

Small-area estimation for the Iowa Seat-Belt Use Survey, 2017–2021

Zirou Zhou¹, Emily Berg²

Abstract

This study develops Bayesian small-area estimates of seat-belt use for Iowa counties from 2017 to 2021. The aim is to produce stable county-by-year estimates and measures of uncertainty for all 99 Iowa counties, although only 15 counties were directly sampled in the Iowa Seat-Belt Use Survey. The analysis focuses on two policy-relevant quantities: the percentage of belted vehicle miles traveled and the percentage of belted vehicle occupants. To address the multivariate, longitudinal, and boundary-inflated structure of the data, we compare several Bayesian hierarchical models for driver and passenger seat-belt-use proportions and combine them with a Poisson log-linear model for occupant counts. Correlated random effects are used to borrow strength across counties, years, and road segments, and posterior inference is implemented using Integrated Nested Laplace Approximations. Model performance is evaluated using WAIC, posterior predictive checks, posterior intervals, Bayesian model averaging, and comparisons with direct estimates. The empirical results show that the Bayesian estimates are generally close to direct estimates in posterior mean, while often producing substantially smaller posterior standard deviations. These gains indicate improved stability for local traffic-safety summaries, although they are interpreted as a bias–variance trade-off rather than as uniform dominance over direct estimation.

Key words: small area estimation, Bayesian hierarchical model, seat-belt survey, one-inflated proportion, INLA.

1. Introduction

Reliable local estimates of seat-belt use are important for traffic-safety planning, but they are difficult to obtain when survey samples are small within local domains. The Iowa Seat-Belt Use Survey provides annual information on seat-belt use, but it was designed primarily to support statewide estimation rather than direct county-by-year inference. The research problem addressed in this paper is therefore to estimate county-by-year seat-belt-use parameters for all 99 Iowa counties from 2017 to 2021 when direct county estimates are unstable or unavailable. Only 15 counties are directly sampled, and each sampled county contains a limited number of road segments observed across years. As a result, direct estimators can have large sampling variation, and direct estimates cannot be obtained for non-sampled counties. In this paper, the inferential target is all county-by-year domains in Iowa. Estimates for counties without sampled road segments are interpreted as model-based

¹Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, IA, USA. E-mail: zzhou@iastate.edu. ORCID: <https://orcid.org/0000-0002-4059-7871>.

²Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, IA, USA. E-mail: emilyb@iastate.edu. ORCID: <https://orcid.org/0000-0002-0962-005X>.



posterior predictions that rely on auxiliary variables and hierarchical borrowing of strength, rather than as design-based direct estimates.

The practical motivation comes from the need to provide the Iowa Governor's Traffic Safety Bureau with more stable summaries of local seat-belt use and possible changes over time. Two related but distinct quantities are of interest. The first is the proportion of vehicle miles traveled for which drivers or passengers are belted, which weights road segments by traffic exposure. The second is the proportion of belted drivers or passengers among vehicle occupants, which depends on both the seat-belt-use proportions and the number of occupants observed on each road segment. These two summaries provide complementary information for program evaluation and traffic-safety planning. The data structure creates several modeling challenges: the responses are multivariate because drivers and passengers are observed jointly; the proportions are bounded and often equal to one; the occupant counts are discrete; and the same road segments are revisited from 2017 to 2021, creating a longitudinal structure.

Small area estimation provides a natural framework for this problem because it combines survey data with model-based information to improve the stability of domain-level estimates (Rao and Molina, 2015). Bayesian hierarchical methods are especially useful when the target parameters are nonlinear functions of unit-level quantities and when uncertainty must be propagated through several stages of prediction (Molina et al., 2014). Recent work has further emphasized Bayesian and unit-level modeling for complex survey data, including informative sampling, non-Gaussian responses, and repeated or longitudinal survey structures (Parker et al., 2023a,b; Vedensky et al., 2025a; McGovern et al., 2026). Previous work has extended small area methods to binary, count, and other non-Gaussian outcomes, including logistic mixed models (González-Manteiga et al., 2007), zero-inflated models with correlated area effects (Lyu et al., 2020), and broader unit-level Bayesian approaches for complex survey data (Parker et al., 2023a; Wu and Stephenson, 2023). Related Bayesian models have also been developed for binomial or beta-binomial structures with spatial or multilevel dependence (Dong and Wakefield, 2021), for applications that jointly model proportions and sample sizes (Oleson et al., 2007), and for longitudinal or multi-type survey responses with domain-level dependence (Vedensky et al., 2025a,b). Bayesian model averaging and related Bayesian model-combination strategies have also been used to account for model uncertainty in small area estimation and spatial disease mapping (Aitkin et al., 2009; Hoeting et al., 1999; Mohebbi et al., 2014; Sugawara et al., 2024). These studies show the value of flexible hierarchical modeling for complex small area problems, but they do not directly address the combination of one-inflated seat-belt-use proportions, occupant counts, repeated road-segment observations, and driver-passenger dependence in a multiyear county-level application.

Several studies are directly related to the Iowa seat-belt survey. Berg (2022) developed databases for small area estimation using the Iowa seat-belt data. Berg (2023) proposed a Bayesian unit-level Poisson model with multivariate random effects for seat-belt counts. Zhou and Berg (2024) developed a unit-level one-inflated beta model for seat-belt-use rates, but the analysis was limited to driver data from one year. These studies provide important foundations, but each addresses only part of the current problem. A count-based model captures traffic volume and occupant counts, but it does not directly model the bounded

seat-belt-use rate. A one-inflated beta model captures the boundary behavior of proportions, but the previous application did not jointly use drivers and passengers across multiple years. The remaining gap is an integrated application-specific Bayesian small area framework that compares alternative proportion models, incorporates occupant counts, and produces county-by-year predictions for both sampled and nonsampled counties.

This paper develops such a framework for the Iowa Seat-Belt Use Survey from 2017 to 2021. We model driver and passenger seat-belt-use proportions using several candidate distributions, including one-inflated beta, beta-epsilon, logit-normal-epsilon, beta-binomial, and binomial specifications. We also model the numbers of drivers and passengers using a Poisson log-linear count model. The proportion and count components are combined to estimate the two policy-relevant small area parameters: the percentage of belted vehicle miles traveled and the percentage of belted occupants. Correlated random effects are used to represent county, county-by-year, and road-segment sources of variation, allowing the model to borrow strength across related domains while preserving the longitudinal and multivariate structure of the survey. The Bayesian implementation is based on latent Gaussian modeling and Integrated Nested Laplace Approximations, which provide computationally feasible posterior inference for these hierarchical models (Rue et al., 2009).

The main contribution of this study is not the introduction of a new general distribution, but the construction and evaluation of an integrated Bayesian small area estimation strategy for a complex traffic-safety survey. Compared with simpler existing models, the proposed framework jointly addresses boundary-inflated proportions, discrete occupant counts, driver-passenger dependence, repeated observations over years, and prediction for non-sampled counties. Model assessment is organized into model selection and goodness-of-fit evaluation. We use WAIC to compare alternative year-effect structures, posterior predictive checks, and empirical coverage rates of posterior intervals to assess distributional fit, and comparisons with direct estimates to evaluate the practical effect of shrinkage. Bayesian model averaging is used to account for remaining model uncertainty among retained models.

The remainder of the paper is organized as follows. Section 2 describes the survey data structure, defines the county-by-year parameters of interest, and presents the Bayesian models for proportions and counts. Section 3 describes model selection, posterior predictive checking, Bayesian model averaging, and the prediction strategy for sampled and nonsampled counties. Section 4 presents the county-by-year estimates and compares the Bayesian predictions with direct estimates. Section 5 discusses the practical findings, limitations, and future extensions, including prior sensitivity analysis, spatial random effects, stronger predictive validation, and alternative count models when overdispersion is present.

2. Methods

This section presents the proposed Bayesian small area estimation framework in a structured order. We first define the county-by-year parameters of interest and present all candidate models considered in the study. We then describe the prior specification, posterior computation, and the criteria used for model selection and goodness-of-fit evaluation. Finally, we describe the Iowa Seat-Belt Use Survey data, the target population and sampling

frame, the auxiliary variables, the exploratory analysis, and the final prediction procedure for sampled and non-sampled counties.

2.1. Methodological framework and parameters of interest

The inferential target is seat-belt use for all 99 Iowa counties from 2017 to 2021. Let $i = 1, \dots, 99$ index counties, $t = 1, \dots, 5$ index survey years, and $j \in U_i$ index road segments in the population frame for county i . The observed survey data are available only for road segments from 15 sampled counties. Therefore, estimates for sampled counties use both observed survey data and model-based borrowing of strength, whereas estimates for non-sampled counties are model-based posterior predictions.

For road segment j in county i and year t , let $N_{d,tij}$ and $N_{p,tij}$ denote the observed numbers of drivers and passengers, respectively. Let $n_{d,tij}$ and $n_{p,tij}$ denote the corresponding numbers of belted drivers and belted passengers. When the denominators are positive, the observed driver and passenger seat-belt-use proportions are

$$y_{d,tij} = \frac{n_{d,tij}}{N_{d,tij}}, \quad (1)$$

$$y_{p,tij} = \frac{n_{p,tij}}{N_{p,tij}}. \quad (2)$$

Passenger proportions are undefined when no passengers are observed. These cases are excluded from the passenger proportion likelihood, while the passenger count information is retained in the count model.

Two types of county-by-year parameters are considered. The first is the vehicle-miles-traveled-weighted seat-belt-use proportion. Let v_{ij} denote vehicle miles traveled for road segment j in county i . The driver and passenger parameters are

$$v_{d,it} = \frac{\sum_{j \in U_i} v_{ij} y_{d,tij}}{\sum_{j \in U_i} v_{ij}}, \quad (3)$$

$$v_{p,it} = \frac{\sum_{j \in U_i} v_{ij} y_{p,tij}}{\sum_{j \in U_i} v_{ij}}. \quad (4)$$

The second type is the proportion of belted occupants. The driver, passenger, and total-occupant parameters are

$$\vartheta_{d,it} = \frac{\sum_{j \in U_i} N_{d,tij} y_{d,tij}}{\sum_{j \in U_i} N_{d,tij}}, \quad (5)$$

$$\vartheta_{p,it} = \frac{\sum_{j \in U_i} N_{p,tij} y_{p,tij}}{\sum_{j \in U_i} N_{p,tij}}. \quad (6)$$

These parameters are nonlinear functions of both seat-belt-use proportions and occupant counts. Therefore, the full framework combines models for proportions with models for counts and propagates uncertainty through both components.

2.2. Candidate models for seat-belt-use proportions

Let $g \in d, p$ denote driver or passenger. Five candidate models are considered for the seat-belt-use response. These models represent different ways of handling bounded proportions, one-inflation, and binomial count information.

For model m , the common linear predictor for the proportion or probability component is written as

$$\eta_{g,tij}^{(m)} = x'_{g,tij} \beta_g^{(m)} + b_{g,i}^{(m)} + b_{g,it}^{(m)} + b_{g,ij}^{(m)}, \quad (7)$$

where $x_{g,tij}$ contains the fixed-effect covariates, including log vehicle miles traveled, road type, and year-related terms when considered. The random effects $b_{g,i}^{(m)}$, $b_{g,it}^{(m)}$, and $b_{g,ij}^{(m)}$ represent county, county-by-year, and road-segment components, respectively. For each random-effect level, driver and passenger effects are allowed to be correlated:

$$b_\ell^{(m)} \sim \text{MVN}\left(0, \Sigma_\ell^{(m)}\right), \quad \ell \in i, it, ij. \quad (8)$$

2.2.1 Model 1: one-inflated beta model

The one-inflated beta model separates the probability of observing a boundary value equal to one from the beta-distributed continuous component. Define

$$w_{g,tij} = \begin{cases} 1, & y_{g,tij} = 1, \\ 0, & y_{g,tij} < 1. \end{cases} \quad (9)$$

For observations with $0 < y_{g,tij} < 1$, the beta component is

$$y_{g,tij} \mid w_{g,tij} = 0 \sim \text{Beta}\left(\mu_{g,tij}^{(1)} \phi_g^{(1)}, 1 - \mu_{g,tij}^{(1)} \phi_g^{(1)}\right), \quad (10)$$

and the one-inflation component is

$$w_{g,tij} \sim \text{Bernoulli}\left(p_{g,tij}^{(1)}\right). \quad (11)$$

The two link functions are

$$\text{logit}\left(\mu_{g,tij}^{(1)}\right) = x'_{g,tij} \beta_g^{(1)} + b_{g,i}^{(1)} + b_{g,it}^{(1)} + b_{g,ij}^{(1)}, \quad (12)$$

$$\text{logit}\left(p_{g,tij}^{(1)}\right) = z'_{g,tij} \alpha_g^{(1)} + u_{g,i}^{(1)} + u_{g,it}^{(1)} + u_{g,ij}^{(1)}. \quad (13)$$

This model extends the unit-level one-inflated beta model in Zhou and Berg (2024) by applying it to both drivers and passengers across multiple years.

2.2.2 Model 2: beta-epsilon model

The beta-epsilon model uses a small adjustment to move boundary values away from one. Let

$$y_{g,tij}^\epsilon = y_{g,tij} - \epsilon, \quad \epsilon = 0.001. \quad (14)$$

Then

$$y_{g,tij}^\epsilon \sim \text{Beta} \left(\mu_{g,tij}^{(2)} \phi_g^{(2)}, 1 - \mu_{g,tij}^{(2)} \phi_g^{(2)} \right), \quad (15)$$

$$\text{logit} \left(\mu_{g,tij}^{(2)} \right) = x'_{g,tij} \beta_g^{(2)} + b_{g,i}^{(2)} + b_{g,it}^{(2)} + b_{g,ij}^{(2)}. \quad (16)$$

This model handles proportions equal to one through a deterministic boundary adjustment rather than a separate one-inflation component.

2.2.3 Model 3: logit-normal-epsilon model

The logit-normal-epsilon model also uses the adjusted response $y_{g,tij}^\epsilon$, but models the transformed response on the logit scale:

$$\text{logit} (y_{g,tij}^\epsilon) \sim \text{Normal} \left(\mu_{g,tij}^{(3)}, \sigma_g^{2(3)} \right), \quad (17)$$

$$\mu_{g,tij}^{(3)} = x'_{g,tij} \beta_g^{(3)} + b_{g,i}^{(3)} + b_{g,it}^{(3)} + b_{g,ij}^{(3)}. \quad (18)$$

This model is included as a normal working model for the transformed proportions.

2.2.4 Model 4: beta-binomial model

The beta-binomial model uses the observed belted counts and denominators directly:

$$n_{g,tij} | p_{g,tij}^{(4)} \sim \text{Binomial} \left(N_{g,tij}, p_{g,tij}^{(4)} \right), \quad (19)$$

$$p_{g,tij}^{(4)} \sim \text{Beta} \left(\mu_{g,tij}^{(4)} \frac{1 - \rho_g^{(4)}}{\rho_g^{(4)}}, 1 - \mu_{g,tij}^{(4)} \frac{1 - \rho_g^{(4)}}{\rho_g^{(4)}} \right), \quad (20)$$

$$\text{logit} \left(\mu_{g,tij}^{(4)} \right) = x'_{g,tij} \beta_g^{(4)} + b_{g,i}^{(4)} + b_{g,it}^{(4)} + b_{g,ij}^{(4)}. \quad (21)$$

The parameter $\rho_g^{(4)}$ controls overdispersion relative to the binomial model. This specification is related to beta-binomial small area models discussed in Rao and Molina (2015) and applications such as Dong and Wakefield (2021).

2.2.5 Model 5: binomial model

The binomial model also uses the observed belted counts and denominators, but without the beta-binomial overdispersion layer:

$$n_{g,tij} \sim \text{Binomial} \left(N_{g,tij}, p_{g,tij}^{(5)} \right), \quad (22)$$

$$\text{logit} \left(p_{g,tij}^{(5)} \right) = x'_{g,tij} \beta_g^{(5)} + b_{g,i}^{(5)} + b_{g,it}^{(5)} + b_{g,ij}^{(5)}. \quad (23)$$

This model is the most direct count-based specification for the observed number of belted drivers or passengers.

2.3. Model for occupant counts

To estimate the proportion of belted occupants, we also model the numbers of drivers and passengers on each road segment. The count model is specified as a Poisson log-linear model:

$$N_{g,tij} \sim \text{Poisson}(\theta_{g,tij}), \quad g \in d, p, \quad (24)$$

$$\log(\theta_{g,tij}) = x'_{g,tij} \beta_{C,g} + c_{g,i} + c_{g,it} + c_{g,ij}. \quad (25)$$

The log link is used because it is the standard link for Poisson regression and ensures that the mean parameter $\theta_{g,tij}$ is positive. As in the proportion models, the driver and passenger random effects are allowed to be correlated at the county, county-by-year, and road-segment levels:

$$c_\ell \sim \text{MVN}(0, \Sigma_{C,\ell}), \quad \ell \in i, it, ij. \quad (26)$$

This count component is combined with the proportion component when estimating $\vartheta_{d,it}$, $\vartheta_{p,it}$, and $\vartheta_{o,it}$.

2.4. Model selection and goodness-of-fit evaluation

The evaluation procedure is organized into two parts: model selection and goodness-of-fit evaluation.

Model selection is based on the Widely Applicable Information Criterion (WAIC). WAIC is used to compare alternative year-effect structures, including fixed year effects, random year effects, no year effects, and random slopes for year-by-road-type combinations. The final year-effect structure is selected by considering both WAIC and model complexity.

Goodness-of-fit evaluation is based on posterior predictive p-values and empirical coverage rates of posterior intervals. For each candidate model, replicated data are generated from the posterior predictive distribution and compared with the observed data. We consider moment-based summaries, including the mean, median, standard deviation, skewness, and kurtosis, as well as Pearson-type discrepancy measures. Posterior predictive p-values close to 0 or 1 indicate a lack of fit. Because several models show poor posterior predictive performance for higher-order moments, we also use empirical coverage rates of 95%

posterior intervals as an additional goodness-of-fit diagnostic. These coverage rates are used for model checking and model comparison, not as a separate uncertainty-measurement procedure.

For the one-inflated beta model, a posterior draw can be written as

$$y_{g,tij}^{(l)} = \left(1 - w_{g,tij}^{(l)}\right) y_{g,tij}^{*,(l)} + w_{g,tij}^{(l)}, \quad (27)$$

where $y_{g,tij}^{*,(l)}$ is drawn from the beta component and $w_{g,tij}^{(l)}$ is drawn from the Bernoulli component. The posterior predictive mean and variance are

$$E\left(y_{g,tij}^{(l)}\right) = \left(1 - p_{g,tij}^{(l)}\right) \mu_{g,tij}^{*,(l)} + p_{g,tij}^{(l)}, \quad (28)$$

$$\text{Var}\left(y_{g,tij}^{(l)}\right) = \left(1 - p_{g,tij}^{(l)}\right) \frac{\mu_{g,tij}^{*,(l)} \left(1 - \mu_{g,tij}^{*,(l)}\right)}{\phi_g^{(l)} + 1} + \left(1 - p_{g,tij}^{(l)}\right) p_{g,tij}^{(l)} \left(1 - \mu_{g,tij}^{*,(l)}\right)^2. \quad (29)$$

The Pearson-type discrepancy is then computed using the posterior predictive mean and variance. For Models 2 and 3, posterior predictive checks are computed on the adjusted proportion scale or logit-adjusted scale. For Models 4 and 5, posterior predictive checks are computed using the belted counts $n_{g,tij}$ conditional on $N_{g,tij}$.

2.5. Bayesian model averaging

The posterior predictive checks and empirical coverage rates in Section 3.2 suggest that no single candidate model is uniformly best for both drivers and passengers. Therefore, Bayesian model averaging is used among the retained models. In the final implementation, Models 2 and 5 are retained because they provide more reasonable goodness-of-fit behavior than the other candidate models.

Let $\mathcal{M} = 2, 5$ denote the retained model set. For posterior draw l , let $\ell_m^{(l)}$ be the conditional log-likelihood of the observed data under model m . The model weight is computed as

$$\omega_m^{(l)} = \frac{\exp\left(\ell_m^{(l)}\right)}{\sum_{h \in \mathcal{M}} \exp\left(\ell_h^{(l)}\right)}, \quad m \in \mathcal{M}. \quad (30)$$

For the two retained models, this becomes

$$\omega_2^{(l)} = \frac{\exp\left(\ell_2^{(l)}\right)}{\exp\left(\ell_2^{(l)}\right) + \exp\left(\ell_5^{(l)}\right)}, \quad \omega_5^{(l)} = 1 - \omega_2^{(l)}. \quad (31)$$

For numerical stability, the weights are computed on the log-likelihood scale and then normalized. The final posterior draws of the small area parameters are obtained by combining the model-specific posterior draws using these weights.

2.6. Final prediction procedure for sampled and non-sampled counties

The final estimates are produced by combining the retained proportion models through Bayesian model averaging and then combining the resulting proportion predictions with the Poisson count predictions. For each posterior draw, the model generates predictions of $y_{d,tij}$, $y_{p,tij}$, $N_{d,tij}$, and $N_{p,tij}$ for road segments in the population frame. These predicted quantities are substituted into the definitions of $v_{d,it}$, $v_{p,it}$, $\vartheta_{d,it}$, $\vartheta_{p,it}$, and $\vartheta_{o,it}$.

For sampled counties, the prediction uses posterior draws of the fixed effects and available county-level random effects. Segment effects for nonsampled road segments are predicted from their posterior predictive distributions. For nonsampled counties, county-specific effects are not directly observed, so the estimates rely more heavily on fixed effects and the posterior predictive distribution of the county effects. Therefore, nonsampled-county estimates are model-based posterior predictions rather than direct design-based estimates.

3. Prior specification and selection

All candidate models are implemented in a Bayesian hierarchical framework. The prior distributions are specified separately for fixed effects, dispersion parameters, overdispersion parameters, and random-effect covariance matrices. For each model $m = 1, \dots, 5$ and each occupant group $g \in \{d, p\}$, the regression coefficients in the proportion component are assigned diffuse normal priors,

$$\beta_g^{(m)} \sim N(0, 10^4 I). \quad (32)$$

For the one-inflation component in Model 1, the fixed-effect coefficients are assigned the same type of diffuse normal prior,

$$\alpha_g^{(1)} \sim N(0, 10^4 I). \quad (33)$$

For the Poisson count model, the fixed-effect coefficients are also assigned diffuse normal priors,

$$\beta_{C,g} \sim N(0, 10^4 I). \quad (34)$$

For the beta precision parameters in the one-inflated beta model and the beta-epsilon model, weakly informative priors are assigned on the log scale to ensure positivity:

$$\log(\phi_g^{(1)}) \sim N(0, 10^2), \quad \log(\phi_g^{(2)}) \sim N(0, 10^2). \quad (35)$$

For the logit-normal-epsilon model, the residual precision is denoted by

$$\tau_g^{(3)} = \frac{1}{\sigma_g^{2(3)}}, \quad (36)$$

and the prior is specified as

$$\log \left(\tau_g^{(3)} \right) \sim N(0, 10^2). \quad (37)$$

For the beta-binomial model, the overdispersion parameter satisfies $0 < \rho_g^{(4)} < 1$. Therefore, a weakly informative prior is assigned on the logit scale:

$$\text{logit} \left(\rho_g^{(4)} \right) \sim N(0, 10^2). \quad (38)$$

The random effects are specified at three levels: county, county-by-year, and road segment. Let

$$\ell \in \{i, it, ij\} \quad (39)$$

denote the random-effect level. For the proportion component, the driver and passenger random effects are modeled jointly as

$$b_\ell^{(m)} = \begin{pmatrix} b_{d,\ell}^{(m)} \\ b_{p,\ell}^{(m)} \end{pmatrix} \sim \text{MVN} \left(0, \Sigma_{B,\ell}^{(m)} \right), \quad m = 1, \dots, 5. \quad (40)$$

For the one-inflation component in Model 1, the corresponding driver and passenger random effects are

$$\begin{pmatrix} b_\ell^{(1)} \\ u_\ell^{(1)} \end{pmatrix} = \begin{pmatrix} b_{d,\ell}^{(1)} \\ b_{p,\ell}^{(1)} \\ u_{d,\ell}^{(1)} \\ u_{p,\ell}^{(1)} \end{pmatrix} \sim \text{MVN} \left(0, \Sigma_{U,\ell}^{(1)} \right). \quad (41)$$

For the Poisson count model, the driver and passenger count random effects are modeled as

$$c_\ell = \begin{pmatrix} c_{d,\ell} \\ c_{p,\ell} \end{pmatrix} \sim \text{MVN} \left(0, \Sigma_{C,\ell} \right). \quad (42)$$

The covariance matrices are assigned inverse-Wishart-type priors. Equivalently, in the INLA implementation, the corresponding precision matrices are assigned Wishart priors. For each covariance matrix

$$\Sigma_{r,\ell}, \quad r \in \{B, C\}, \quad (43)$$

we define

$$\Omega_{r,\ell} = \Sigma_{r,\ell}^{-1}, \quad (44)$$

and assign

$$\Omega_{r,\ell} \sim \text{Wishart}_2(\nu, R), \quad \nu = 5, \quad R = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}. \quad (45)$$

For $\Sigma_{U,\ell}$, we can define $\Omega_{U,\ell} = \Sigma_{U,\ell}^{-1}$, and assign

$$\Omega_{U,\ell} \sim \text{Wishart}_4(\nu, R), \quad \nu = 11, \quad R = \begin{pmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 \\ 0 & 0 & 0.01 & 0 \\ 0 & 0 & 0 & 0.01 \end{pmatrix}. \quad (46)$$

This prior allows the two to four occupant-specific random effects to be correlated while keeping the covariance structure weakly informative. The use of this inverse-Wishart-type covariance prior is motivated by the prior sensitivity analysis in Section 3.1, where the inverse-Wishart specification better recovered the positive dependence between the driver and passenger random effects than the half-Cauchy alternatives.

Posterior inference is implemented using Integrated Nested Laplace Approximations (INLA) for latent Gaussian models (Rue et al., 2009). Posterior means are used as point estimates, and posterior standard deviations and 95% credible intervals are used to summarize uncertainty.

3.1. Prior sensitivity test

To evaluate the sensitivity of posterior inference to the prior specification for the county-level covariance structure, we compared three prior choices: a half-Cauchy prior with scale 2.5, a half-Cauchy prior with scale 1000, and an inverse-Wishart prior with hyperparameters (5, 0.01, 0.01, 0). The simulation study was based on the one-inflated beta regression model defined in Section 2.2.1, specifically equations (9) to (13). The beta precision parameter was fixed at $\phi = 42$. The true values of the random-effect parameters were chosen to match values estimated from the empirical data analysis, with

$$\sigma_b = 0.09164, \quad \sigma_m = 0.077, \quad \rho = 0.6. \quad (47)$$

Table 1 reports the prior sensitivity results on the standard deviation scale. The two half-Cauchy priors give nearly identical posterior summaries. Both half-Cauchy specifications overestimate the marginal standard deviations of the county-level random effects. For example, the posterior medians of σ_b and σ_m are about 0.355 and 0.539, respectively, compared with the true values 0.09164 and 0.077. In addition, the posterior median of the correlation is close to zero, and the posterior interval is wide, indicating that the half-Cauchy priors do not recover the positive dependence between the two county-level random-effect components.

Table 1. Prior sensitivity analysis for county-level standard deviations and correlation

Prior		True	Posterior median	95% posterior interval
Half-Cauchy(2.5)	σ_b	0.09164	0.3550	(0.2548, 0.5054)
	σ_m	0.0770	0.5390	(0.2814, 1.1581)
	ρ	0.6000	0.0000	(-0.6278, 0.6308)
Half-Cauchy(1000)	σ_b	0.09164	0.3550	(0.2548, 0.5054)
	σ_m	0.0770	0.5390	(0.2814, 1.1581)
	ρ	0.6000	0.0000	(-0.6278, 0.6308)
Inverse-Wishart	σ_b	0.09164	0.1734	(0.0698, 0.3824)
	σ_m	0.0770	0.0890	(0.0000, 0.5379)
	ρ	0.6000	0.8616	(-0.5171, 0.9971)

The inverse-Wishart prior gives a different pattern. The posterior medians of the standard deviations are closer to the true values, especially for the second component, and the posterior median of the correlation is positive. This suggests that the inverse-Wishart prior better preserves the positive dependence structure in the simulation. However, the posterior interval for ρ remains wide, and the posterior interval for σ_m is also wide. This indicates that the county-level covariance structure is still weakly identified.

Overall, the sensitivity analysis shows that the beta precision parameter is relatively stable across prior specifications, but the county-level variance and correlation parameters are sensitive to the covariance prior. We use the inverse-Wishart prior in the subsequent analysis because it better captures the positive dependence between the driver and passenger county-level random effects, which is important for the proposed multivariate small area model.

4. Survey data exploratory analysis

The Iowa Seat-Belt Use Survey is conducted annually to support estimates of seat-belt use. The data used in this study cover 2017–2021. A probability-proportional-to-size sample of road segments was selected from 15 Iowa counties, and the selected road segments were revisited across years. Although only 15 counties were directly sampled, the inferential target is all 99 Iowa counties. Therefore, estimates for the remaining 84 counties are model-based posterior predictions that rely on auxiliary variables and the fitted hierarchical model.

For each sampled road segment and year, observers recorded the number of drivers, passengers, and total occupants, as well as the corresponding numbers wearing seat belts. The data therefore contain both proportions and counts. The proportions are bounded between 0 and 1 and often equal to 1, motivating the one-inflated and epsilon-adjusted proportion models. The count outcomes are nonnegative integers, motivating the Poisson count model.

The main auxiliary variables are vehicle miles traveled, road type, year, county, and road-segment identifiers. Vehicle miles traveled is used both as a covariate in the regression models and as a weight in the definition of the vehicle-miles-traveled parameter. Road type is included because seat-belt-use patterns and traffic volume may differ across road classes.

Year is considered because the survey is longitudinal and because changes in seat-belt use over time are of interest for traffic-safety planning.

4.1. Exploratory analysis of seat-belt-use proportions

Following Zhou and Berg (2024), Figure 1 examines the relationship between the logit of observed seat-belt-use proportions and log vehicle miles traveled. The plots suggest an approximately linear association between the logit of the proportion and log vehicle miles traveled. They also suggest that the relationship may vary by road type and year. These patterns motivate the inclusion of log vehicle miles traveled, road type, and year-related terms in the candidate models.

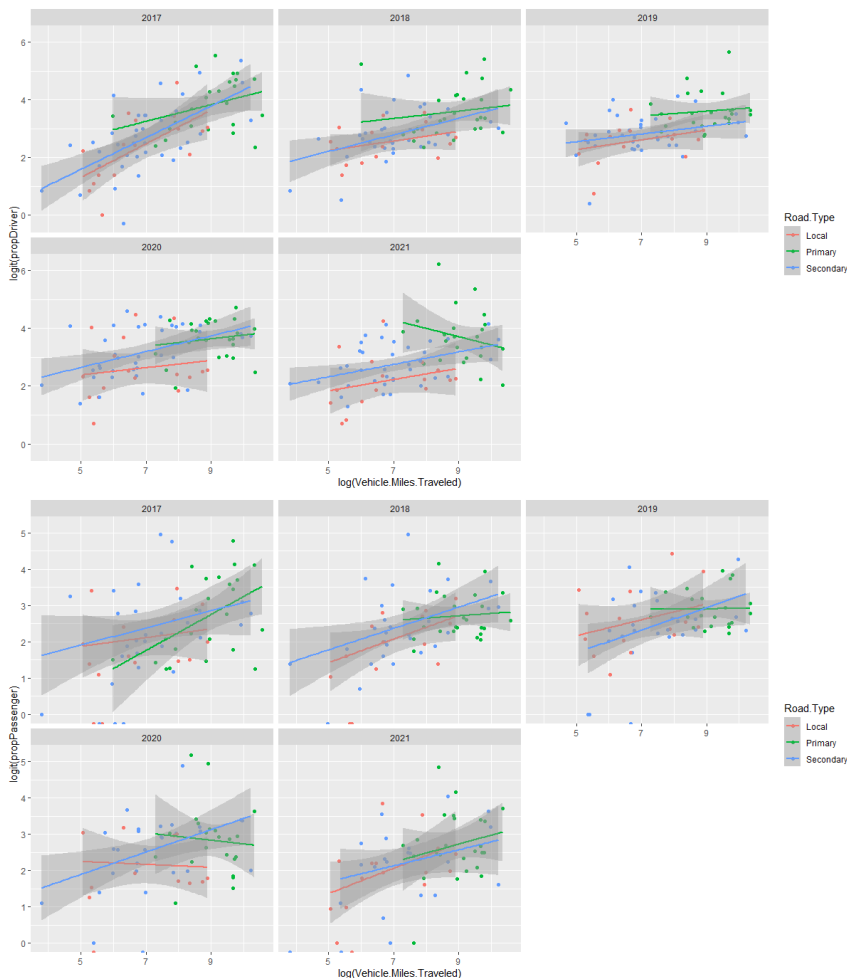


Figure 1. Scatter plots for the logit of proportions for drivers and passengers

4.2. Exploratory analysis of occupant counts

To assess whether a Poisson count model is reasonable, vehicle miles traveled were stratified into ten groups based on deciles. For each group, the sample mean and standard deviation of the count outcome were calculated. The counts were also grouped by road type. We then regressed the logarithm of the standard deviation on the logarithm of the group mean. Under a Poisson mean–variance relationship, this log–log regression has a slope close to 0.5.

Table 2. Box–Cox transformation regression: log standard deviation versus log mean

Grouping	Outcome	λ
Road type	Total driver count	0.4627
Road type	Total passenger count	0.5233
VMT category	Total driver count	0.4772
VMT category	Total passenger count	0.5957

As shown in Table 2, the estimated slopes are close to 0.5, suggesting that the observed mean–variance relationship is broadly consistent with a Poisson structure. Therefore, the Poisson log-linear model is used as the baseline count model.

5. Model evaluation results

5.1. Model selection for year effects

We first evaluate whether year effects should enter the model as fixed effects, random effects, both, or neither. This comparison is conducted using the binomial model, which is later shown to be one of the retained models.

Table 3. Posterior summaries for fixed year effects

	Year	Mean	SD	95% posterior interval
Drivers	2018	-0.133	0.154	(-0.438, 0.170)
	2019	-0.100	0.155	(-0.405, 0.204)
	2020	0.154	0.156	(-0.152, 0.461)
	2021	-0.186	0.155	(-0.491, 0.120)
Passengers	2018	-0.033	0.163	(-0.354, 0.286)
	2019	0.141	0.164	(-0.182, 0.464)
	2020	0.134	0.164	(-0.188, 0.457)
	2021	0.034	0.165	(-0.290, 0.358)

The posterior intervals in Table 3 contain zero for all fixed year effects, suggesting limited evidence of strong year-specific fixed effects. We then compare alternative year-effect structures using WAIC.

Table 4. WAIC for alternative year-effect structures.

Model	WAIC
Fixed year effects	3636.05
Random year effects	3632.49
No year effects	3634.34
Random slopes for years and road type	3630.16

Although the random-slope model has the smallest WAIC in Table 4, the differences among the candidate structures are small. Therefore, parsimony and interpretability are also considered. Figure 2 shows posterior density plots for the variance and correlation parameters associated with random year effects.

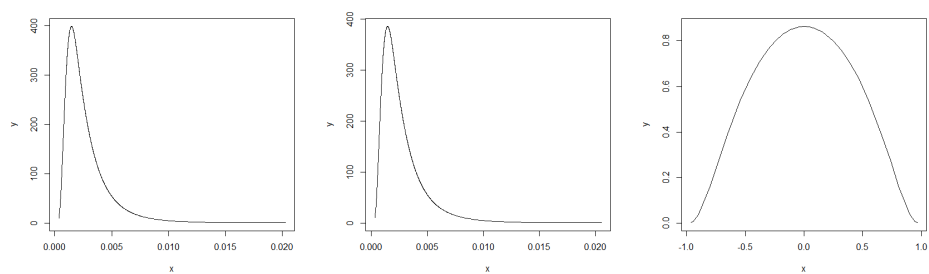


Figure 2. Posterior densities for σ^2 and ρ in the random year-effect model

The variance and correlation parameters in Figure 2 are nearly zero, indicating that the random year effects add little extra variation. Thus, the final model excludes the year effect for both fix and random effects as it may be an unnecessary random-effect component.

5.2. Model comparison

Posterior predictive p-values are used to compare the five candidate response models with random slopes for road types. The posterior p-values in Table 5 compare observed summaries with summaries generated from posterior predictive draws.

Table 5. Posterior predictive p-values for alternative models

Specification		Mean	Median	SD	Skewness	Kurtosis	Pearson test
Model 1	Driver	22.8%	22.6%	0.9%	99.9%	0.1%	45.5%
	Passenger	59.87%	12.47%	0.03%	100%	0%	0%
Model 2	Driver	57.8%	87.2%	3.07%	100%	0.67%	0.87%
	Passenger	50.5%	94.3%	8.2%	100%	0%	6.3%
Model 3	Driver	51.8%	99.7%	0%	0%	0%	0%
	Passenger	26.53%	52.63%	100%	0%	0%	100%
Model 4	Driver	78.2%	80.47%	62.47%	37.43%	36%	56.23%
	Passenger	99.87%	7.8%	56.3%	1.23%	2.7%	37.53%
Model 5	Driver	80.9%	60.03%	30.7%	2.7%	5.06%	35.43%
	Passenger	81.23%	11.4%	98.78%	100%	100%	99.17%

Table 5 suggests that no model is uniformly best across all summaries and both response groups. Model 1 works reasonably for some driver summaries but performs poorly for some passenger summaries. Models 2 and 5 provide more stable performance across both drivers and passengers.

Since the posterior p-values are not great for the higher order moments, we also evaluate the empirical coverage of 95% posterior intervals as a goodness-of-fit test. The results are shown in Table 6.

Table 6. Coverage rates of 95% posterior intervals

Specification		Coverage rate
Model 1	Driver	89.95%
	Passenger	73.94%
Model 2	Driver	95.45%
	Passenger	96.28%
Model 3	Driver	87.08%
	Passenger	53.10%
Model 4	Driver	88.51%
	Passenger	44.73%
Model 5	Driver	94.73%
	Passenger	97.27%

Models 2 and 5 have coverage rates closest to the nominal 95% level for both drivers and passengers. Therefore, these two models are retained for Bayesian model averaging.

Table 7. Summary of model checking and final model decision

Model	Response form	Checking result	Decision
M1	One-inflated beta	Reasonable for some driver summaries; weak for passenger summaries	Not retained
M2	Beta-epsilon	Reasonable PPCs; near-nominal coverage	Retained
M3	Logit-normal-epsilon	Poor PPCs for several summaries	Not retained
M4	Beta-binomial	Weak passenger coverage; unstable checks	Not retained
M5	Binomial	Reasonable PPCs; near-nominal coverage	Retained

Table 7 summarizes the model-checking evidence and the final model decision. The results indicate that no single candidate model is uniformly adequate for all summaries and both response groups. Model 1 captures some features of the driver proportions but performs weakly for passenger summaries. Model 3 has poor posterior predictive performance for several summaries, indicating that the logit-normal working model does not adequately reproduce the observed data. Model 4 allows extra-binomial variation, but its passenger coverage rate is weak, and its checking results are less stable. In contrast, Models 2 and 5 provide more balanced goodness-of-fit behavior across drivers and passengers, with reasonable posterior predictive checks and empirical coverage rates close to the nominal level. Therefore, these two models are retained and combined through Bayesian model averaging in the final estimation procedure.

5.3. Bayesian model averaging results

Figures 3 and 4 show two selected county examples comparing Models 2 and 5. Model 5 performs better than Model 2 in some counties, while Model 2 performs better in others. This supports the use of Bayesian model averaging rather than selecting a single retained model.

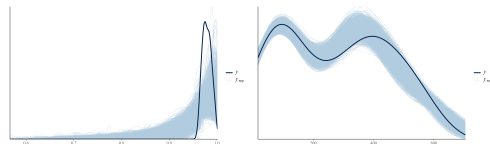


Figure 3. Selected county example where Model 5 performs better than Model 2

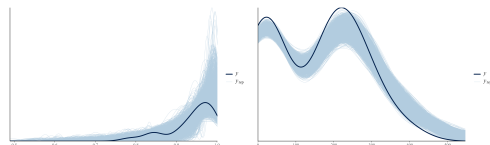


Figure 4. Selected county example where Model 2 performs better than Model 5

The posterior predictive checks indicate that no single candidate model is uniformly best for both drivers and passengers across all counties. Therefore, Bayesian model averaging is used among the retained models. In the final implementation, Models 2 and 5 are retained because they provide reasonable posterior predictive checks and credible interval behavior.

6. Results and Comparison

This section reports the final Bayesian small area estimates and compares them with direct estimates when available. The results are organized around the two parameters of interest: the proportion of belted vehicle miles traveled and the proportion of belted occupants. The supplement includes figures and tables for state-level estimations of the proportion of belted vehicle miles traveled, along with comparisons to direct estimations. Similarly, for the proportion of belted occupants, the supplement provides figures for county-level estimations, as well as state-level figures and tables, with comparisons to direct estimations. Additionally, a comparison between the two approaches is included in the supplement.

6.1. Proportion of vehicle miles traveled

For drivers and passengers, the vehicle-miles-traveled-weighted parameters $v_{d,it}$ and $v_{p,it}$ are estimated using posterior draws from the Bayesian model averaging procedure.

6.1.1 Posterior means and credible intervals

Figure 5 shows the posterior means and 95% credible intervals for the 15 sampled counties across years. All county results are included in the supplementary material. The county-level estimates show moderate year-to-year variation, but most credible intervals overlap substantially, suggesting limited evidence of strong county-specific temporal changes.

Figure 6 shows year-to-year differences based on VMT. The reference line at zero helps assess whether annual changes are meaningfully different from zero. Most intervals overlap zero, suggesting that annual fluctuations are generally within posterior uncertainty.

6.1.2 Comparison with direct estimates

Figure 7 in the supplement compares Bayesian estimates with direct estimates. The posterior means are generally close to the direct estimates, indicating that the model preserves the main empirical pattern in the observed data. The posterior standard deviations are often smaller than the direct-estimator standard deviations, reflecting the stabilizing effect of borrowing strength across counties, years, and road segments.

The reduction in posterior standard deviation should be interpreted as a bias–variance trade-off. The Bayesian model improves stability by shrinking noisy county-year estimates toward patterns explained by covariates and random effects. However, shrinkage can also introduce bias if important local variation is not captured by the model. Therefore, the comparison in Figure 7 should be interpreted as evidence of improved precision, not as proof that the Bayesian estimates are uniformly superior to direct estimates.

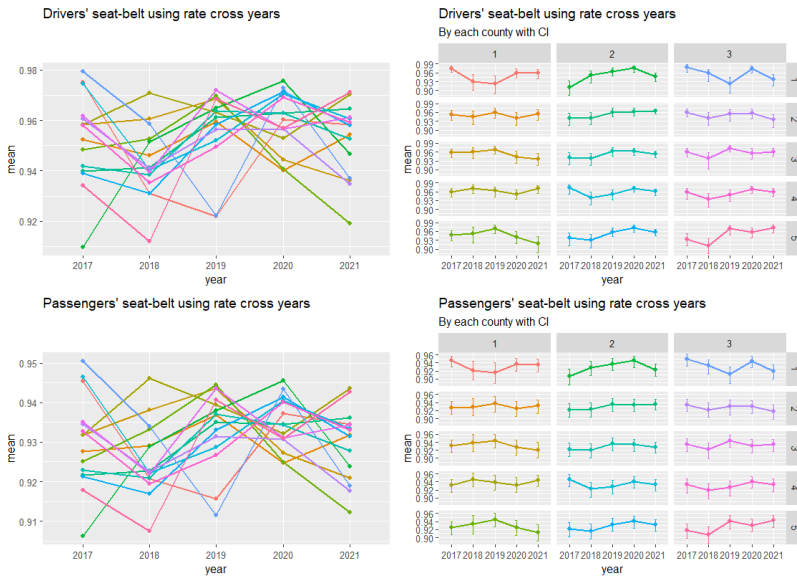


Figure 5. Posterior means and 95% credible intervals for 15 sampled counties across years

6.1.3 State-level estimation

The state-level posterior intervals for the vehicle-miles-traveled-weighted proportions are shown in Table 8 in the supplement. Driver estimates are generally higher and more stable than passenger estimates.

6.1.4 Estimation for all 99 counties

Heat maps for the lower bound, posterior mean, and upper bound of the 95% posterior intervals are provided in the Supplemental material. Moran's I tests are used to assess whether the county-level posterior means show spatial dependence. The results are shown in Table 9 in the supplement.

The Moran's I tests do not show strong evidence of spatial dependence for most years. Some borderline results may reflect spatial patterning in vehicle miles traveled or road characteristics, which are included as auxiliary information in the prediction model.

6.2. Proportion of belted occupants

The proportion of belted occupants is estimated by combining predictions from the proportion models with predictions from the Poisson count model. This procedure estimates $\vartheta_{d,it}$ and $\vartheta_{p,it}$.

Based on the counts, the changes in proportions over the years, the 95% posterior intervals, and comparisons against direct estimations, including potential outliers, are also included in the supplementary material.

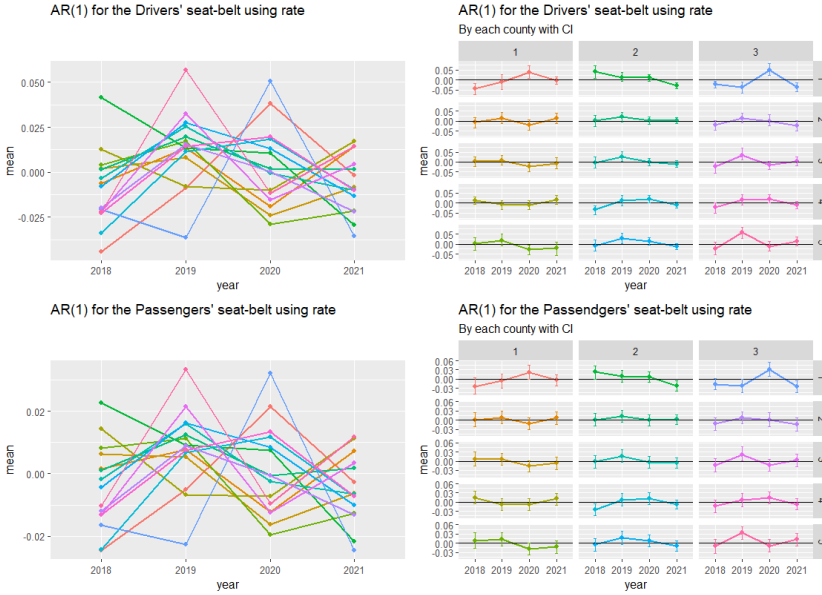


Figure 6. Posterior means and 95% credible intervals for year-to-year differences in 15 sampled counties

6.2.1 Estimation for all 99 counties

Moran's I tests for the belted-occupant county means are shown in Table 11 in the supplement.

The Moran's I results are similar to those from the vehicle-miles-traveled procedure. Overall, there is limited evidence of strong residual spatial dependence in the county-level posterior means, although spatial random effects remain a useful direction for future work.

7. Discussion and conclusion

This study developed an integrated Bayesian small area estimation framework for estimating county-by-year seat-belt use in Iowa from 2017 to 2021. The target of inference was all 99 Iowa counties, although the survey directly sampled road segments from only 15 counties. To address this sparse and complex data structure, we considered two related estimation procedures: one for the proportion of belted vehicle miles traveled and one for the proportion of belted vehicle occupants. The framework combines models for one-inflated seat-belt-use proportions with a Poisson log-linear model for occupant counts, incorporates correlated random effects, and uses Bayesian model averaging to account for model uncertainty among retained candidate models.

The main empirical finding is that the Bayesian estimates are generally close to the direct estimates in posterior mean, while often producing smaller posterior standard deviations. In many county-year domains, the hierarchical model reduces posterior uncertainty

by borrowing strength across counties, years, and road segments. This is practically useful because direct county estimates can be unstable when only a small number of road segments are observed. The proposed framework, therefore, provides a structured model-based approach for producing more stable local summaries of seat-belt use, including predictions for nonsampled counties where direct estimates are unavailable.

However, these gains should be interpreted carefully. A smaller posterior standard deviation does not automatically imply better estimation if the shrinkage introduces bias. Thus, the improvement from the Bayesian model should be viewed as a bias–variance trade-off rather than as uniform dominance over direct estimation. The model improves stability by shrinking noisy county-year estimates toward patterns explained by the covariates and random effects, but this shrinkage may also reduce local variation that is not fully captured by the model. For non-sampled counties, the estimates are model-based posterior predictions that depend on the assumed hierarchical structure and available auxiliary variables, rather than design-based direct estimates.

The model comparison results also show that no single response distribution performed best for all aspects of the data. The Iowa seat-belt survey contains several challenging features, including boundary values in the proportions, discrete occupant counts, repeated road-segment observations, and dependence between driver and passenger outcomes. Although the retained models provide reasonable summaries, they remain simplified representations of the underlying data-generating process. In particular, county-specific year trends are difficult to estimate precisely with only five annual waves and a limited number of sampled counties.

Another limitation concerns the passenger analysis. Observations with zero passengers were excluded from the passenger-use component. Therefore, passenger-related estimates should be interpreted as conditional on at least one passenger being present, rather than as unconditional estimates over all vehicles. If passenger presence is associated with road type, traffic volume, time of day, county characteristics, or seat-belt behavior itself, this conditioning may affect the passenger estimates and county-year comparisons.

Future work should extend the model in several directions. First, spatial random effects could be incorporated to borrow information across neighboring counties and to better reflect geographic dependence in traffic-safety behavior (Muchie et al., 2022). Second, stronger predictive validation, such as cross-validation, bootstrap comparisons, or design-based benchmarking, would help assess whether the reduction in posterior uncertainty is accompanied by acceptable bias. Third, prior sensitivity analysis requires improvement, particularly for beta-epsilon and Bernoulli regression. The current prior sensitivity test is simplistic and relies on a single simulation iteration for beta one-inflated regression. Finally, alternative count models, such as negative binomial, zero-inflated, or hurdle-type formulations, should be considered if over-dispersion or structural zeros are present in the occupant counts. These extensions would strengthen the reliability and interpretability of Bayesian small area estimates for future seat-belt surveys.

Supplementary Material

The supplementary material is available at <https://doi.org/10.5281/zenodo.21275327>.

References

- Aitkin, M., Liu, C. C., and Chadwick, T., (2009). Bayesian model comparison and model averaging for small-area estimation. *The Annals of Applied Statistics*, 3(1), pp. 199–221. doi:10.1214/08-AOAS205.
- Berg, E., (2022). Construction of databases for small area estimation. *Journal of Official Statistics*, 38, pp. 673–708.
- Berg, E., (2023). Small area prediction of seat-belt use rates using a Bayesian hierarchical unit-level Poisson model with multivariate random effects. *Stat*, 12, e544.
- Dong, T. Q., Wakefield, J., (2021). Modeling and presentation of vaccination coverage estimates using data from household surveys. *Vaccine*, 39, pp. 2584–2594.
- González-Manteiga, W., M. J. Lombardía, I. Molina, D. Morales, and Santamaría, L., (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational Statistics & Data Analysis*, 51(5), pp. 2720–2733.
- Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T., (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4), pp. 382–417. doi:10.1214/ss/1009212519.
- Lyu, X., E. J. Berg, and Hofmann, H., (2020). Empirical Bayes small area prediction under a zero-inflated lognormal model with correlated random area effects. *Biometrical Journal*, 62, pp. 1859–1878.
- McGovern, A., Wilson, K., and Wakefield, J., (2026). Direct-assisted Bayesian unit-level modeling for small area estimation of rare event prevalence. *Journal of Survey Statistics and Methodology*, 14(1), pp. 178–208. doi:10.1093/jssam/smaf037.
- Mohebbi, Mohammadreza, Rory Wolfe, and Forbes A., (2014). Disease mapping and regression with count data in the presence of overdispersion and spatial autocorrelation: a Bayesian model averaging approach. *International journal of environmental research and public health*, 11, No. 1, pp. 883–902.
- Molina, I., B. Nandram and Rao, J. N. K., (2014). Small area estimation of general parameters with application to poverty indicators: A hierarchical Bayes approach. *The Annals of Applied Statistics*, 8(2), pp. 852–885.
- Muchie, K. F., A. K. Wanjoya and Mwalili, S. M., (2022). On hierarchical Bayesian spatial small area model for binary data under spatial misalignment. *Journal of Probability and Statistics*, 3865626.

- Oleson, J. J., C. He, D. Sun and Sheriff, S., (2007). Bayesian estimation in small areas when the sampling design strata differ from the study domains. *Survey Methodology*, 33, pp. 173–185.
- Parker, P. A., Janicki, R. and Holan, S. H., (2023a). A comprehensive overview of unit-level modeling of survey data for small area estimation under informative sampling. *Journal of Survey Statistics and Methodology*, 11(4), pp. 829–857. doi:10.1093/jssam/smad020.
- Parker, P. A., Janicki, R. and Holan, S. H., (2023b). Comparison of unit-level small area estimation modeling approaches for survey data under informative sampling. *Journal of Survey Statistics and Methodology*. doi:10.1093/jssam/smad022.
- Rao, J. N. K., Molina, I. , (2015). *Small Area Estimation*. Hoboken, NJ: John Wiley & Sons.
- Rue, H., S. Martino and Chopin, N., (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* 71, pp.319–392.
- Sugasawa, S., Kobayashi, G., and Kawakubo, Y., (2024). Bayesian benchmarking small area estimation via entropic tilting. arXiv:2407.17848.
- Vedensky, D., Parker, P. A. and Holan, S. H., (2025a). Bayesian unit-level models for longitudinal survey data under informative sampling: An analysis of expected job loss using the Household Pulse Survey. *Journal of the Royal Statistical Society: Series A, Statistics in Society*.
- Vedensky, D., Parker, P. A. and Holan, S. H., (2025b). Bayesian unit-level modeling of categorical survey data with a longitudinal design. arXiv:2502.15112.
- Wu, S. M., Stephenson, B. J. K., (2023). Bayesian estimation methods for survey data with potential applications to health disparities research. *WIREs Computational Statistics*, 16(1), e1633. doi:10.1002/wics.1633.
- Zhou, Z., Berg, E., (2024). A unit-level one-inflated beta model for small area prediction of seat-belt use rates. *Journal of Applied Statistics*, pp. 1–24.

