fails to capture the third wave. Boxplots represent 1000 simulated mortality trajectories and the blue curve represents reported deaths.

Taken together, Tables 1–6 demonstrate that the choice of numerical integration algorithm had only a minor influence on model fit and parameter estimation. Likewise, modifying the temperature-response formulation produced only limited changes in likelihood and mainly affected the temperature-response parameter itself. In contrast, replacing the original behavioral-response formulation consistently resulted in substantial reductions in log-likelihood, larger shifts in parameter estimates, and failure to reproduce the third mortality wave in a subset of administrative units. Similar results were obtained when both the temperature-response and behavioral-response terms were modified simultaneously.

Collectively, these findings identify the behavioral-response formulation as the principal mechanistic component underlying the model’s ability to reproduce the multiple mortality waves observed during the 1918–1919 influenza pandemic.

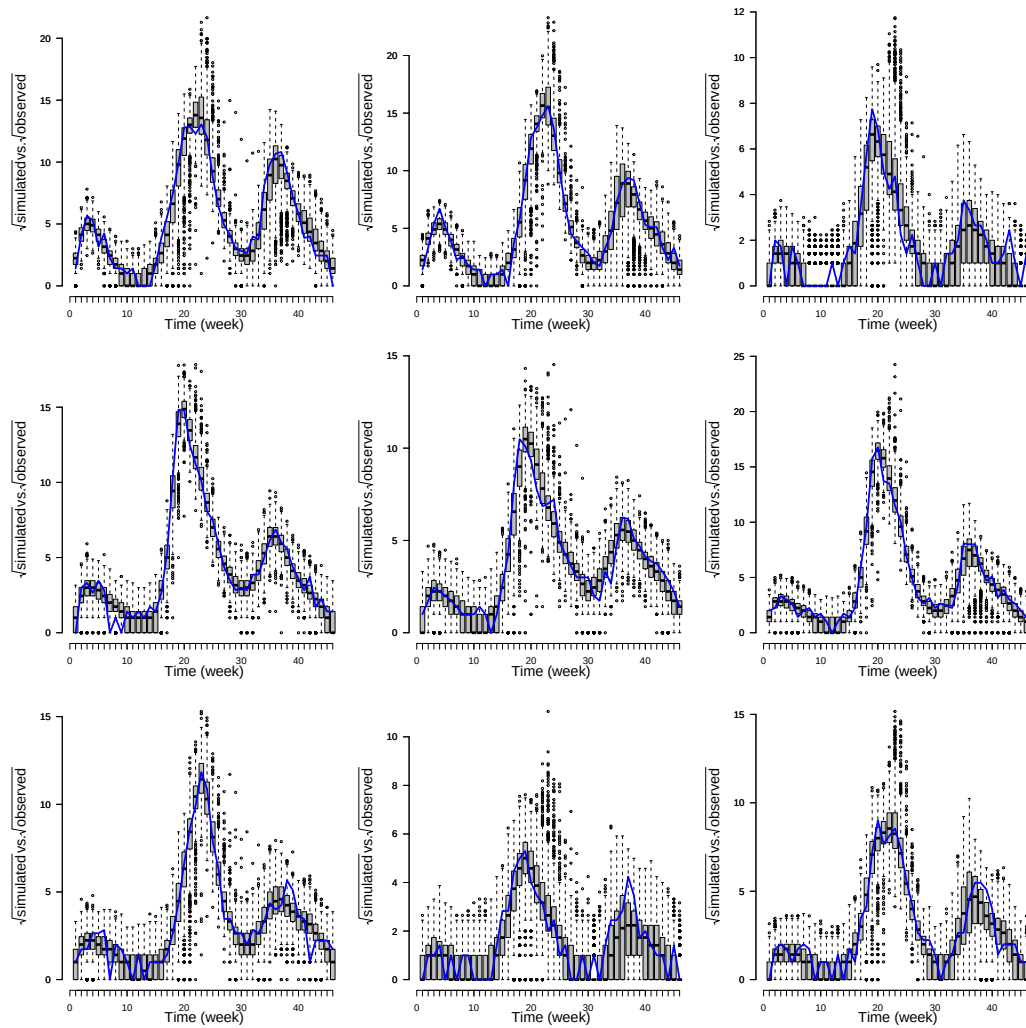


Figure 2. Cities where the model captures the third wave. Boxplots represent 1000 simulated mortality trajectories and the blue curve represents reported deaths.

Table 6. Impacts of simultaneously replacing the temperature-response and behavioral-response terms.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
ΔLL	-96.814	-8.989	-0.562	-6.901	0.063	1.065
$\Delta\beta_0$	-0.749	-0.147	0.016	0.096	0.132	7.915
$\Delta\alpha$	-0.656	-0.071	0.020	0.048	0.123	2.047
$\Delta\xi$	-0.929	-0.407	-0.308	-0.155	-0.044	8.500
$\Delta\lambda$	-0.901	-0.227	-0.040	-0.077	0.066	1.671
$\Delta\phi$	-0.656	-0.036	0.008	1.672	0.142	46.371
$\Delta\kappa$	-0.996	-0.862	-0.161	-0.325	0.017	2.899
Δb_1	-0.235	-0.015	0.001	0.066	0.048	2.393
Δb_2	-0.362	-0.026	0.005	0.154	0.171	1.744

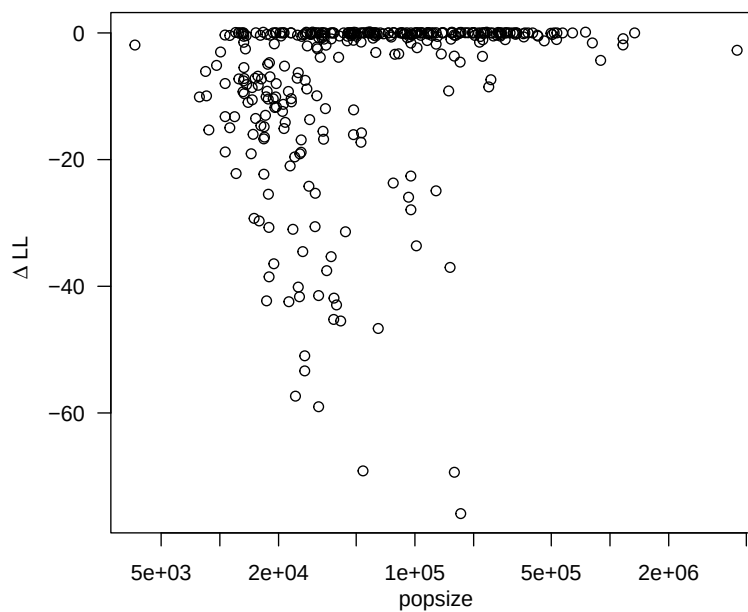


Figure 3. Relationship between population size and log-likelihood difference. The x -axis shows population size and the y -axis shows ΔLL .

4. Discussion

Previous studies on the 1918 influenza pandemic have primarily focused on identifying the roles of behavioral changes, school-term effects, and climatic drivers in shaping epidemic dynamics [4, 6]. However, the relative influence of different mathematical assumptions underlying mechanistic epidemic models has received comparatively less attention. In this study, we systematically investigated uncertainty arising from model formulation by maintaining identical observation models, likelihood functions, optimization procedures, initial conditions, and parameter constraints across all comparisons. This controlled framework allowed us to isolate the effects of individual modeling choices on parameter inference and model performance.

Our results demonstrate that numerical implementation choices and mechanistic assumptions influence epidemic model performance to different extents. Alterations in numerical integration algorithms affected the estimated likelihood values, but these effects were generally limited. Replacing the multinomial Euler algorithm with deterministic Euler or RK4 integration resulted in modest improvements in log-likelihood, suggesting that part of the likelihood difference may arise from the additional demographic stochasticity introduced by the multinomial formulation. Nevertheless, the relatively small changes in parameter estimates indicate that when the underlying mechanistic structure remains unchanged, alternative numerical integration schemes do not substantially alter the main epidemiological interpretation. This finding highlights the importance of considering both statistical performance and biological realism when selecting computational approaches for stochastic epidemic models.

The importance of behavioral responses in shaping pandemic dynamics has been demonstrated in previous studies of the 1918 influenza pandemic. Reactive social-distancing models assume that

individuals modify their contact patterns according to perceived infection risk, creating a feedback mechanism between epidemic severity and transmission intensity [6, 7]. Such endogenous behavioral feedback provides a plausible explanation for multiple epidemic waves without requiring substantial temporal variation in intrinsic transmissibility. Yu et al. further demonstrated that behavioral functions based on the recent proportion of influenza mortality provided better explanations of the observed multi-wave patterns across 334 administrative units in England and Wales compared with alternative formulations [6]. Our findings support and extend this interpretation by showing that modifying the mathematical structure of the behavioral-response function resulted in substantially larger changes in model performance than modifying numerical integration schemes or temperature-response formulations. These results suggest that accurately representing adaptive human behavior is a critical component of mechanistic models of epidemic transmission.

The strong sensitivity to behavioral formulation also has broader implications for modeling recurrent epidemic waves. Behavioral responses introduce endogenous feedback into disease transmission systems: increasing mortality may reduce social contacts and suppress transmission, whereas declining perceived risk may lead to behavioral relaxation and subsequent resurgence. Therefore, behavioral assumptions should be regarded as structural components of epidemic models rather than merely additional covariates. In particular, inappropriate specification of behavioral mechanisms may lead to substantial biases in model fitting and interpretation, even when other model components are accurately represented.

Several limitations should be acknowledged. First, behavioral responses during the 1918 influenza pandemic cannot be directly observed, and therefore all behavioral functions represent simplified approximations of complex social processes. Second, the temperature-response analysis relied on a common temperature series applied across administrative units, which may not fully capture local climatic heterogeneity. Third, other factors influencing human behavior, including public health interventions, information dissemination, social norms, and population mobility, were not explicitly incorporated into the model. Future studies integrating additional historical, behavioral, and mobility-related information may provide a more comprehensive representation of adaptive human responses during pandemics.

In conclusion, this study demonstrates that different modeling assumptions contribute unequally to uncertainty in epidemic inference. While numerical integration schemes and temperature-response formulations produced relatively limited effects on model interpretation, the behavioral-response formulation emerged as a key determinant of model performance. These findings emphasize that mechanistic epidemic models should balance statistical optimization with biological plausibility, and that realistic representation of human behavioral adaptation is essential for understanding multi-wave epidemic dynamics.

Data, materials, and software availability

All code and processed datasets used in this study are publicly available in the Zotero group library (https://www.zotero.org/groups/6628341/1918_influenza_pandemic). This research analyzed previously published data from He et al. [4]. Daily Central England Temperature (CET) data were obtained from the UK Met Office (<https://www.metoffice.gov.uk>). Weekly reported influenza deaths during the 1918–1919 pandemic across 334 administrative units of England and Wales were retrieved from the UK Data Archive (Johnson [5]; <http://dx.doi.org/10.5255/UKDA-SN-4350-1>).

Use of AI tools declaration

The authors declare that no generative artificial intelligence (Generative-AI) tools were used in the writing of the scientific text, core analysis, or data generation of this manuscript. Standard grammar assistance software was utilized solely for language polishing and proofreading to enhance readability.

Acknowledgments

The authors thank the University Research Facility in Big Data Analytics (UBDA) for the computational resources, the Smart Cities Research Institute and the Hong Kong Polytechnic University for the financial support. The work was partially supported by the Hong Kong Research Grants Council Theme-based Research Scheme fund (T24-508/22-N), the Hong Kong Research Grants Council General Research Fund (15307626) and the Hong Kong Research Grants Council Collaborative Research Fund (C5079-21G).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. He D, Ionides EL, King AA, (2010) Plug-and-play inference for disease dynamics: Measles in large and small populations as a case study. *J R Soc Interface* 7: 271–283. <https://doi.org/10.1098/rsif.2009.0151>
2. King AA, Nguyen D, Ionides EL, (2016) Statistical inference for partially observed Markov processes via the R package pomp. *J Stat Softw* 69: 1–43. <https://doi.org/10.18637/jss.v069.i12>
3. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA, (2015) Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc Natl Acad Sci USA* 112: 719–724. <https://doi.org/10.1073/pnas.1410597112>
4. He D, Dushoff J, Day T, Ma J, Earn DJD, (2013) Inferring the causes of the three waves of the 1918 influenza pandemic in England and Wales. *Proc R Soc B* 280: 20131345. <https://doi.org/10.1098/rspb.2013.1345>
5. Johnson N, (2001) *Aspects of the Historical Geography of the 1918–19 Influenza Pandemic in Britain*, PhD thesis, University of Cambridge. <https://doi.org/10.17863/CAM.27727>
6. Yu D, Lin Q, Chiu APY, He D, (2017) Effects of reactive social distancing on the 1918 influenza pandemic. *PLoS ONE* 12: e0180545. <https://doi.org/10.1371/journal.pone.0180545>
7. Caley P, Philp DJ, McCracken K, (2008) Quantifying social distancing arising from pandemic influenza. *J R Soc Interface* 5: 631–639. <https://doi.org/10.1098/rsif.2007.1197>



© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)