

Bayesian Analysis of Removal Hazards and Dynamic System Behavior in Ebola Outbreak

Zuchen Li

December 15, 2025

1 Introduction

Understanding the rate at which infectious individuals are removed from the transmission pool is critical for studying epidemic dynamics and anticipating system load. In this project, I utilize individual-level data from an Ebola outbreak to develop a sequence of Bayesian models that investigate removal hazards and reconstruct system-level trajectories, aligning with methods.

The analysis proceeds in three stages: First, I employ a Bayesian exponential survival model to estimate baseline removal rates, using data augmentation to infer the latent removal times of right-censored cases. Second, these individual-level posterior samples are used to reconstruct discrete-time Susceptible-Infected-Removed (SIR) trajectories, (Kermack and McKendrick, 1927), thereby quantifying uncertainty in the epidemic curve. Finally, I aggregate these data to the block level to capture dynamic complexities, fitting a time-varying random-walk model and a state-dependent logistic regression model to assess how removal hazards evolve with calendar time and system load. This approach also bridges the gap toward recent literature on state-dependent dynamic systems in epidemiology (Bu et al., 2023; Morsomme and Xu, 2025).

2 Model and Methods

2.1 Individual-level removal model with censoring

The dataset records symptom onset and removal times for patients $i = 1, \dots, N$. Let $T_i = \text{hosp}_i - \text{onset}_i$ be the infectious duration. For censored cases, T_i is latent and known only to exceed a censoring time C_i . I assume an exponential model $T_i \mid \gamma \sim \text{Exp}(\gamma)$. The likelihood contribution L_i distinguishes between observed and censored events:

$$L_i(\gamma, T_i) = \begin{cases} \gamma e^{-\gamma t_i} & \text{if observed at } t_i \\ e^{-\gamma C_i} & \text{if censored at } C_i \end{cases}$$

I assign a weakly informative prior $\gamma \sim \text{Gamma}(0.01, 0.01)$. To handle censoring, I use Bayesian data augmentation (Tanner and Wong, 1987): latent T_i for censored cases are treated as parameters constrained to (C_i, ∞) and sampled jointly with γ via Markov chain Monte Carlo (MCMC) (Metropolis et al., 1953; Hastings, 1970). This yields posterior samples for γ (mean ≈ 0.22) and completes the dataset for the SIR reconstruction.

2.2 Discrete-time SIR reconstruction

Using the posterior draws of $\{T_i\}$, I reconstruct the epidemic trajectory on a daily grid. For each iteration m , the imputed removal time is $\text{hosp}_i^{(m)} =$

$\text{onset}_i + T_i^{(m)}$. I then calculate daily system states by aggregating individuals:

$$I_k^{(m)} = \sum_{i=1}^N \mathbf{1}\{\text{onset}_i \leq t_k < \text{hosp}_i^{(m)}\}, \quad R_k^{(m)} = \sum_{i=1}^N \mathbf{1}\{\text{hosp}_i^{(m)} \leq t_k\}.$$

New infections and removals are tabulated similarly. Aggregating these over all samples yields the posterior mean and credible intervals for the epidemic curve, propagating individual-level censoring uncertainty to the system level.

2.3 Block-level time-varying removal rate γ_k

To analyze temporal dynamics, I aggregate the data into daily blocks where n_k is the "at-risk" count and y_k is the number of removals. Assuming a piecewise-constant rate, I model $y_k \sim \text{Bin}(n_k, p_k)$ where $p_k = 1 - \exp(-\gamma_k)$. To induce temporal smoothness, I place a random-walk prior on the log-scale:

$$\log \gamma_k \mid \log \gamma_{k-1}, \tau_{\text{rw}} \sim \mathcal{N}(\log \gamma_{k-1}, \tau_{\text{rw}}^{-1}), \quad k = 2, \dots, K.$$

The initial state $\log \gamma_1$ and precision τ_{rw} receive weak priors. It borrows strength across time to estimate a smooth trajectory for the removal hazard.

2.4 State-dependent hazard via logistic regression

Finally, I investigate whether removal efficiency depends on system load. Re-using the daily counts, I fit a Bayesian logistic regression to the removal probability p_k :

$$\text{logit}(p_k) = \alpha_0 + \alpha_1 \left(\frac{I_k}{N} \right) + \alpha_2 t_k^*.$$

Here, I_k/N represents the fraction infectious (system load) and t_k^* is a standardized time index to control for secular trends. The coefficients α receive normal priors. This model quantifies the association between epidemic prevalence and removal hazard.

3 Data Analysis

3.1 Data preparation and baseline estimation

The analysis utilizes individual-level records from a single Ebola outbreak wave ($N \approx 1000$). Data cleaning involved removing inconsistent records (e.g., negative durations) and aligning the time origin. Symptom onset is treated as fully observed, while removal times are subject to right-censoring.

The individual-level Bayesian exponential model yielded a posterior mean removal rate of $\gamma \approx 0.217$ per day (95% CI: 0.20, 0.23), corresponding to a mean infectious period of approximately 4.6 days. Convergence diagnostics were excellent, with trace plots showing good mixing, $\hat{R} \approx 1.0$, and effective sample sizes exceeding 50,000. These baseline estimates align with clinical expectations for this outbreak wave.

3.2 SIR reconstruction and uncertainty

Using the posterior samples of latent removal times T_i , we reconstructed the daily system states. The resulting posterior epidemic curve $I(t)$ exhibits a characteristic unimodal shape, rising to a peak before declining. Crucially, the 95% credible bands around $I(t)$ and $R(t)$ are moderately wide, explicitly quan-

tifying how the ambiguity in censored removal times propagates to system-level estimates. This confirms that deterministic SIR models likely underestimate the uncertainty regarding the true peak burden of the epidemic.

3.3 Dynamic and state-dependent hazards

We further analyzed aggregated daily counts (n_k, y_k) to investigate temporal and state-dependent dynamics. The time-varying model with a random-walk prior revealed moderate fluctuations in removal intensity around the baseline rate, with the smoothing prior effectively capturing periods of elevated hazard while filtering out day-to-day stochastic noise.

For the state-dependent logistic regression, MCMC diagnostics indicated that while the regression coefficients exhibited high autocorrelation (lag-1 AC ≈ 0.97) due to covariate correlation, the chains achieved stationarity ($\hat{R} = 1.0$) with sufficient effective sample sizes ($> 1,800$) to support robust inference.

Posterior summaries reveal two distinct mechanisms driving the removal process. First, the coefficient for the infectious fraction is significantly positive ($\alpha_1 = 6.48$, 95% CI: 1.61, 11.45). This implies that the removal probability *increases* substantially as the system load rises, suggesting an adaptive response where isolation efforts were intensified during periods of high prevalence. Concurrently, the time trend coefficient is also positive ($\alpha_2 = 0.22$, 95% CI: 0.13, 0.32), reflecting a gradual improvement in removal efficiency over the calendar period independent of the specific case load.

4 Discussion

The analysis yields three principal insights regarding the removal process.

First, the explicit modeling of censoring prevents the downward bias in infectious period estimation that typically arises from naive complete-case analysis. The baseline model identifies a mean infectious period of approximately 4.6 days. The subsequent SIR reconstruction highlights that uncertainty in these latent removal times propagates significantly to the aggregate epidemic curve. The result offers a more honest basis for risk assessment than deterministic trajectories.

Second, the block-level analysis detects a strong state-dependence in the removal hazard. Contrary to the hypothesis that high case loads might overwhelm the healthcare system (which would result in a negative α_1), we estimate a strongly positive coefficient for the infectious fraction ($\alpha_1 = 6.48$). This suggests an adaptive public health response: as the prevalence of infection increased, the speed of isolation and hospitalization accelerated. This phenomenon likely reflects the mobilization of emergency resources, intensified contact tracing, and heightened community awareness triggered by the visible surge in cases.

Finally, the positive secular trend ($\alpha_2 = 0.22$) indicates that removal efficiency improved over time even after controlling for system load. This points to a learning effect or procedural optimization in outbreak management independent of the immediate crisis severity. Together, these results demonstrate that epidemic parameters are not static constants but dynamic quantities responsive to both system state and operational experience.

References

- Morsomme, R. and Xu, J. (2025). Exact Bayesian inference for fitting stochastic epidemic models to partially observed incidence data. *The Annals of Applied Statistics*, **19**(3), 2279–2293.
- Bu, F., Aiello, A. E., Xu, J., and Volfovsky, A. (2023). Likelihood-based inference for partially observed epidemics on dynamic networks. *Journal of the American Statistical Association*, **118**(541), 102–116.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**(6), 1087–1092.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**(398), 528–540.
- Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, **115**(772), 700–721.

5 Tables & Figures

Table 1: Posterior summaries for the logistic regression result of the state-dependent hazard model coefficients.

Parameter	Mean	Median	2.5%	97.5%
α_0 (Intercept)	-1.84	-1.84	-2.12	-1.57
α_1 (Fraction Infectious I_k/N)	6.48	6.45	1.61	11.45
α_2 (Time Trend t^*)	0.22	0.22	0.13	0.32

Table 2: MCMC convergence diagnostics. $\hat{R}$ values close to 1.0 indicate convergence. While Model 3 parameters show high autocorrelation (AC) at Lag 1, the Effective Sample Sizes (ESS) remain sufficient for inference.

Model	Parameter	$\hat{R}$	ESS	AC (Lag 1)	AC (Lag 10)
Censored Exp	γ	1.000	53,678	0.301	-0.003
RW Gamma	μ_γ	1.000	131,545	0.025	0.005
	τ_γ	1.001	77,996	0.244	0.002
	τ_{RW}	1.000	5,636	0.864	0.457
State Hazard	α_0	1.000	1,852	0.969	0.776
	α_1	1.000	1,823	0.972	0.780
	α_2	1.000	4,826	0.586	0.310

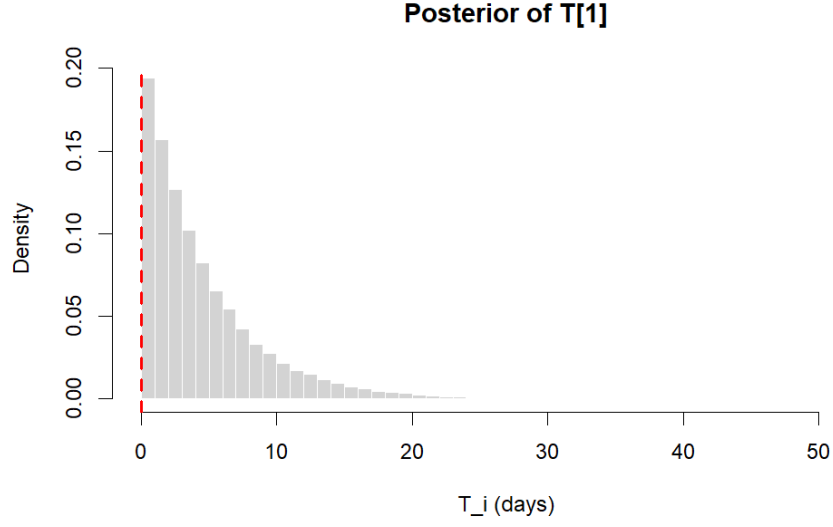


Figure 1: Posterior distribution of the latent removal time T_i for individual

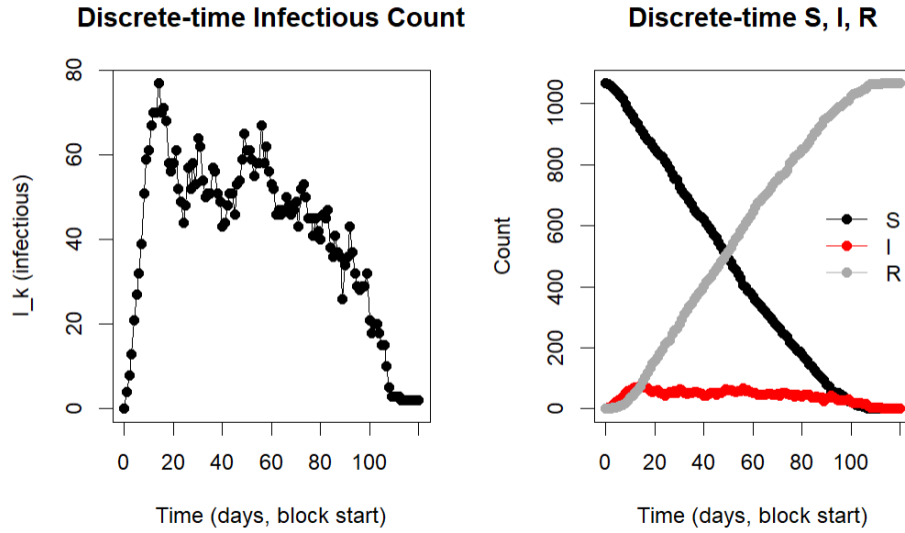


Figure 2: Reconstructed discrete daily infectious count and SIR trajectories.

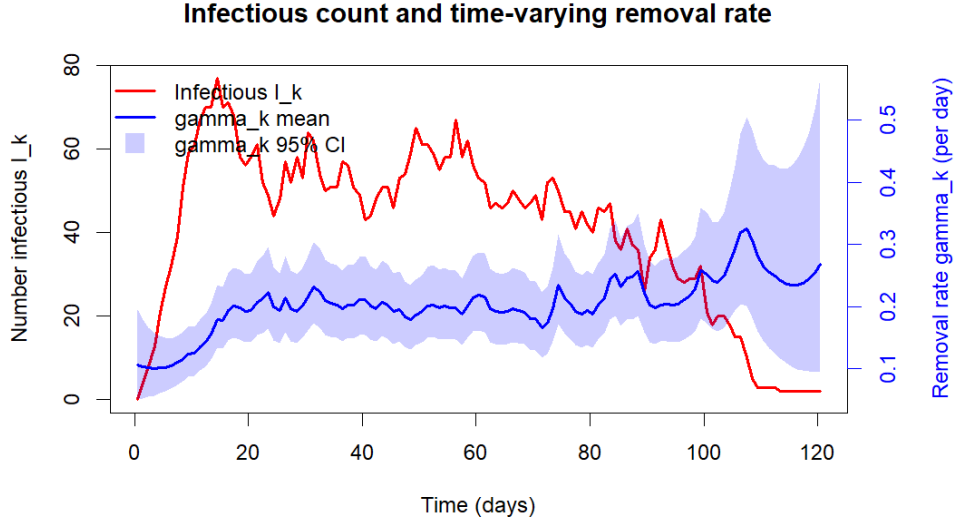


Figure 3: The red line shows the number of actively infectious individuals I_k each day. The blue line shows the posterior mean of the time-varying removal rate γ_k . The shaded blue region is the 95% posterior credible interval for γ_k .

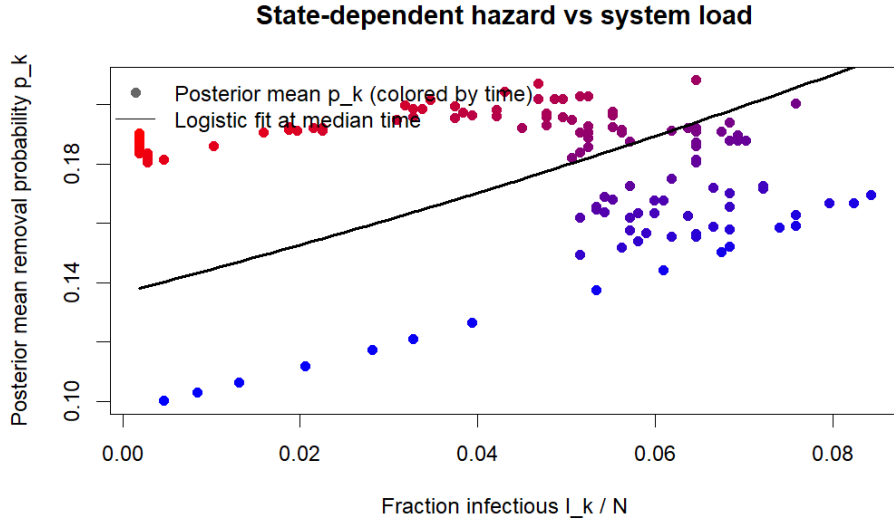


Figure 4: Points display the posterior mean removal probability p_k versus the infectious fraction I_k/N colored by time, overlaid with the fitted logistic regression curve holding the time covariate fixed.