

Article

Bayesian Disease Mapping Beyond Aggregated Counts: A BYM2 Framework for Individual-Level Inference

Chengwei Zhang ¹, Hui Yan ², Zihao Jia ¹ and Danchen Aaron Yang ^{2,*} 

¹ College of Science, Nanjing Agricultural University, Nanjing 210095, China; zhangcw@stu.njau.edu.cn (C.Z.); leojia10@stu.njau.edu.cn (Z.J.)

² College of Veterinary Medicine, Nanjing Agricultural University, Nanjing 210095, China; yanhui@stu.njau.edu.cn

* Correspondence: d.a.yang@njau.edu.cn

Abstract

Bayesian disease mapping models are formulated for aggregated areal counts, although veterinary epidemiological studies commonly collect individual-level data from animals located in geographical units. Aggregating such data to the areal level may change the estimand and make individual-level interpretation inappropriate. This study extends the original Besag–York–Mollié 2 (BYM2) model to a multilevel modeling framework for individual-level binary outcomes by embedding a BYM2 random effect within a logistic regression model. The proposed model retains individual animals as the unit of analysis, estimates covariate associations, and accounts for residual structured and unstructured areal variation. We present the work as a practical guide, covering likelihood and priors specifications, BYM2 scaling, estimation algorithm, model diagnostics, posterior predictive checks, practical model comparison, sensitivity analysis, and interpretation of fixed and areal effects. The model is directly applicable to individual-level binary outcomes in a cross-sectional study or a cohort study with a fixed follow-up period, while we also demonstrated that the proposed model can estimate associations generated from an open cohort time-to-event process under appropriate sampling frameworks including probability proportional to size sampling and incidence density sampling. This paper provides a reproducible workflow for applying BYM2 disease mapping to individual-level epidemiological data and clarifies the study designs under which the proposed binary modeling framework can be appropriately used.

Keywords: Bayesian; spatial modeling; Stan; veterinary epidemiology; non-aggregation

MSC: 62P10



Academic Editor: Davide Valenti

Received: 24 June 2026

Revised: 14 July 2026

Accepted: 16 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Spatial variation is common in veterinary epidemiological studies. Disease occurrence in animals, farms, or other study subjects is often influenced not only by measured individual-level or subject-level exposures but also by unmeasured regional factors, such as environmental conditions, veterinary service accessibility, and local disease ecology [1,2]. When such factors are spatially patterned, study subjects located in neighboring areas may have more similar outcome risks than those located farther apart. Ignoring this residual spatial dependence may lead to incomplete adjustment for areal heterogeneity [3] and may obscure regional patterns that are relevant for disease control and surveillance. Bayesian disease mapping models provide a useful framework for representing this type of areal

heterogeneity [4]. In traditional areal disease mapping, the outcome variable is commonly expressed as the number of cases within each spatial unit, and spatial random effects are used to smooth regional estimates and describe residual spatial structure [5]. Among these approaches, the Besag–York–Mollié (BYM) model [6] and its reparameterized form, the BYM2 model [7], are widely used in veterinary epidemiology to decompose areal variation into structured and unstructured components [8,9]. The BYM2 formulation improves the interpretability and computational stability of the original BYM model by scaling the structured spatial effect and explicitly modeling the proportion of spatially structured variation [10]. This approach is appropriate when the research question focuses on areal-level risk or associations [11–13].

However, many veterinary epidemiological studies are not conducted with aggregated regional counts. Conducted studies, including routinely collected animal-health databases, often contain individual-level information on outcome status and exposures. When the objective is to estimate associations between exposures and disease occurrence at the animal level, aggregation to the areal level is likely to create ecological fallacy [14,15]. A more direct strategy is therefore to retain the study subject as the unit of analysis and to introduce the areal spatial structure through a random effect component. For binary outcomes, logistic regression is a familiar tool in veterinary epidemiology. Their natural applications include cross-sectional studies, where the outcome represents disease or infection status at the time of sampling, and risk-based closed cohort studies, where all study subjects are followed over a defined risk period and the outcome records whether the event occurred during that period. In these settings, a natural extension of ordinary logistic regression is to allow the model to accommodate the multilevel spatial structure of the data by embedding a BYM2 areal random effect. The fixed effects describe associations between measured exposures and the binary outcome, while the BYM2 component describes residual spatial heterogeneity among the areas to which study subjects belong. The use of such a model is, however, less straightforward when the source population is dynamic. In an open cohort, study subjects may enter and leave the population at different times, and their time at risk may differ substantially. In this setting, disease frequency is more naturally described by rates rather than risks, and survival or rate-based methods are usually the primary analytical tools. Nevertheless, if the aim is to estimate relative exposure effects, a case–control design based on appropriate sampling methods can provide an efficient analytical representation of the underlying rate-based process. Under the usual assumptions of valid control selection and stable exposure definition, the odds ratio from the resulting case–control data can be interpreted as a relative measure of the event rate rather than as an absolute disease risk [16].

Rather than introducing a new spatial model, this paper reformulates the original BYM2 disease mapping framework for modeling individual-level binary outcomes nested within areal units. The objectives are therefore twofold. First, we describe the Bayesian hierarchical specification of the individual-level Bernoulli logistic BYM2 model and provide the posterior estimation procedure using the No-U-Turn Sampler implemented in Stan [17,18]. Second, we present a practical data analysis workflow to demonstrate the required data structure, model implementation, posterior diagnostics, posterior predictive checking, model comparison, sensitivity analysis and interpretation of results.

2. Materials and Methods

2.1. The Model

Suppose there are N individuals and C regions. For the i -th individual, let Y_i denote the binary event outcome, where $Y_i = 1$ indicates that the individual experiences the target

event and p_i denote the probability that the i -th individual experiences the target event. The individual-level Bernoulli likelihood is written as

$$Y_i \sim \text{Bernoulli}(p_i), \quad i = 1, \dots, N. \quad (1)$$

The model incorporates areal random effects into the linear predictor of the individual-level logistic regression model:

$$\text{logit}(p_i) = \eta_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + b_{c(i)}, \quad (2)$$

where, α is the intercept term, representing the baseline log-odds when the covariates are at their reference levels. $\mathbf{x}_i$ is the covariate vector for the i -th individual, which may include binary variables, continuous variables, and/or categorical variables encoded by dummy variables. $\boldsymbol{\beta}$ is the corresponding vector of regression coefficients, used to characterize the association between the covariates and the risk of event occurrence. The subscript $c(i)$ denotes the region to which the i -th individual belongs, and $b_{c(i)}$ denotes the corresponding areal random effect.

To simultaneously characterize structured spatial correlation and unstructured areal heterogeneity, this model adopts the BYM2 reparameterization to construct the areal random effect. For region $c = 1, \dots, C$, the composite areal random effect is defined as follows [19]:

$$b_c = \sigma \left(\sqrt{1 - \rho} \theta_c + \sqrt{\rho/s} \phi_c \right) \quad (3)$$

where θ_c denotes unstructured areal heterogeneity, which describes independent regional variation that does not depend on spatial adjacency. The scaling factor s is determined by the areal adjacency structure. It is used to scale the intrinsic conditional autoregressive (ICAR) structured spatial effect ϕ_c so that the structured spatial component and the unstructured component θ_c are placed on comparable variance scales. The computation of s will be illustrated in later text. The standard deviation (SD) σ represents the overall strength of areal heterogeneity, namely the overall magnitude of residual areal-level variation after adjusting for covariates. The parameter $\rho \in [0, 1]$ represents the proportion of the structured spatial component in the composite areal random effect.

For C regions, the spatial effect is a vector of length C , such that $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_C)^\top$. Under the Gaussian Markov random field framework, if the precision matrix $\mathbf{Q}$ is symmetric positive definite, then the spatial effect vector $\boldsymbol{\phi}$ can be expressed as a multivariate normal distribution with $\mathbf{Q}$ as the precision matrix, that is, $\boldsymbol{\phi} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$ [20]. In the ICAR model, the precision matrix is constructed from the spatial adjacency structure of the areal units. Specifically, it is defined as

$$\mathbf{Q} = \tau(\mathbf{D} - \mathbf{A}), \quad (4)$$

where τ is the precision parameter, $\mathbf{D} = \text{diag}(n_1, \dots, n_C)$ is the degree matrix, and n_c denotes the number of regions neighboring region c . The matrix $\mathbf{A} = (A_{cd})$ is a symmetric binary adjacency matrix, where $A_{cd} = 1$ if regions c and d are neighbors and $A_{cd} = 0$ otherwise, with $A_{cc} = 0$ for all $c = 1, \dots, C$. Under this specification, each row of $\mathbf{Q}$ sums to zero. Therefore, a constant vector lies in the null space of $\mathbf{Q}$, meaning that $\mathbf{Q}$ is rank-deficient and has no ordinary inverse. As a result, the ICAR model does not define a proper multivariate normal distribution with covariance matrix $\mathbf{Q}^{-1}$ in the usual sense [21]. Instead, it is used as an intrinsic Gaussian Markov random field prior for the spatial random effect.

The implication of this singularity can be seen more clearly from the quadratic form associated with the ICAR prior. Because $\mathbf{Q} = \tau(\mathbf{D} - \mathbf{A})$ is constructed from the graph Laplacian, the corresponding quadratic form can be written as

$$\boldsymbol{\phi}^\top \mathbf{Q} \boldsymbol{\phi} = \tau \sum_{\substack{c < d \\ A_{cd}=1}} (\phi_c - \phi_d)^2. \quad (5)$$

This expression shows that the ICAR prior penalizes pairwise differences between the spatial effects of neighboring regions. However, the penalty involves only differences such as $\phi_c - \phi_d$ and does not contain any term that fixes the absolute level of $\boldsymbol{\phi}$ as a whole. Therefore, if the same constant is added to all regional spatial effects, all neighboring differences remain unchanged, and the ICAR penalty is also unchanged. In practical Bayesian spatial modeling, this issue is commonly addressed by imposing a sum-to-zero constraint,

$$\sum_{c=1}^C \phi_c = 0. \quad (6)$$

Under this constraint, the intercept represents the overall level of the outcome, while the ICAR component represents relative spatial deviations among regions [19].

The preceding model specification defines the Bernoulli likelihood, the fixed effect structure, and the BYM2 areal random effect component. To complete the Bayesian hierarchical model and make posterior inference possible, prior distributions must be assigned to the unknown parameters before posterior estimation. The prior specifications for the parameters are listed:

$$\begin{aligned} \alpha &\sim N(0, 1), \\ \beta_k \mid \lambda &\sim \text{DoubleExp}(0, 1/\lambda), \quad k = 1, \dots, P, \\ \lambda &\sim \text{Gamma}(3, 1), \\ \sigma &\sim N^+(0, 1), \\ \zeta_\rho &\sim N(0, 1), \quad \rho = \text{logit}^{-1}(\zeta_\rho), \\ \theta_c &\sim N(0, 1), \quad c = 1, \dots, C, \\ \boldsymbol{\phi} \mid \mathbf{A} &\sim \text{ICAR}(\mathbf{A}), \quad \text{with } \mathbf{1}^\top \boldsymbol{\phi} \sim N\left(0, (0.001C)^2\right). \end{aligned} \quad (7)$$

where the double exponential priors provide a Bayesian Lasso formulation for the regression coefficients; under this parameterization, larger values of λ correspond to stronger shrinkage toward zero [22,23].

2.2. Parameter Estimation

Under the model and prior specifications described above, the unknown parameters to be estimated can be denoted as $\Theta = (\alpha, \boldsymbol{\beta}, \lambda, \boldsymbol{\theta}, \boldsymbol{\phi}, \sigma, \zeta_\rho)$. This study uses the No-U-Turn Sampler (NUTS) in Stan to sample from the posterior distribution [18]. NUTS is an adaptive Hamiltonian Monte Carlo (HMC) method that uses gradient information from the log posterior density to construct efficient sampling trajectories [17]. Before posterior sampling, the BYM2 scaling factor s is calculated from the diagonal elements of the generalized inverse of the $\mathbf{Q}_* = \mathbf{D} - \mathbf{A}$ matrix and is given by

$$s = \exp\left\{\frac{1}{C} \sum_{c=1}^C \log[(\mathbf{Q}_*^-)_{cc}]\right\}, \quad (8)$$

where $\mathbf{Q}_*^-$ denotes the generalized inverse of $\mathbf{Q}_*$. The scaling factor s corresponds to the geometric mean of the marginal variances implied by the ICAR structure. Because s

depends only on the spatial adjacency matrix, it is treated as a fixed quantity throughout posterior sampling. In the BYM2 formulation used here, the area-level random effect for region c is expressed as a scaled mixture of an unstructured component and a structured spatial component, therefore s enters the likelihood through the linear predictor η_i rather than appearing as a separate prior term in the log posterior density.

Combining the Bernoulli logistic likelihood with the prior distributions specified above, the log posterior density can be written, up to an additive constant C_0 , as

$$\begin{aligned} \log \pi(\Theta \mid \mathbf{y}) = & \sum_{i=1}^N [y_i \eta_i - \log\{1 + \exp(\eta_i)\}] - \frac{1}{2} \alpha^2 \\ & + \sum_{k=1}^P (\log \lambda - \lambda |\beta_k|) + 2 \log \lambda - \lambda \\ & - \frac{1}{2} \theta^\top \theta - \frac{1}{2} \phi^\top Q_* \phi - \frac{1}{2} \sigma^2 - \frac{1}{2} \zeta_\rho^2 - \frac{(\mathbf{1}^\top \phi)^2}{2(0.001C)^2} + C_0. \end{aligned} \quad (9)$$

The constrained parameters are then mapped to an unconstrained space before NUTS is applied. In this model, positive parameters such as λ and σ are transformed to the real line using logarithmic transformations. The structured spatial effect ϕ is represented as a C -dimensional parameter vector. To address the intrinsic non-identifiability of the ICAR prior, a soft sum-to-zero constraint is imposed by assigning a narrow normal prior to the sum of the spatial effects. This formulation encourages the structured spatial effects to be centered around zero without imposing an exact deterministic constraint. Other unconstrained parameters, including ζ_ρ , α , β , and θ , are sampled directly on the real scale.

Let $\mathbf{q}$ denote the vector of unconstrained parameters. The target density on the unconstrained scale is then given by

$$\log \pi_q(\mathbf{q} \mid \mathbf{y}) = \log \pi_\Theta(T^{-1}(\mathbf{q}) \mid \mathbf{y}) + \log \left| \det \frac{\partial T^{-1}(\mathbf{q})}{\partial \mathbf{q}} \right|. \quad (10)$$

where $T^{-1}(\mathbf{q})$ maps the unconstrained parameters back to their original constrained space. The second term is the Jacobian adjustment associated with this transformation and is handled automatically by Stan. NUTS uses the same Hamiltonian framework as HMC. An auxiliary momentum vector $\mathbf{m}$ is introduced in the unconstrained parameter space, and the Hamiltonian is defined as

$$H(\mathbf{q}, \mathbf{m}) = -\log \pi_q(\mathbf{q} \mid \mathbf{y}) + \frac{1}{2} \mathbf{m}^\top \mathbf{M}^{-1} \mathbf{m}, \quad (11)$$

where $\mathbf{M}$ is the mass matrix. In each iteration, the momentum is sampled from $\mathbf{m} \sim N(\mathbf{0}, \mathbf{M})$ and the parameter position and momentum are updated using leapfrog integration given a step size ϵ :

$$\mathbf{m}^{(r+1/2)} = \mathbf{m}^{(r)} + \frac{\epsilon}{2} \nabla_{\mathbf{q}} \log \pi_q(\mathbf{q}^{(r)} \mid \mathbf{y}), \quad (12)$$

$$\mathbf{q}^{(r+1)} = \mathbf{q}^{(r)} + \epsilon \mathbf{M}^{-1} \mathbf{m}^{(r+1/2)}, \quad (13)$$

$$\mathbf{m}^{(r+1)} = \mathbf{m}^{(r+1/2)} + \frac{\epsilon}{2} \nabla_{\mathbf{q}} \log \pi_q(\mathbf{q}^{(r+1)} \mid \mathbf{y}). \quad (14)$$

Here, $\nabla_{\mathbf{q}} \log \pi_q(\mathbf{q} \mid \mathbf{y})$ is the gradient of the log posterior density on the unconstrained scale, which Stan evaluates using automatic differentiation. In standard HMC, the number of leapfrog steps must be specified in advance. In contrast, NUTS dynamically builds the Hamiltonian trajectory and stops its expansion when a U-turn is detected. During

warm-up, Stan adapts the step size ϵ and the mass matrix $\mathbf{M}$, thereby reducing the need for manual tuning of the sampler. After the warm-up stage, Stan fixes the step size and mass matrix and enters the formal sampling stage. Suppose the formal sampling stage contains L independent chains. For each chain, S HMC iterations are performed to obtain posterior samples of the unconstrained parameter vector $\mathbf{q}^{(s,l)} : s = 1, \dots, S, l = 1, \dots, L$, which are then transformed back to the original parameter scale according to $\Theta = T^{-1}(\mathbf{q})$. Finally, the posterior samples are used to obtain point and interval estimates for each parameter.

2.3. Model Diagnostics

Sampling convergence and numerical stability of NUTS need to be assessed to evaluate whether the posterior distributions are adequately explored. These diagnostics include trace plots, $\hat{R}$, effective sample size (ESS), divergent transitions, maximum tree-depth warnings, and the estimated Bayesian fraction of missing information (E-BFMI) [24]. Because these sampling diagnostics assess the numerical behavior of the sampler rather than the adequacy of the fitted model, posterior predictive checks are also conducted [19,25]. For each retained posterior draw, replicated outcomes are generated from the posterior predictive distribution under the fitted model. The observed data are then compared with the replicated datasets using several summaries relevant to the study design. First, the total number of observed cases is compared with the posterior predictive distribution of the total number of replicated cases, and a posterior predictive p -value is calculated as the proportion of replicated datasets in which the total number of replicated cases was greater than or equal to the observed total. Second, observed and posterior predictive mean case counts are compared at the areal level to assess whether the model reproduced geographical variation in case counts. Third, observations or strata with similar posterior predicted probabilities are grouped, and the observed case proportions are compared with the corresponding mean posterior predicted probabilities to assess calibration. Where appropriate, covariate-stratified posterior predictive checks are further used to examine whether the fitted model captured important covariate patterns. This part will be further demonstrated in the illustrative example.

2.4. Illustrative Example

For cross-sectional studies or closed-population cohort studies with a fixed risk period, application of the proposed Bernoulli logistic BYM2 model is straightforward. This illustrative example was therefore designed to show how an open cohort can be converted into case-control analytical datasets under two sampling strategies and how the proposed individual-level Bernoulli logistic BYM2 model can then be applied to the resulting data. The purpose was to guide readers through the analysis and interpretation steps that would arise in a real-world application, rather than to make epidemiological claims about a specific population or to present a simulation benchmark demonstrating universal superiority of the proposed model over competing approaches.

Figure 1 summarizes the construction of the illustrative analytical datasets and the subsequent model-fitting workflow. The source population was constructed as an open cohort with a time-to-event outcome. Each individual had an areal identifier linked to one of $C = 77$ cities of East China, individual-level covariates, an entry time, an event time, and an observed follow-up time. The city-level adjacency structure required by the BYM2 model was obtained using the Queen contiguity criterion, under which two cities were defined as neighbors if they shared a boundary or a vertex. This adjacency graph was used to generate and model the spatially structured component of the areal random effects. For the source population, individual event times were generated from a Weibull proportional hazards model. The linear predictor included two individual-level covariates

and a composite city-level random effect generated from a BYM2-type structure. The binary covariate x_{1i} was generated from a Bernoulli distribution with probability 0.3 of taking the value 1, and the continuous covariate x_{2i} was generated from a standard normal distribution. Individuals were assigned to the 77 city-level areas according to an empirical city-level population size distribution derived from a veterinary insurance dataset [26], so that areas with larger populations contributed more individuals to the simulated source population. Delayed entry was introduced by allowing individuals to enter the cohort at different calendar times, and administrative censoring was imposed at the end of the study period. The true parameter values and data-generating distributions used to generate the open cohort source population are summarized in Table 1. More detail of data generation can be found at <https://github.com/aaronpopolar/multilevelBYM2> (accessed on 14 July 2026). Thus, the source population retained the main features of an open cohort, including staggered entry, unequal follow-up time, event indicators, individual-level covariates, and areal membership. The open cohort dataset was not used directly as the input for the proposed model. Instead, it served as the source population from which two case–control analytical datasets were constructed.

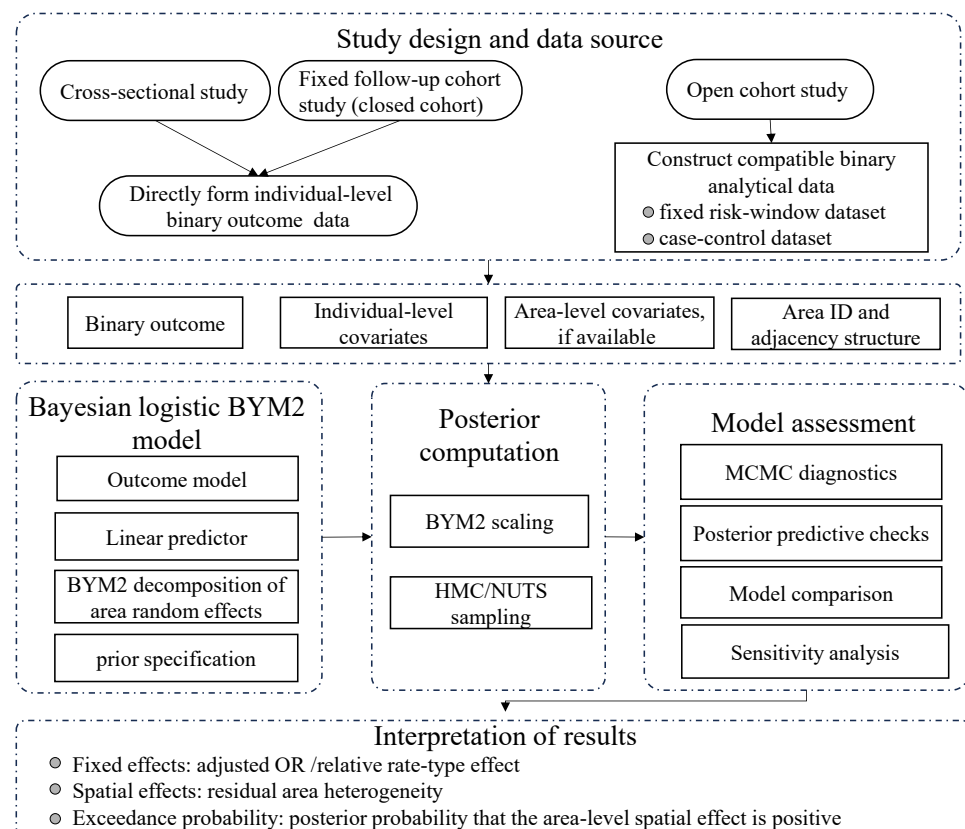


Figure 1. Schematic workflow of the Bernoulli logistic BYM2 modeling framework.

Two case–control analytical datasets were constructed from the open cohort, one using probability proportional to size sampling (PPS) [27] and the other using incidence density sampling (IDS) [16]. In both designs, all individuals who experienced the event in the source population were included as cases, and the number of controls was fixed by design with the case-to-control ratio 1:2. Therefore, proportion of cases in the resulting analytical datasets did not estimate the event incidence in the source population, and the intercept from the fitted model is influenced by the sampling scheme and should not be interpreted as the population baseline disease probability. The main quantities of interest are the regression coefficients and the residual areal random effects.

Table 1. True parameter values used to simulate the open cohort source population.

Parameter	Description	True Value
N_0	Source population size	30,000
C	Number of city-level areas	77
ρ	Spatial mixing parameter	0.7
σ	Overall standard deviation of BYM2 random effect	1.5
β_1	Fixed effect of the binary covariate x_{1i}	1.0
β_2	Fixed effect of the continuous covariate x_{2i}	0.5
<i>shape</i>	Weibull shape parameter	1.2
<i>scale</i> ₀	Baseline Weibull scale parameter	5000

The intercept α was estimated in the fitted Bernoulli logistic BYM2 model, but it was not assigned a true value in the open cohort data generating mechanism because event times were generated using a Weibull proportional hazards model.

Under the PPS design, the aim was to construct a case–control dataset while allowing individuals with longer observed follow-up to contribute more heavily to the control sample. The measure of size was defined as the observed follow-up time t_i , representing the amount of time contributed by individual i to the source population. Controls were sampled with replacement from all individuals who contributed positive observed follow-up time, regardless of their final event status. Thus, an individual who experienced the event could also be sampled one or more times as a control record. Under the IDS design, the aim was to construct a case–control dataset while preserving the risk-set structure of the event process. For each case, a separate risk set was constructed at the calendar time of event occurrence. For the i -th case, the risk set consisted of all individuals who had already entered the cohort and were still under observation at the time when case i experienced the event. The case individual was excluded from their own control set, and two controls were sampled from this risk set. Because controls were selected according to their risk status at each case event time rather than their final event status, an individual selected as a control at an earlier event time could subsequently become a case, and the same individual could also be selected as a control for multiple cases [28]. The proposed model was fitted separately to the two case–control analytical datasets. Since the sampled control records represent the distribution of individual-time at risk in the source population with stable covariates (measured at the beginning of the study and not time-varying), the exponentiated coefficient, $\exp(\beta_k)$, can be interpreted as a relative rate-type effect [16]. The areal random effects describe residual spatial differences after adjustment for covariates.

As the final part of the illustrative workflow, we fitted two additional individual-level regional models for comparison with the proposed BYM2 model. The first was a standard Bayesian logistic mixed effects model with independent regional random intercepts, representing purely unstructured regional heterogeneity. The second was an ICAR spatial logistic model, representing purely spatially structured regional heterogeneity. These two models correspond to the limiting cases of the BYM2 formulation, with $\rho = 0$ and $\rho = 1$, respectively. The three models were compared using posterior summaries and approximate leave-one-out cross-validation with Pareto-smoothed importance sampling (PSIS-LOO), including LOOIC and pointwise Pareto- k diagnostics [29,30]. To assess whether the main posterior inferences were sensitive to the prior specification of the BYM2 hyperparameters, we conducted a prior sensitivity analysis for σ and ρ . In the sensitivity analysis, we refitted the model under alternative priors for σ , including a more concentrated prior, $\sigma \sim N^+(0, 0.5)$, and a more diffuse prior, $\sigma \sim N^+(0, 2)$. We also considered a more diffuse prior for the mixing parameter, $\text{logit}(\rho) \sim N(0, 2)$. The resulting posterior summaries were compared with those from the model with the main priors, focusing on the interpretation of the covariate effects β , σ , and ρ . All analyses were performed in R

version 4.5.1 using the `cmdstanr` package, which interfaced with CmdStan version 2.37.0 for posterior computation.

3. Results

3.1. Description of the Analytical Datasets

The synthetic source population contained 3837 individuals who experienced the event, accounting for 12.79%. Both PPS and IDS generated 11,511 case–control records, including all cases and 7674 controls. The PPS dataset contained 10,108 unique individuals, among which the control records corresponded to 6762 unique individuals. The IDS dataset contained 10,105 unique individuals, among which the control records corresponded to 6720 unique individuals. Since both sampling methods allowed individuals to enter the control set repeatedly, 850 control individuals were repeatedly selected in PPS sampling, while 857 control individuals were repeatedly selected in IDS. In both datasets, a single control individual was selected at most 4 times. In addition, 491 individuals in the PPS dataset contributed records as both controls and cases, while the corresponding number in the IDS dataset was 452.

3.2. Model Output

Both models were sampled using 4 chains, with 1000 warm-up iterations and 2500 posterior sampling iterations per chain, yielding 10,000 retained posterior samples for each analysis. For both models, there were 0 divergent transitions, 0 max treedepth hits, and the minimum E-BFMI across chains were 0.781 for PPS and 0.778 for IDS. For both analyses, the $\hat{R}$ values for the main parameters were (close to) 1, and the effective sample sizes were adequate. The trace plots for ρ , σ , λ , α , β_1 , and β_2 showed good mixing, with no obvious between-chain separation or long-term drifting (see Appendix A). These diagnostics suggest that posterior computation was stable for both sampled datasets.

The two case–control sampling designs produced similar fitted results in this illustrative example. The posterior results of the main parameters based on the PPS dataset are shown in Table 2, while those based on the IDS dataset are shown in Table 3. For the fixed effects, the PPS estimates were $\hat{\beta}_1 = 1.060$ and $\hat{\beta}_2 = 0.512$, while the IDS estimates were $\hat{\beta}_1 = 0.989$ and $\hat{\beta}_2 = 0.499$. In both analyses, the estimated effect of x_1 was larger than that of x_2 , which was consistent with the intended covariate pattern in the illustrative dataset. The BYM2 hyperparameters were also similar across the two analyses. The posterior means of σ were 1.590 under PPS sampling and 1.670 under IDS, while the posterior means of ρ were 0.675 and 0.660, respectively. These results indicated that the fitted model behaved as expected in this illustrative workflow.

Table 2. Posterior summaries of the main parameters from the Bernoulli spatial logistic BYM2 model fitted to the PPS dataset.

Parameter	Mean	SD	95% CrI	$\hat{R}$	Bulk ESS	Tail ESS
ρ	0.6750	0.1460	[0.3520, 0.9090]	1.0000	1799	3060
σ	1.5900	0.1650	[1.3000, 1.9500]	1.0000	3383	5813
α	−2.0100	0.1270	[−2.2600, −1.7600]	1.0000	8125	7137
β_1	1.0600	0.0500	[0.9620, 1.1600]	1.0000	22,156	6893
β_2	0.5120	0.0248	[0.4630, 0.5610]	1.0000	21,195	6474
λ	1.9500	0.8810	[0.6290, 4.0400]	1.0000	19,935	6666

Mean and SD denote the posterior mean and posterior standard deviation, respectively. CrI denotes the Bayesian credible interval. $\hat{R}$ is the convergence diagnostic, and bulk ESS and tail ESS denote the effective sample sizes for the central part and tails of the posterior distribution, respectively.

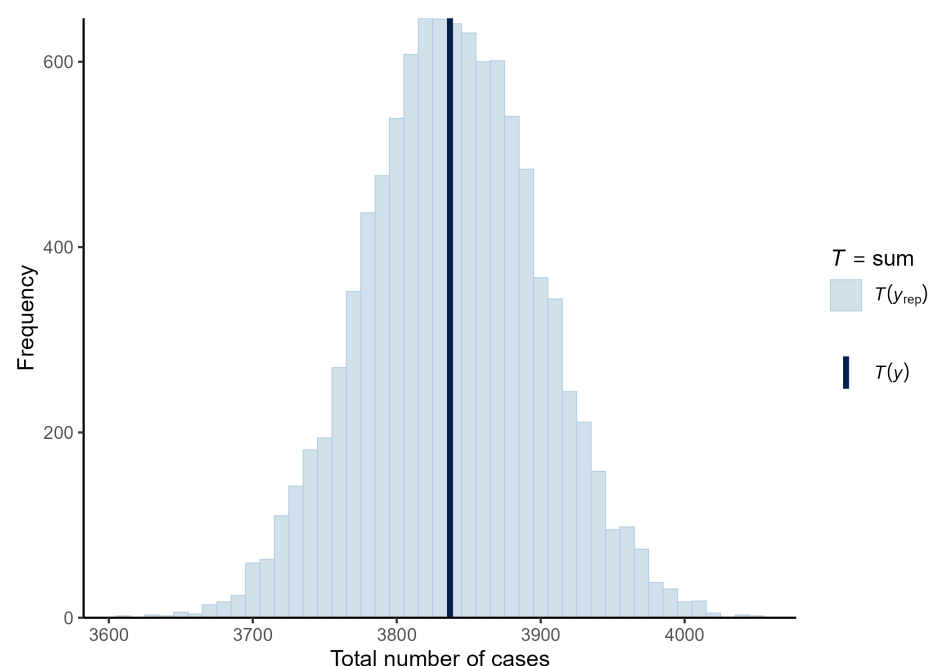
Table 3. Posterior summaries of the main parameters from the Bernoulli spatial logistic BYM2 model fitted to the IDS dataset.

Parameter	Mean	SD	95% CrI	$\hat{R}$	Bulk ESS	Tail ESS
ρ	0.6600	0.1440	[0.3460, 0.9010]	1.0000	1672	3344
σ	1.6700	0.1760	[1.3600, 2.0500]	1.0000	3436	5495
α	−1.9500	0.1340	[−2.2200, −1.6900]	1.0000	6970	6704
β_1	0.9890	0.0491	[0.8930, 1.0800]	1.0000	21,540	6902
β_2	0.4990	0.0246	[0.4510, 0.5480]	1.0000	23,730	7006
λ	2.0000	0.8960	[0.6420, 4.0800]	1.0000	18,835	6715

3.3. Posterior Predictive Checks

Posterior predictive checks were used to examine whether the fitted model could reproduce key features of the constructed analytical dataset. Consistent with the diagnostic strategy described in the Materials and Methods section, the checks compared the observed incidence density sampling dataset with replicated datasets generated from the posterior predictive distribution. For illustrative purposes, these checks were applied to the IDS dataset.

First, the total number of cases was compared with its posterior predictive distribution. If the model reproduces the overall case frequency in the sampled dataset, the observed total falls within the main body of the replicated distribution rather than in its extreme tails. In Figure 2, the observed total number of cases per 10,000 observations is located near the central region of the posterior predictive distribution. This indicates that the fitted model reproduces the overall number of cases in the IDS dataset.

**Figure 2.** Posterior predictive check for the total number of cases in the IDS dataset. The histogram shows the posterior predictive distribution of $T(y_{\text{rep}}) = \sum_i y_{\text{rep},i}$, and the vertical line marks the observed statistic $T(y) = \sum_i y_i$.

Second, posterior predictive performance was assessed at the city level by comparing the observed number of cases with the posterior predictive mean number of cases for each city. In this plot, good agreement is indicated by points lying close to the 45-degree reference line. Figure 3 shows that most cities are distributed near this line, suggesting that the fitted model reproduces the main pattern of geographical variation in sampled case counts.

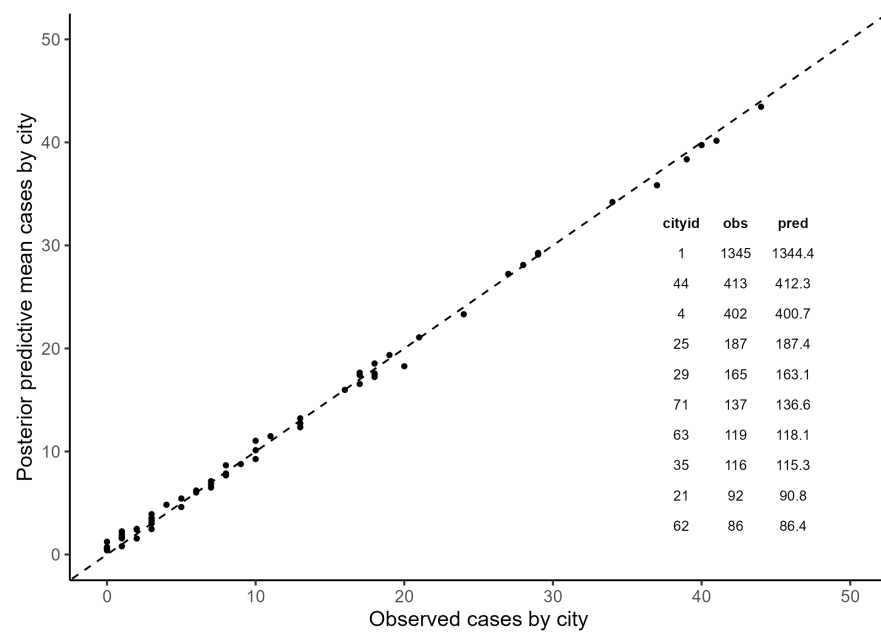


Figure 3. City-level posterior predictive check for case counts in the IDS dataset. Each point represents one city with no more than 50 observed cases. The horizontal axis shows the observed case count, the vertical axis shows the posterior predictive mean case count, and the dashed line denotes the 45-degree reference line. Cities with more than 50 observed cases are summarized in the inset table.

Third, a binned calibration check was used to compare observed case proportions with posterior predicted probabilities. Observations were grouped according to similar posterior predicted probabilities, and the observed case proportion within each group was compared with the corresponding mean posterior predicted probability. If the model is well calibrated within the sampled dataset, the observed proportions are close to the predicted probabilities across bins. Figure 4 shows that the observed case proportions are close to the posterior predicted probabilities across the probability groups, suggesting that the fitted model reproduces the overall probability pattern in the IDS dataset.

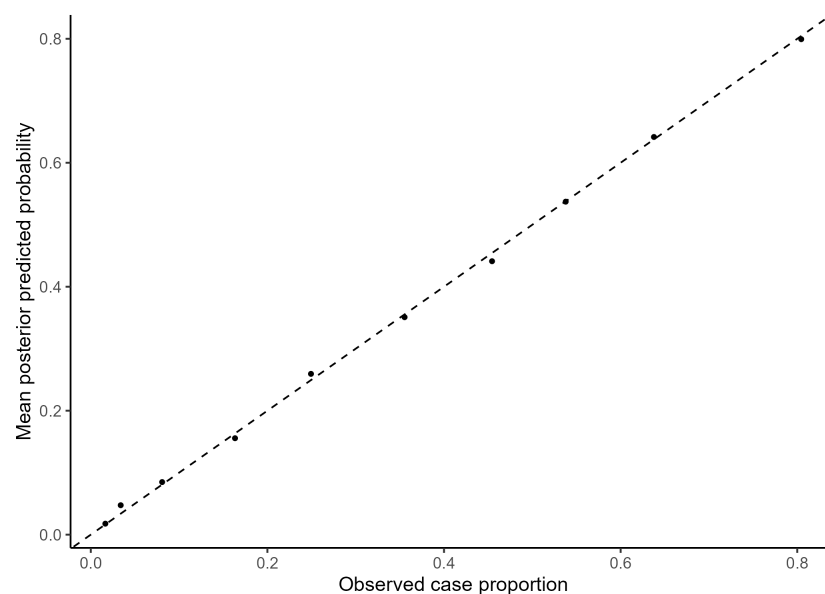


Figure 4. Binned calibration of posterior predicted probabilities in the IDS dataset. Observations are grouped into bins according to similar posterior predicted probabilities. Each point represents one bin, and the dashed line denotes the 45-degree reference line.

Fourth, covariate-stratified posterior predictive checks were used to assess whether the fitted model reproduced the covariate pattern. The observed and predicted case proportions were compared across groups of x_2 , separately within the two strata of $x_1 = 0$ and $x_1 = 1$. If the model captures the covariate structure adequately, the observed and predicted trends are similar within each stratum. Figure 5 shows that the observed and predicted patterns are broadly consistent across x_2 groups in both x_1 strata, indicating that the model reproduces the main covariate pattern built into the data.

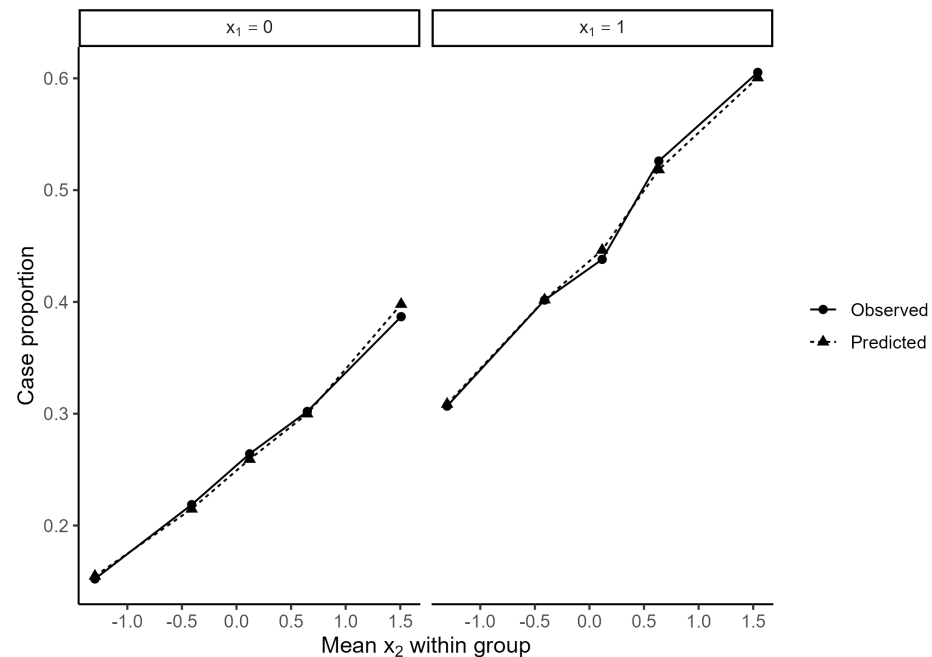


Figure 5. Covariate-stratified posterior predictive check in the IDS dataset. The plot compares observed and posterior predicted case proportions across five groups of the continuous covariate x_2 , separately for the two strata of the binary covariate x_1 .

3.4. Model Comparison

For both the IDS and PPS datasets, the three models with different regional random effect specifications produced nearly identical fixed effect estimates. The estimated standard deviation of each random effect were broadly comparable, although the pure spatial model produced largest posterior mean (Table 4). The LOOIC values were also similar, and all maximum Pareto- k values were below 0.7, indicating reliable PSIS-LOO approximations. The BYM2 model additionally estimated the relative contribution of structured and unstructured regional variation. The posterior means of ρ were 0.660 for IDS and 0.675 for PPS, with 95% credible intervals of [0.346, 0.901] and [0.352, 0.909], respectively. The data indicated that residual heterogeneity is compatible with spatial structure, but the exact proportion attributable to the structured component is estimated with greater uncertainty. Therefore, a model ignoring spatial adjacency is too simple, while a purely spatial ICAR model is also overly restrictive.

3.5. Sensitivity Analysis

The sensitivity analysis indicated that the posterior estimates were stable across the alternative prior specifications. For the IDS dataset, the covariate effects β were almost identical across the main prior and the three sensitivity analysis scenarios (Table 5). The BYM2 hyperparameters showed only modest changes across the alternative hyperpriors, and these changes followed the expected direction without altering the interpretation. When the prior for σ was made more diffuse in Scenario 1, the posterior mean of σ increased

slightly from 1.67 to 1.71, whereas the more concentrated prior in Scenario 2 reduced it to 1.55. The corresponding posterior means of ρ were 0.67 and 0.61, compared with 0.66 under the main prior. Under the more diffuse prior for ρ in Scenario 3, the posterior mean of ρ increased to 0.75 and the credible interval became wider. The PPS sensitivity results, reported in Appendix B Table A1, reveal the same pattern.

Table 4. Comparison of individual-level regional models fitted to datasets based on incidence density sampling (IDS) and probability proportional to size (PPS).

Model	β_1	β_2	Regional SD	LOOIC	Max Pareto- k	$k > 0.7$	Execution Time
<i>IDS dataset</i>							
BYM2	0.989 [0.893, 1.080]	0.499 [0.451, 0.548]	1.67 [1.36, 2.05]	10,852.52	0.516	0	2016.7
IID	0.988 [0.893, 1.080]	0.499 [0.451, 0.548]	1.63 [1.35, 1.97]	10,856.42	0.560	0	283.6
Spatial	0.988 [0.894, 1.080]	0.499 [0.452, 0.547]	1.95 [1.61, 2.35]	10,851.13	0.478	0	343.4
<i>PPS dataset</i>							
BYM2	1.060 [0.962, 1.160]	0.512 [0.463, 0.561]	1.59 [1.30, 1.95]	10,771.71	0.509	0	2165.3
IID	1.060 [0.964, 1.160]	0.512 [0.463, 0.561]	1.55 [1.30, 1.86]	10,777.22	0.422	0	347.3
Spatial	1.060 [0.964, 1.160]	0.512 [0.465, 0.561]	1.83 [1.51, 2.22]	10,770.74	0.475	0	347.4

Values for β_1 , β_2 , and regional SD are posterior means with 95% credible intervals in brackets. Regional SD denotes the standard deviation for the BYM2 model, the standard Bayesian logistic regression with iid regional random intercepts (IID), and the ICAR spatial logistic model (Spatial). LOOIC denotes the leave-one-out information criterion estimated using Pareto-smoothed importance sampling leave-one-out cross-validation. Pareto- k values were used to assess the reliability of the PSIS-LOO approximation. Execution time is reported in seconds.

Table 5. Posterior summaries of the main parameters from the Bernoulli spatial logistic BYM2 model fitted to the IDS dataset under the main prior and three sensitivity analysis scenarios.

Scenario	Parameter	Mean	SD	95% CrI
Main prior	ρ	0.6600	0.1440	[0.3460, 0.9010]
	σ	1.6700	0.1760	[1.3600, 2.0500]
	α	−1.9500	0.1340	[−2.2200, −1.6900]
	β_1	0.9890	0.0491	[0.8930, 1.0800]
	β_2	0.4990	0.0246	[0.4510, 0.5480]
	λ	2.0000	0.8960	[0.6420, 4.0800]
Scenario 1	ρ	0.6695	0.1453	[0.3536, 0.9053]
	σ	1.7095	0.1862	[1.3871, 2.1146]
	α	−1.9580	0.1327	[−2.2178, −1.6923]
	β_1	0.9886	0.0482	[0.8961, 1.0821]
	β_2	0.4993	0.0245	[0.4505, 0.5466]
	λ	2.0122	0.9122	[0.6465, 4.1345]
Scenario 2	ρ	0.6125	0.1477	[0.3009, 0.8656]
	σ	1.5471	0.1452	[1.2876, 1.8570]
	α	−1.9366	0.1299	[−2.1965, −1.6837]
	β_1	0.9874	0.0480	[0.8950, 1.0817]
	β_2	0.4987	0.0252	[0.4499, 0.5476]
	λ	2.0185	0.8929	[0.6732, 4.1175]
Scenario 3	ρ	0.7455	0.1678	[0.3584, 0.9812]
	σ	1.7324	0.1949	[1.3937, 2.1517]
	α	−1.9646	0.1220	[−2.2003, −1.7202]
	β_1	0.9886	0.0476	[0.8961, 1.0838]
	β_2	0.4988	0.0250	[0.4500, 0.5480]
	λ	2.0096	0.9014	[0.6702, 4.1032]

Main prior denotes the prior specification used in the main analysis. Scenario 1 changes only the prior for σ , setting it to $\sigma \sim N^+(0, 2)$; Scenario 2 changes only the prior for σ , setting it to $\sigma \sim N^+(0, 0.5)$; and Scenario 3 changes only the prior for ρ , setting it to $\rho \sim \text{logit-Normal}(0, 2)$.

3.6. Interpretation

Figure 6 presents three complementary summaries of the posterior spatial effects from the Bayesian spatial model. The three panels answer related but distinct questions. Panel A answers: “What is the estimated spatial effect, and how large is it?” Panel B answers: “How much posterior probability supports a positive spatial effect?” Panel C answers: “Does the 95% credible interval lie entirely on one side of zero?” A careful interpretation should avoid dichotomizing the results into simply “significant” and “non-significant” wording [31,32]. Instead, the maps should be read jointly. Areas with large positive posterior means in Panel A, high posterior probabilities in Panel B, and positive 95% credible interval classifications in Panel C provide the most consistent evidence for positive spatial effects. Areas with large negative posterior means, low posterior probabilities of being greater than zero, and negative 95% credible interval classifications provide the most consistent evidence for negative spatial effects. Areas where the three panels are less aligned should be interpreted more cautiously because they may reflect small effect sizes, greater posterior uncertainty, or estimates close to zero.

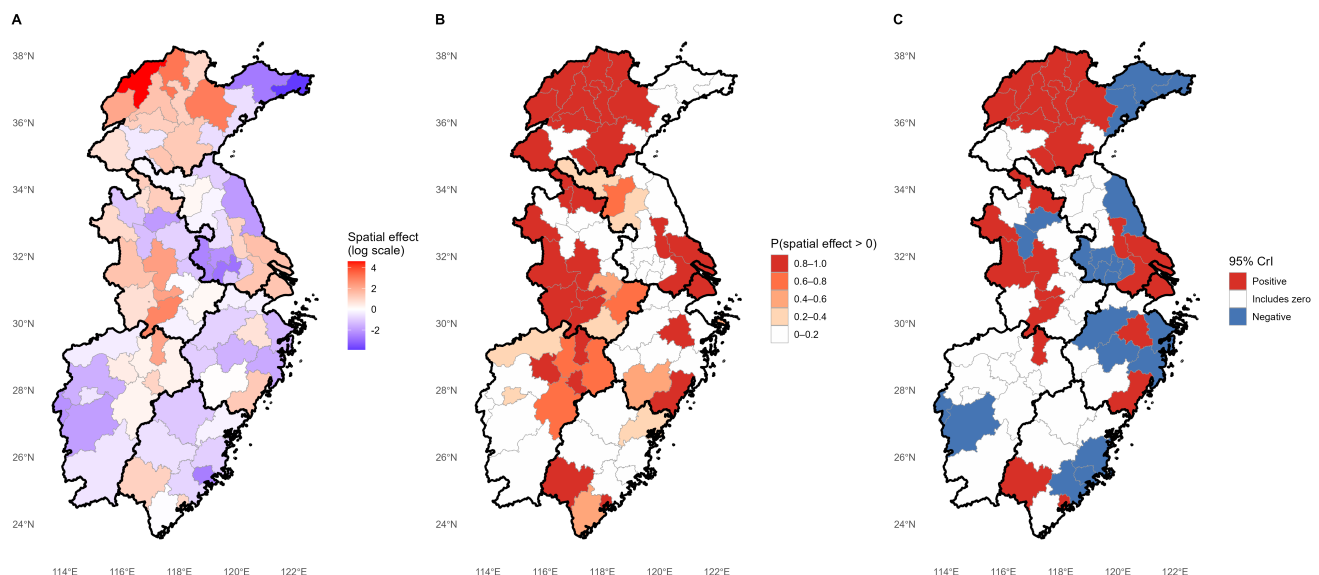


Figure 6. Posterior summaries of city-level BYM2 spatial effects under IDS. Panel (A) shows the posterior median of b_c . Panel (B) shows the posterior exceedance probability $P(b_c > 0 \mid \text{data})$. Panel (C) summarizes the 95% credible interval classification of b_c . Boundary shapefile obtained from National Platform for Common GeoSpatial Information Services, Map Approval No.: GS (2024) 0650.

The fixed effect describes covariate associations after adjustment for residual areal heterogeneity. For covariate x_k , β_k is the adjusted contrast on the linear predictor scale, conditional on the other covariates and the BYM2 areal random effect. Because the fitted model is a Bernoulli logistic model, the exponentiated coefficient is, from the model perspective, an adjusted odds ratio. For ordinary cross-sectional or fixed-period cohort binary outcomes, $\exp(\beta_k)$ is an adjusted odds ratio for the event, comparing individuals who differ by one unit in x_k but are otherwise similar with respect to the other covariates and the areal random effect. In the present illustrative example, its design-based interpretation is a relative rate-type measure of association under the incidence density sampling assumptions [16]. The posterior mean or median provides the central posterior estimate of the covariate association. The 95% credible interval contains 95% of the posterior probability for β_k . Under the fitted Bayesian model and given the observed data, there is a 95% chance that β_k lies within this interval. The interval should be interpreted as posterior uncertainty

around the magnitude of the association, rather than as a rule for declaring whether an association is present or absent [31,32].

4. Discussion

This study extended the established BYM2 disease mapping approach to a multilevel modeling framework by embedding the scaled BYM2 random effect within a Bernoulli logistic regression model. The model was implemented in Stan and illustrated through an open-cohort example in which time-to-event source data were converted into binary analytical datasets using case–control study design. The example was used as a worked analysis rather than as a simulation benchmark, showing how the model can be fitted, checked, compared with alternative regional random effect specifications, and interpreted in practice. The following discussion focuses on the statistical rationale, interpretation, limitations, potential model extensions and practical use of this workflow in veterinary epidemiological applications.

4.1. Basis of the Individual-Level BYM2 Formulation

The motivation for the present model was to allow a disease mapping model to handle individual-level predictors and outcome variables, while preserving the ability to account for residual spatial variation. This formulation is supported by the broader spatial multilevel modeling literature. Standard multilevel logistic regression can account for clustering within areas through independent area-level random intercept, but it does not explicitly model spatial dependence between neighboring areas. Xu compared standard multilevel, pure spatial, and hybrid spatial models for binary outcomes and showed that models accounting for both within-area and between-area dependence can differ particularly in their estimation of random-effect variation [33]. Arcaya et al. similarly emphasized that geographical membership and spatial proximity can represent distinct sources of area variation [34]. Browne and Goldstein also showed, from a general multilevel modeling perspective, that cluster-level residuals need not be independent and may be modeled with more general correlation structures [35]. These studies provide conceptual support for replacing an independent area-level random intercept with a spatially structured disease mapping component when residual geographical dependence is expected. Recent medical studies have used related spatial random-effect ideas, but for different inferential purposes. Davis et al. incorporated scaled BYM spatial random effects into propensity score and outcome models to address geographic confounding in health disparity studies [3]. Janssens et al. used a logistic model with patient-level covariates, a spatially structured municipality-level random effect, and an unstructured practice-level random effect for disease mapping using primary care registry data [36]. These studies support the broader need to account for residual spatial variation in individual-level veterinary epidemiological data.

4.2. Relationship to Poisson BYM2 and Sparse Region Behavior

This individual-level perspective also clarifies how the proposed Bernoulli logistic BYM2 formulation should be distinguished from the more familiar aggregated Poisson BYM2 disease mapping model [13]. The present model does not introduce a different spatial smoothing mechanism. The borrowing of information across neighboring regions is inherited from the scaled BYM2 component [7]. What changes is the likelihood through which the data inform the areal random effects. Consequently, the global dispersion parameter σ has the same structural role in both models, but it is not generally numerically equivalent across the two likelihoods. The connection between the two formulations is closest under rare outcome conditions where the sum of individual Bernoulli outcomes within a region can be approximated by a Poisson count. When individual-level covariates

are absent or homogeneous within regions, the individual-level Bernoulli model approaches an aggregated Poisson representation.

Outside this restricted setting, however, aggregation changes the unit of analysis and can change the estimand [14,15]. Therefore, the purpose of the present framework is not to demonstrate that the individual-level BYM2 model is universally superior to aggregated Poisson BYM2 model. Rather, the practical advantage is estimand alignment. When the data support a research question concerning individual-level covariate associations meanwhile adjusting for spatial correlation, the Bernoulli BYM2 formulation preserves the individual as the unit of analysis while retaining the spatial smoothing component. Aggregated Poisson BYM2 remains appropriate when the scientific target is regional disease burden.

This distinction also clarifies the expected behavior of the model in sparse regions. In a Poisson model, the information contributed by region c for its areal effect is approximately proportional to the expected regional count. In the Bernoulli logistic model, the corresponding information is approximately $I_c^L = \sum_{i:c(i)=c} p_i(1 - p_i)$. This quantity is bounded above by $I_c^L \leq \frac{n_c}{4}$, and under rare-event conditions, it is approximately $\sum_{i:c(i)=c} p_i$, that is, the expected number of cases in region c . Therefore, when n_c is small or when events are rare, the likelihood contribution for b_c is weak, so the posterior for b_c is more strongly determined by the BYM2 hierarchical prior, including the adjacency structure and the posterior distribution of the hyperparameters σ and ρ [7,37]. If σ and ρ are also weakly identified by the data, then their hyperpriors may have greater influence [38]. For a region with $n_c = 0$, there is no direct likelihood contribution, and the posterior areal effect should be interpreted as a spatially smoothed prediction rather than as a data-supported estimate. This behavior is not a special weakness of the logistic formulation, it is the expected consequence of combining sparse regional data with a spatial smoothing prior [6,7]. This also motivates the discussion on prior sensitivity analysis in the later text.

4.3. Design Consideration

Extreme sparsity should also be considered as a study design limitation rather than only a model fitting challenge. As we mentioned previously, the BYM2 prior can regularize unstable regional estimates and provide spatially smoothed predictions, but it cannot create direct likelihood information for regions where few or no individuals are observed. This distinction is consistent with the small area estimation literature, where model-based borrowing of strength is used to improve unstable small area estimates but does not remove the need for sufficient information at an appropriate spatial resolution [39–41]. In applied studies, such settings should be addressed through design choices before model fitting, for example by increasing the total sample size, extending the observation period, using a coarser spatial resolution, combining regions with very limited information when scientifically defensible, or using an efficient sampling design for rare events. These choices increase the amount of information available to estimate regional heterogeneity. The role of the BYM2 component is then to account for residual areal variation.

Beyond spatial sparsity, an additional design consideration arises when the source data come from an open cohort, where the outcome is naturally represented by event time rather than by a single binary response. To apply the proposed Bernoulli BYM2 model in this setting, the time-to-event source data must first be translated into a binary analytical structure that preserves the target interpretation [42–44]. In the illustrative example, this was done by using IDS and PPS sampling designs to construct analytical datasets from the open cohort. The results showed that the modeling and sampling framework behaved as expected under controlled conditions. These findings should be interpreted as a methodological demonstration rather than as evidence that either sampling design is universally preferable. The main point is that, when an open cohort source

population contains spatially structured event-time data, an appropriately constructed binary analytical dataset can allow the individual-level Bernoulli BYM2 model to estimate the rate-type association while adjusting for the spatial heterogeneity.

4.4. Practical Model Comparison

We compared the proposed model with a standard Bayesian logistic mixed effects model and a ICAR spatial logistic model. This comparison should be interpreted as a practical model assessment step, as would be done in a real-world data analysis where researchers collect the data but do not know the underlying data-generating mechanism. It should therefore not be interpreted as a formal attempt to prove that one model is universally superior to the others. The standard and ICAR spatial models represent two limiting assumptions about residual regional heterogeneity [7]. The BYM2 model is useful because it does not require the analyst to fix either assumption in advance but instead provides a data-supported estimate of the relative contribution of structured and unstructured regional variation through ρ .

We found a wide 95% credible interval of ρ . This is expected because ρ is a hyperparameter governing the decomposition of the latent region level random effects. This decomposition is inferred indirectly from the residual pattern across the regions. Therefore, the ρ estimates should be interpreted as evidence that residual heterogeneity is compatible with spatial structure. However, when the number of regions is moderate, e.g., 77 in this study, one should not expect ρ to provide a highly precise estimate of the exact proportion of spatially structured variation. The regional standard deviation estimates should also be compared cautiously across models because they represent different forms of regional variation. In the standard mixed effects logistic model, this variation is represented by independent area-level deviations, in the ICAR model, it is represented by a spatially smoothed field defined by the adjacency graph [6], and in the BYM2 model, it is represented by a combination of structured and unstructured components [7]. Thus, the standard deviation estimates are informative about the scale of residual regional variation within each model, but they should not be interpreted as directly equivalent measures across models.

In both IDS and PPS analyses, the fixed effect estimates were nearly unchanged across the three regional effect specifications, suggesting that the covariate associations in this illustrative example were not sensitive to the regional random effect structures. However, this should not be interpreted as a general guarantee that fixed effects are robust to arbitrary spatial misspecification. The similar LOOIC values also indicated that pointwise predictive performance alone did not provide a decisive basis for choosing among these spatial hierarchical models. These, however, do not make model comparison unnecessary; rather, it illustrates the type of assessment that is needed in practical data analysis. In applied settings, the analyst needs to examine whether the choice of regional random effect structure changes the fixed effect estimates, whether predictive criteria provide evidence for one specification over another, whether the PSIS-LOO diagnostics are reliable, and whether the BYM2 estimate of ρ supports the flexibility of the structured/unstructured mixture or instead favors a more parsimonious model.

4.5. Sensitivity Analysis

The sensitivity analysis provides a check on how different parts of the model are informed by the likelihood and by the prior. The fixed effect estimates were stable across the different sensitivity analyses, which was expected because these coefficients were assigned diffuse priors and were informed by the individual-level likelihood from a large analytical dataset. By contrast, the BYM2 hyperparameters σ and ρ require more careful consideration. These two quantities are informed mainly by the regional pattern of residual

variation. Therefore, they are naturally more sensitive to the number of regions, the balance of regional sample sizes, the strength of the spatial pattern, and the chosen hyperpriors. In the present analysis, the changes in σ and ρ followed the direction implied by the alternative priors, but changes were not large enough to alternate the interpretation.

Although σ and ρ are usually not the primary inferential targets, their prior sensitivity remains important, especially for ρ , as it helps assess, together with fixed effect stability, PSIS-LOO, and scientific plausibility, whether the flexible BYM2 decomposition is needed or whether a more parsimonious regional effect model may be sufficient. This role of ρ is most useful when its posterior is not dominated by a strongly informative prior. Under weakly informative priors, the posterior distribution of ρ is allowed to move towards the unstructured or structured side if the residual regional pattern supports such behavior. Therefore, if the posterior of ρ remains broadly consistent across reasonable prior choices, it can provide another evidence about whether the data support the additional BYM2 decomposition.

More generally, prior choice should be guided by the scale on which prior knowledge is most interpretable. Regression coefficients are often difficult to specify directly; therefore, prior information can be elicited on the response scale under meaningful covariate patterns and then transformed to the coefficient scale, as in previous Bayesian veterinary epidemiological applications [45,46]. The same principle applies to the BYM2 hyperparameters. For ρ , prior beliefs can be expressed on the probability scale as the expected proportion of residual regional variation attributable to spatial structure. An analyst may specify a best estimate together with a probability statement, such as being 95% confident that ρ is below or above a chosen value, and then convert this information into a logit-normal or beta prior [45,47,48]. For σ , prior choice can be guided by prior predictive reasoning. Before model fitting, the analyst can examine the range of regional effects generated under a proposed prior, taking into account the BYM2 scaling, the prior for ρ , and the adjacency structure, and then assess whether this range is plausible for the disease given the areal units used in the analysis [49,50]. However, such prior knowledge is often unavailable in applied spatial analysis, which again highlights the importance of sensitivity analysis.

4.6. Stan Implementation, Computation, and Future Extensions

Spatial disease mapping models are often fitted using the integrated nested Laplace approximation (INLA) [51–53], which is considered computationally efficient for latent Gaussian spatial models. However, the exact individual-level model specification and prior structure used in this work were not directly available through standard INLA formula syntax. In contrast, this study used full Bayesian approach via HMC/NUTS in Stan [17,18]. The main motivation was flexibility rather than computational speed since the execution times in the present analysis were manageable for an applied research work (Table 4). Being flexible and user-controllable are useful because it makes the complete Bayesian workflow transparent, including model formulation, posterior computation, convergence assessment and posterior predictive checking.

The flexibility of Stan is also relevant for future extensions of the proposed framework [19]. This is particularly useful for individual-level spatial epidemiological data, where studies may differ in their data structures, outcome types, and dependence assumptions. For example, the Bernoulli likelihood used here could be extended to longitudinal binary outcomes, spatiotemporal disease mapping, multinomial or ordinal disease states, time-to-event or recurrent events outcome. However, whether a particular extension is practically useful will depend on the data structure, parameter identifiability, computational performance, and the ability of the model to pass appropriate diagnostic and predictive

checks. Therefore, such extensions should be treated as a separate model development rather than as automatic consequences of the present framework.

5. Conclusions

In conclusion, this study extends the BYM2 disease mapping framework to model individual-level binary outcome data. The proposed model allows individual-level co-variate associations to be estimated while accounting for residual structured and unstructured regional heterogeneity. Through an illustrative example, the study showed how time-to-event source data can be converted into binary analytical datasets through appropriate case–control sampling methods when the target is a relative association rather than an absolute risk. The Stan implementation further demonstrated a complete Bayesian workflow, including posterior computation, convergence assessment, posterior predictive checking, model comparison, prior sensitivity analysis, and interpretation. The framework is potentially adaptable to more complex hierarchical data structures and alternative outcome types, supporting future applications of individual-level spatial modeling in veterinary epidemiology.

Author Contributions: Conceptualization, D.A.Y.; methodology, C.Z., Z.J. and D.A.Y.; software, D.A.Y. and C.Z.; validation, C.Z. and D.A.Y.; formal analysis, C.Z.; data curation, C.Z., D.A.Y. and H.Y.; writing—original draft, C.Z., H.Y. and Z.J.; writing—review and editing, C.Z., H.Y. and D.A.Y.; visualization, D.A.Y., H.Y. and C.Z.; supervision, D.A.Y.; project administration, D.A.Y.; funding acquisition, D.A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Young Scientist C) under fund no. 32503075.

Data Availability Statement: The data presented in this study are openly available in Github at <https://github.com/aaronpopolar/multilevelBYM2> (accessed on 14 July 2026).

Acknowledgments: The authors thank the reviewers for their constructive comments and Nanjing Agricultural University for providing internet access and computational resources.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BYM	Besag–York–Mollié
BYM2	Besag–York–Mollié 2
E-BFMI	Estimated Bayesian fraction of missing information
HMC	Hamiltonian Monte Carlo
ICAR	Intrinsic conditional autoregressive
IDS	Incidence density sampling
IID	Independent and Identically Distributed
INLA	Integrated nested Laplace approximation
MCMC	Markov chain Monte Carlo
NUTS	No-U-Turn Sampler
PPS	Probability proportional to size sampling

Appendix A. MCMC Trace Plots

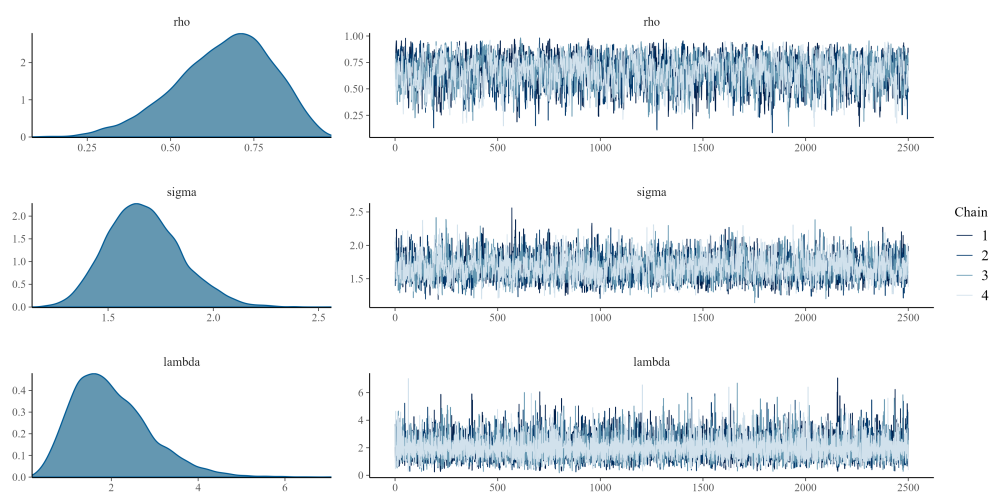


Figure A1. Trace plots for the IDS sampling model, part 1.

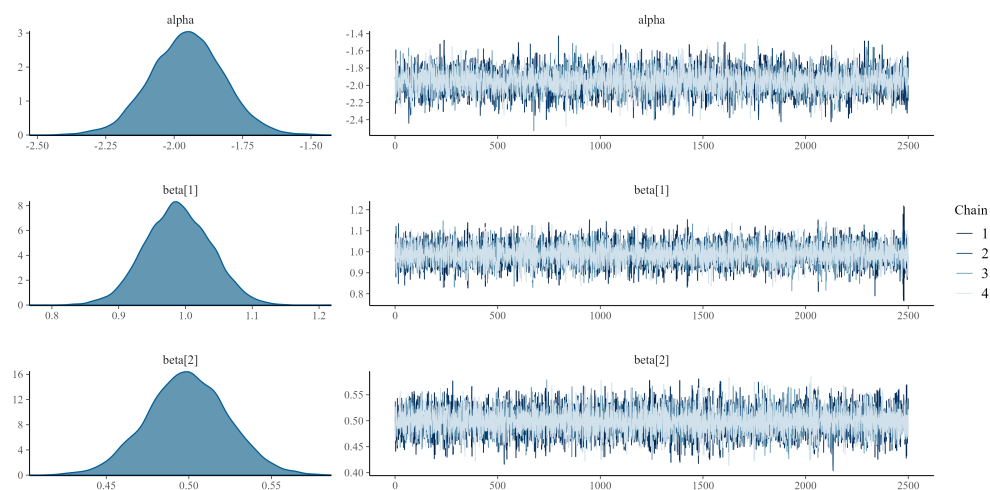


Figure A2. Trace plots for the IDS sampling model, part 2.

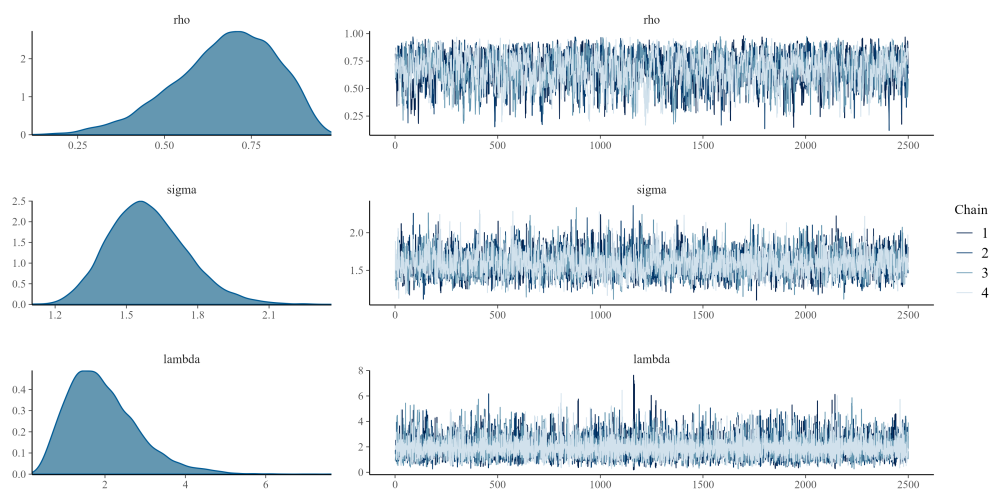


Figure A3. Trace plots for the PPS sampling model, part 1.

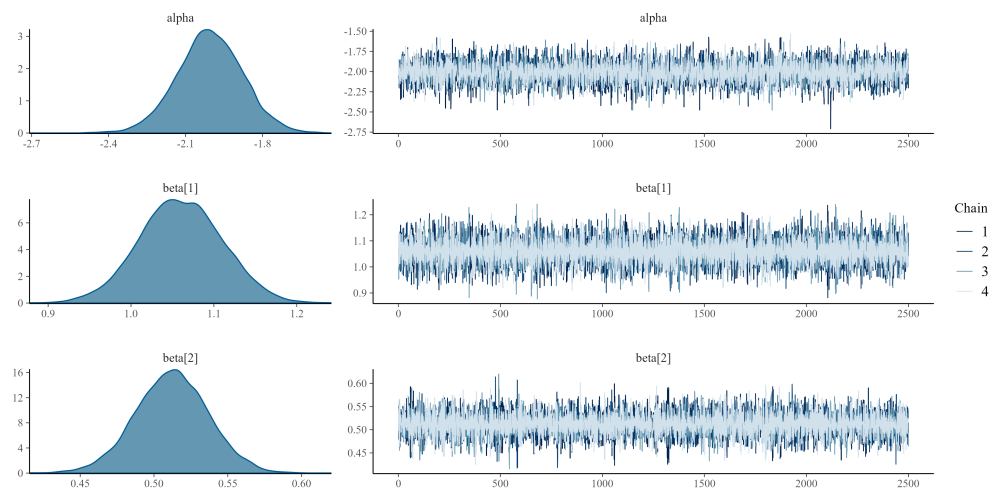


Figure A4. Trace plots for the PPS sampling model, part 2.

Appendix B. Prior Sensitivity Analysis for the PPS Dataset

Table A1. Posterior summaries of the parameters from the Bernoulli logistic BYM2 model fitted to the PPS dataset under three sensitivity analysis scenarios.

Scenario	Parameter	Mean	SD	95% CrI
Scenario 1	ρ	0.6954	0.1404	[0.3759, 0.9116]
	σ	1.6284	0.1738	[1.3260, 2.0073]
	α	−2.0117	0.1259	[−2.2568, −1.7620]
	β_1	1.0606	0.0487	[0.9660, 1.1567]
	β_2	0.5122	0.0249	[0.4639, 0.5614]
	λ	1.9516	0.8871	[0.6352, 4.0238]
Scenario 2	ρ	0.6451	0.1437	[0.3410, 0.8912]
	σ	1.4847	0.1405	[1.2319, 1.7880]
	α	−1.9876	0.1225	[−2.2284, −1.7469]
	β_1	1.0587	0.0494	[0.9613, 1.1555]
	β_2	0.5115	0.0249	[0.4623, 0.5607]
	λ	1.9476	0.8837	[0.6284, 4.0061]
Scenario 3	ρ	0.7918	0.1520	[0.4249, 0.9883]
	σ	1.6611	0.1833	[1.3340, 2.0435]
	α	−2.0173	0.1132	[−2.2330, −1.7892]
	β_1	1.0598	0.0502	[0.9610, 1.1591]
	β_2	0.5120	0.0249	[0.4635, 0.5615]
	λ	1.9511	0.8687	[0.6458, 3.9886]

Scenario 1 changes only the prior for σ , setting it to $\sigma \sim N^+(0, 2)$; Scenario 2 changes only the prior for σ , setting it to $\sigma \sim N^+(0, 0.5)$; and Scenario 3 changes only the prior for ρ , setting it to $\rho \sim \text{logit-Normal}(0, 2)$.

References

1. Zhang, Y.; Yang, D.A.; Yang, M.; Pi, M.; Zhang, Y.; Su, Z. Prevalence and Associated Risk Factors of Helminth Infections in the Digestive Tract of Camels in Xinjiang, China. *Vet. Sci.* **2024**, *11*, 503. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Li, Y.; Huyen, N.T.; Cuc, N.T.K.; Dinh, N.C.; Thong, T.T.; Ha, L.T.; Nguyen, N.P.T.; Hernandez-Jover, M.; Kelly, J.; Manyweathers, J. Incidence and Risk Factors of Infectious Diseases in Smallholder Pig Farms in Vietnam. *Res. Vet. Sci.* **2025**, *193*, 105726. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Davis, M.L.; Neelon, B.; Nietert, P.J.; Burgette, L.F.; Hunt, K.J.; Lawson, A.B.; Eggede, L.E. Propensity Score Matching for Multilevel Spatial Data: Accounting for Geographic Confounding in Health Disparity Studies. *Int. J. Health Geogr.* **2021**, *20*, 10. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, Y.; Chen, X.; Xue, F. A Review of Bayesian Spatiotemporal Models in Spatial Epidemiology. *ISPRS Int. J.-Geo-Inf.* **2024**, *13*, 97. [\[CrossRef\]](#)

5. MacNab, Y.C. Bayesian Disease Mapping: Past, Present, and Future. *Spat. Stat.* **2022**, *50*, 100593. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Besag, J.; York, J.; Mollié, A. Bayesian Image Restoration, with Two Applications in Spatial Statistics. *Ann. Inst. Stat. Math.* **1991**, *43*, 1–20. [\[CrossRef\]](#)
7. Riebler, A.; Sørbye, S.H.; Simpson, D.; Rue, H. An Intuitive Bayesian Spatial Model for Disease Mapping That Accounts for Scaling. *Stat. Methods Med. Res.* **2016**, *25*, 1145–1165. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Teng, K.T.; Martínez Avilés, M.; Ugarte-Ruiz, M.; Barcena, C.; de la Torre, A.; Lopez, G.; Moreno, M.A.; Dominguez, L.; Alvarez, J. Spatial Trends in Salmonella Infection in Pigs in Spain. *Front. Vet. Sci.* **2020**, *7*, 345. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Ayubi, E.; Shirzadi, M.R.; Amiri, B.; Pourmozafari, J.; Faghih Soleimani, M.; Khosronejad, S.; Khazaei, S. Exploring the Spatial Pattern of Animal Bites in Iran (2021–2022). *Prev. Vet. Med.* **2023**, *221*, 106056. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Morales-Otero, M.; Núñez-Antón, V. Comparing Bayesian Spatial Conditional Overdispersion and the Besag–York–Mollié Models: Application to Infant Mortality Rates. *Mathematics* **2021**, *9*, 282. [\[CrossRef\]](#)
11. Ward, T.; Morris, M.; Gelman, A.; Carpenter, B.; Ferguson, W.; Overton, C.; Fyles, M. Bayesian Spatial Modelling of Localised SARS-CoV-2 Transmission through Mobility Networks across England. *PLoS Comput. Biol.* **2023**, *19*, e1011580. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Bovo, E.; Belloni, P.; Sottosanti, A.; Boccuzzo, G. A Model-Based Scan Statistic with Enhanced Specificity for Detecting Spatial Clusters of High Mortality Risk. *Environ. Ecol. Stat.* **2025**, *32*, 1489–1518. [\[CrossRef\]](#)
13. Mussafo, A.J.A.; da Cruz, R.F.; de Arruda Silveira, L.V.; Govone, J.S. Climatic and Social Health Determinants of Malaria Incidence in Tete, Mozambique: Bayesian Spatio-Temporal Modeling of Routine Data (2016–2022). *Spat.-Spatio-Temporal Epidemiol.* **2026**, *57*, 100802. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Greenland, S. Ecologic versus Individual-Level Sources of Bias in Ecologic Estimates of Contextual Health Effects. *Int. J. Epidemiol.* **2001**, *30*, 1343–1350. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Roumeliotis, S.; Abd ElHafeez, S.; Jager, K.J.; Dekker, F.W.; Stel, V.S.; Pitino, A.; Zoccali, C.; Tripepi, G. Be Careful with Ecological Associations. *Nephrology* **2021**, *26*, 501–505. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Dohoo, I.; Martin, W.; Stryhn, H. *Veterinary Epidemiologic Research*, 2nd ed.; VER Inc.: Charlottetown, PE, Canada, 2009. Available online: <https://projects.upei.ca/ver/> (accessed on 19 June 2026).
17. Hoffman, M.D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623. Available online: <https://jmlr.org/papers/v15/hoffman14a.html> (accessed on 19 June 2026).
18. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A Probabilistic Programming Language. *J. Stat. Softw.* **2017**, *76*, 1–32. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Morris, M.; Wheeler-Martin, K.; Simpson, D.; Mooney, S.J.; Gelman, A.; DiMaggio, C. Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan. *Spat.-Spatio-Temporal Epidemiol.* **2019**, *31*, 100301. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Rue, H.; Held, L. *Gaussian Markov Random Fields: Theory and Applications*; Chapman and Hall/CRC: New York, NY, USA, 2005. [\[CrossRef\]](#)
21. Besag, J.; Kooperberg, C. On Conditional and Intrinsic Autoregressions. *Biometrika* **1995**, *82*, 733–746. [\[CrossRef\]](#)
22. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288. [\[CrossRef\]](#)
23. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [\[CrossRef\]](#)
24. Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* **2021**, *16*, 667–718. [\[CrossRef\]](#)
25. Grekov, P.; Pustejovsky, J.E. A Gentle Introduction to Bayesian Posterior Predictive Checking for Single-Case Researchers. *J. Behav. Educ.* **2026**, in press. [\[CrossRef\]](#)
26. Zhang, K.; Laven, R.; Wang, C.; Li, B.; Li, M.; Yang, D.A. Retrospective Survival Analysis of Feline Lower Urinary Tract Disease in China Using Insurance Data. *Vet. J.* **2026**, *317*, 106673. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Lohr, S.L. *Sampling: Design and Analysis*, 3rd ed.; Chapman and Hall/CRC: New York, NY, USA, 2021. [\[CrossRef\]](#)
28. Partlett, C.; Hall, N.J.; Leaf, A.; Juszcak, E.; Linsell, L. Application of the Matched Nested Case-Control Design to the Secondary Analysis of Trial Data. *BMC Med. Res. Methodol.* **2020**, *20*, 117. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Vehtari, A.; Gelman, A.; Gabry, J. Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC. *Stat. Comput.* **2017**, *27*, 1413–1432. [\[CrossRef\]](#)
30. Vehtari, A.; Simpson, D.; Gelman, A.; Yao, Y.; Gabry, J. Pareto Smoothed Importance Sampling. *J. Mach. Learn. Res.* **2024**, *25*, 1–58.
31. Greenland, S.; Senn, S.J.; Rothman, K.J.; Carlin, J.B.; Poole, C.; Goodman, S.N.; Altman, D.G. Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. *Eur. J. Epidemiol.* **2016**, *31*, 337–350. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Laven, R.; Yang, D.A. Common Misinterpretations of Statistical Significance and P-Values in Dairy Research. *JDS Commun.* **2025**, *6*, 721–726. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Xu, H. Comparing Spatial and Multilevel Regression Models for Binary Outcomes in Neighborhood Studies. *Sociol. Methodol.* **2014**, *44*, 229–272. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Arcaya, M.; Brewster, M.; Zigler, C.M.; Subramanian, S.V. Area Variations in Health: A Spatial Multilevel Modeling Approach. *Health Place* **2012**, *18*, 824–831. [[CrossRef](#)] [[PubMed](#)]
35. Browne, W.; Goldstein, H. MCMC Sampling for a Multilevel Model With Nonindependent Residuals Within and Between Cluster Units. *J. Educ. Behav. Stat.* **2010**, *35*, 453–473. [[CrossRef](#)]
36. Janssens, A.; Vaes, B.; Van Pottelbergh, G.; Libin, P.J.K.; Neyens, T. Model-Based Disease Mapping Using Primary Care Registry Data. *Spat.-Spatio-Temporal Epidemiol.* **2024**, *49*, 100654. [[CrossRef](#)] [[PubMed](#)]
37. Sørbye, S.H.; Rue, H. Scaling Intrinsic Gaussian Markov Random Field Priors in Spatial Modelling. *Spat. Stat.* **2014**, *8*, 39–51. [[CrossRef](#)]
38. Simpson, D.; Rue, H.; Martins, T.G.; Riebler, A.; Sørbye, S.H. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Stat. Sci.* **2017**, *32*, 1–28. [[CrossRef](#)]
39. Ghosh, M.; Rao, J.N.K. Small Area Estimation: An Appraisal. *Stat. Sci.* **1994**, *9*, 55–93. [[CrossRef](#)]
40. Rao, J.N.K.; Molina, I. *Small Area Estimation*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2015. [[CrossRef](#)]
41. Pfeffermann, D. New Important Developments in Small Area Estimation. *Stat. Sci.* **2013**, *28*, 40–68. [[CrossRef](#)]
42. Shin, Y.E.; Saegusa, T. Nested Case–Control Sampling without Replacement. *Lifetime Data Anal.* **2024**, *30*, 776–799. [[CrossRef](#)] [[PubMed](#)]
43. Langholz, B. Case-Control Studies = Odds Ratios: Blame the Retrospective Model. *Epidemiology* **2010**, *21*, 10–12. [[CrossRef](#)] [[PubMed](#)]
44. Labrecque, J.A.; Hunink, M.M.G.; Ikram, M.A.; Ikram, M.K. Do Case-Control Studies Always Estimate Odds Ratios? *Am. J. Epidemiol.* **2020**, *190*, 318–321. [[CrossRef](#)] [[PubMed](#)]
45. Vink, W.D.; Jones, G.; Johnson, W.O.; Brown, J.; Demirkan, I.; Carter, S.D.; French, N.P. Diagnostic assessment without cut-offs: Application of serology for the modelling of bovine digital dermatitis infection. *Prev. Vet. Med.* **2009**, *92*, 235–248. [[CrossRef](#)] [[PubMed](#)]
46. Yang, D.A.; Gates, M.C.; Müller, K.R.; Laven, R.A. Bayesian analysis of herd-level risk factors for bovine digital dermatitis in New Zealand dairy herds. *BMC Vet. Res.* **2019**, *15*, 125. [[CrossRef](#)] [[PubMed](#)]
47. Yang, D.A.; Heuer, C.; Laven, R.; Vink, W.D.; Chesterton, R.N. Estimating the true prevalence of bovine digital dermatitis in Taranaki, New Zealand using a Bayesian latent class model. *Prev. Vet. Med.* **2017**, *147*, 158–162. [[CrossRef](#)] [[PubMed](#)]
48. Yang, D.A.; Johnson, W.O.; Müller, K.R.; Gates, M.C.; Laven, R.A. Estimating the herd and cow level prevalence of bovine digital dermatitis on New Zealand dairy farms: A Bayesian superpopulation approach. *Prev. Vet. Med.* **2019**, *165*, 76–84. [[CrossRef](#)] [[PubMed](#)]
49. Gelman, A.; Simpson, D.; Betancourt, M. The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy* **2017**, *19*, 555. [[CrossRef](#)]
50. Gabry, J.; Simpson, D.; Vehtari, A.; Betancourt, M.; Gelman, A. Visualization in Bayesian Workflow. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **2019**, *182*, 389–402. [[CrossRef](#)]
51. Colwell, S.; Chalise, P.; Gajewski, B.; Ratnayake, I.; Mudaranthakam, D.P. Spatial and Temporal Modeling of Breast Cancer Mortality in Kansas: An R-INLA Approach. *PLoS ONE* **2026**, *21*, e0347607. [[CrossRef](#)] [[PubMed](#)]
52. Lindgren, F.; Rue, H. Bayesian Spatial Modelling with R-INLA. *J. Stat. Softw.* **2015**, *63*, 1–25. [[CrossRef](#)]
53. Nadifar, M.; Bekker, A.; Arashi, M.; Ramoelo, A. Bayesian Semi-Parametric Spatial Modeling of Dispersed Count Data with INLA: Applications in Public Health and Climate Risk. *Stoch. Environ. Res. Risk Assess.* **2026**, *40*, 48. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.