Bayesian criterion-based assessments of recurrent event models with applications to commercial truck driver behavior studies

Yiming Zhang¹ | Ming-Hui Chen¹ | Feng Guo²

¹Department of Statistics, University of Connecticut, Storrs, Connecticut, ²Department of Statistics, Virginia Tech, Blacksburg, Virginia.

Correspondence
Feng Guo, Department of Statistics, Virginia Tech, Blacksburg, VA, USA.
Email: feng.guo@vt.edu

Summary
Multitype recurrent events are commonly observed in transportation studies, since commercial truck drivers may encounter different types of safety critical events (SCEs) and take different lengths of on-duty breaks in a driving shift. Bayesian nonhomogeneous Poisson process models are a flexible approach to jointly model the intensity functions of the multitype recurrent events. For evaluating and comparing these models, the deviance information criterion (DIC) and the logarithm of the pseudo-marginal likelihood (LPML) are studied and Monte Carlo methods are developed for computing these model assessment measures. We also propose a set of new concordance indices (C-indices) to evaluate various discrimination abilities of a Bayesian multitype recurrent event model. Specifically, the within-event C-index quantifies adequacy of a given model in fitting the recurrent event data for each type, the between-event C-index provides an assessment of the model fit between two types of recurrent events, and the overall C-index measures the model’s discrimination ability among multiple types of recurrent events simultaneously. Moreover, we jointly model the incidence of SCEs and on-duty breaks with driving behaviors using a Bayesian Poisson process model with time-varying coefficients and time-dependent covariates. An in-depth analysis of a real dataset from the commercial truck driver naturalistic driving study is carried out to demonstrate the usefulness and applicability of the proposed methodology.

KEYWORDS
Bayesian model assessment, concordance index, multitype recurrent event, penalized spline, truck driving safety

1 | INTRODUCTION

Commercial truck crashes are a major traffic safety concern, as there are more than 4000 fatalities and more than 100 000 injuries annually in the United States. Among many crash risk factors, driver-related factors are the largest contributor, bearing responsibility for the majority of crashes.¹ Commercial truck drivers are professionals with revenue-generating...
driving tasks that require extended working and driving hours. The fatigue and health issues associated with long service hours are a major threat to safety and well-being of truck drivers. Understanding drivers’ sleeping patterns, driving and working schedules, and rest breaks can provide valuable insight into their behaviors, help in developing safety countermeasures, and support safety regulations.

As professional drivers, commercial truck drivers are subject to the operation and safety management of the fleet they serve. Their schedules are influenced by the demand of clients and fleet operation characteristics. In addition, Hours-of-Service (HOS) rules set by the Federal Motor Carrier Safety Administration limit the total working hours, total driving time, as well as nondriving breaks per working shift. For example, the current HOS rules limit the total number of working hours per shift to 14 hours and total driving hours to 11 hours. In addition, drivers should take a 34-hour off-duty period (with two 1:00 AM to 5:00 AM periods) after 60 on-duty driving hours in 7 consecutive days or 70 hours in 8 consecutive days. The HOS rules are constantly revised with the advancement in traffic safety research.

Substantial research has been conducted to evaluate truck driver schedules, sleeping, and safety performance. Prolonged working and driving hours within a shift could lead to driving fatigue, which is a major driving risk factor. Total sleeping time prior to a working shift has been shown to have a direct impact on driving risk. Besides the total sleep duration, other characteristics, such as sleep in the early stage of the non-work period, and less sleep between 1:00 AM and 5:00 AM, also influence driving risk. Blanco et al evaluated the change in driving risk over the 11-hour on-duty driving period using a large-scale naturalistic driving study (NDS) and showed that the risk in the 11th hour is significantly higher than in the first 2 hours of driving. The study also found that safety critical events (SCEs) increased over the 14-hour working period. Using a different data source and study design, Chen and Xie confirmed elevated crash risk at the 11th hour of driving. Driver adaption and countermeasures could mitigate adverse effects of fatigue, including sleeping more prior to a long trip, taking rest breaks, and consuming caffeine. For example, rest breaks during long driving session could reduce driving risk.

Modeling the truck driving data is challenging due to the complex nature of driving behaviors and safety events. Various types of SCEs, on-duty and off-duty breaks, and nonworking periods occur within a driving shift. Driving and rest behaviors have both direct and indirect effects on the risk of SCEs. In addition, the effect of a covariate may not be constant over time. For instance, the impact of off-duty sleeping time before a driving task might change over a driving shift due to increasing fatigue over time. Thus, recurrent event models with time-varying coefficients are preferred to model complex transportation data. The Bayesian joint modeling approach for multitype recurrent event data has been developed in the recent literature due to its flexibility to handle the complicated data structures.

Model assessment and diagnosis are the two important components in Bayesian inference. Commonly used approaches, such as the deviance information criterion (DIC), the conditional predictive ordinate (CPO), and the logarithm of the pseudo-marginal likelihood (LPML) have been applied to Bayesian recurrent event models. However, high dimensionality of the random effects in the multitype recurrent event models often leads to a high cost in computing model assessment criteria. There is a need to develop reliable and computationally feasible model assessment tools for Bayesian multitype recurrent event models. The discrimination ability is another essential aspect to assess the model fit. A frequently used approach, especially popular in survival analysis, is the concordance index (C-index). However, discrimination ability is not broadly discussed for recurrent event models. To the best of our knowledge, only Kim et al developed a C-index for the proportional rate/mean model for recurrent event data within the frequentist framework, which is estimated by the proportion of pairs where the risk prediction scores and the observed event counts are concordant.

In this paper, we jointly model the incidence of SCEs and on-duty breaks using a Bayesian time-varying coefficient model for multitype recurrent event data to account for the dependent structure between SCEs and on-duty breaks. In order to investigate time-variant driving behaviors such as driving between 1:00 AM and 5:00 AM, we extend the model proposed by Liu and Guo to incorporate time-dependent covariates. For Bayesian multitype recurrent event models, we introduce the type-specific DIC to assess the goodness-of-fit of the model to the data for each recurrent event type. The proposed type-specific DICs are attractive in the sense that (i) the sum of the type-specific DICs reduces to the overall DIC, which can be used to evaluate the model fit for the whole data, and (ii) they are easily computed using readily available Markov chain Monte Carlo (MCMC) samples from the joint posterior distribution without resorting any analytically intractable integrals. We also consider the LPML for model comparison since it is based on the predictive distribution of the recurrent event data after integrating out of the random effects, which captures the predictive aspect of the model. In addition, an easy-to-implement algorithm of the Monte Carlo method based on the CPO Identity is developed for computing the LPML for the complex multitype recurrent event models. Three types of C-indices, within-event C-index,
between-event C-index, and overall C-index are proposed to quantify discrimination abilities of Bayesian multitype recurrent event models. Each proposed C-index is a function of the model parameters including the random effects, which is computed at each MCMC sample. Thus, the posterior distributions of the C-indices are easily obtained as a byproduct of MCMC sampling.

The remainder of the paper is organized as follows. Section 2 introduces the data structure of the commercial truck driver NDS study, the types of recurrent events, and the potentially useful time-independent and time-varying predictors. The formulation of the Bayesian multitype recurrent event model with time-varying coefficients and time-dependent covariates, the prior specification, and the resulting posterior are given in Section 3. Section 4 presents the detailed development of the Bayesian model assessment criteria and corresponding computational algorithm. In the same section, we propose three types of C-indices for the Bayesian multitype recurrent event models. Simulation studies are presented in Section 5 and an in-depth analysis of the NDS data is carried out in Section 6. Concluding remarks and discussions are provided in Section 7.

2 COMMERCIAL TRUCK DRIVER NDS

The commercial truck driver NDS is a large-scale prospective cohort study with the primary goal of evaluating truck driver safety.7 A total of 100 commercial truck drivers each drove an instrumented truck for a 4-week period (four drivers were excluded due to missing data issues). Each truck was instrumented with an advanced data acquisition system. The system collected dozens of key driving parameters, and included multichannel cameras, three-dimensional accelerometers, GPS, radar, etc. The continuous driving data were collected from ignition-on to ignition-off at high frequency (eg 10 Hz for videos and accelerometers). The objectively collected NDS data have provided opportunities to evaluate various risk factors, particularly driver behavior factor due to the comprehensive information collected.20

In addition to the driving data, the study also included a detailed daily activity register. The data include 15 on-duty and off-duty tasks, such as driving, working, eating, sleep, leisure activities, etc. The participants were asked to record the time of each activity. The activity register data were compared with the objectively recorded driving data to create a set of hybrid activity data containing the metrics of the register data’s completeness and the driving data’s accuracy.

The data for this study were comprised of 1848 on-duty driving shifts from the 96 drivers, among whom 74 (77.1%) worked primarily on long-haul tasks and 21 (22.9%) were line-haul drivers. There were 91 (94.8%) male and 5 (5.2%) female drivers with an average age of 44 years, ranging from 21 to 73 years. Drivers’ experience in driving commercial vehicles varied from 4 weeks to 54 years. A maximum of 11 hours of driving time or a maximum of 14 hours total on-duty time for each driving shift were observed due to the HOS guidelines set by the Federal Motor Carrier Safety Administration.

2.1 Multitype recurrent events

The SCEs and on-duty breaks were considered as recurrent events that took place during and up to 11 driving hours for each driving shift. The NDS study collected four types of SCEs from the continuous driving data, namely, unintentional lane deviations (ULD), crashes, near-crashes, and crash-relevant conflict (CRC) events. These events were identified by an initial screen of kinematic signatures and were confirmed by video recordings. Due to the small number of crash and near-crash events, these were grouped with CRC events to avoid rare event issues.9 As shown in Figure 1, the duration of on-duty breaks had a median length of 45 minutes with a range from 0.01 to 4 hours. Furthermore, the on-duty breaks were dichotomized into short breaks (SB) and long breaks (LB) based on the median break duration. Thus, two types of SCEs (ULD and CRC) and two types of on-duty break (SB and LB) events were considered in the real data analysis. In total, we observed 1124 ULD events from 343 shifts, 1112 CRC events from 548 shifts, 474 SB events from 383 shifts and 436 LB events from 367 shifts.

The incidence rates of the four types of recurrent events at each driving hour are displayed in Figure 2. The SCE rates were generally higher than the on-duty break rates. Both ULD and CRC event rates climbed from the start of driving shifts to the seventh driving hour. After a slight drop, the SCE rates increased again at the last two driving hours. The event rate curves for the on-duty breaks were more flat compared to the SCEs. Particularly, the LB event rate reached its highest at the sixth hour and decreased in the last several driving hours. This likely indicates that drivers might not take a LB toward the later driving hours since this would be close to the end of a driving shift.
2.2 Time-independent variables

One of the primary interests is to investigate if taking a compensatory long off-duty rest after an intensive on-duty period helps commercial truck drivers improve their driving performance. In the NDS data, we identified whether drivers needed a 34-hour off-duty break before a driving shift by checking to see if they drove for at least 60 hours in a 7-day rolling window or at least 70 hours in an 8-day rolling window before the end of the previous driving shift. If a driver did not meet the criterion, we considered the corresponding driving shift to be “no need to restart.” Otherwise, the corresponding
driving shift was grouped to “restarting driving” or “not restarting driving” based on if the driver took an off-duty break for at least 34 hours. The 1848 driving shifts contained 1323 (71.6%) “no need to restart” shifts, 80 (4.3%) “restarting driving” shifts and 445 (24.1%) “not restarting driving” shifts. For the 525 driving shifts after a long concentrated working period, the drivers in 84.8% of those shifts did not take an off-duty break for at least 34 hours, which indicates a low compliance rate for this HOS recommendation.

The off-duty sleeping time, the number of nondriving tasks, the shift starting time, and the driver type were also considered as time-independent predictors for influencing the incidence rates of SCEs and on-duty breaks. The off-duty sleeping time before a driving shift was categorized into three groups: normal sleeping (7-9 hours), insufficient sleeping (less than 7 hours), and abundant sleeping (more than 9 hours). This study consisted of 1112 (60.2%) shifts in the normal sleeping group, 271 (14.7%) shifts in the insufficient sleeping group, and 465 (25.2%) shifts in the abundant sleeping group. It has been shown that the off-duty sleeping groups have different SCE risk patterns over the 11-hour driving time.9,10 The number of nondriving tasks had a mean of 2.86 (SD = 2.08) and a range of frequency from 0 to 13. We considered three shift starting time groups: 667 (36.1%) shifts started in the morning (5:00 AM-9:00 AM), 768 (41.6%) shifts started in the daytime (9:00 AM-8:00 PM), and 413 (22.3%) shifts started at night (8:00 PM-5:00 AM).

Table 1 presents the descriptive statistics of the driving behaviors across the three restarting groups. The drivers in the “not restarting driving” group reported a higher chance of getting abundant off-duty sleep before a driving shift, and a lower percentage of starting driving at night compared to the other two groups. This shows that the drivers who did not take a long off-duty break to recover from an intensive driving period may need a longer off-duty sleep duration to compensate for resting and avoid tiring late-night driving shift. In addition, most of the driving shifts (99.2%) following with a long consecutive on-duty period were from the long-haul drivers, which explains that those shifts had a longer shift duration and fewer number of nondriving tasks.

### Table 1  Descriptive table of shift-level characteristics in the three restarting groups

| Shift-level characteristics | No need to restart | Restarting driving | Not restarting driving | P-value $^a$ |
|-----------------------------|--------------------|--------------------|------------------------|-------------|
| **Length of shift**         | 7.61 (2.52)        | 8.37 (2.47)        | 8.05 (2.69)            | < 0.001     |
| **Number of nondriving work** | 3.04 (2.19)    | 2.85 (1.93)        | 2.33 (1.65)            | < 0.001     |
| **Sleeping Group**          |                    |                    |                        |             |
| Insufficient                | 235 (17.8%)        | 10 (12.5%)         | 26 (5.8%)              | < 0.001     |
| Normal                      | 782 (59.1%)        | 50 (62.5%)         | 280 (62.9%)            |             |
| Abundant                    | 306 (23.1%)        | 20 (25.0%)         | 139 (31.2%)            |             |
| **Starting Time**           |                    |                    |                        |             |
| Morning                     | 435 (32.9%)        | 25 (31.2%)         | 207 (46.5%)            | < 0.001     |
| Daytime                     | 546 (41.3%)        | 38 (47.5%)         | 184 (41.3%)            |             |
| Night                       | 342 (25.9%)        | 17 (21.2%)         | 54 (12.1%)             |             |
| **Driver Type**             |                    |                    |                        |             |
| Line-haul                   | 399 (30.2%)        | 3 (3.8%)           | 1 (0.2%)               | < 0.001     |
| Long-haul                   | 924 (69.8%)        | 77 (96.3%)         | 444 (99.8%)            |             |

$^a$ P-values are from the Kruskal-Wallis test for continuous variables and the Pearson Chi-square/Fisher’s exact test for categorical variables.

#### 2.3 Time-dependent variables

The local times of the driving activities and the occurrence of any events were also collected in the NDS data. Driving at 1:00 AM to 5:00 AM is coded as a binary time-dependent covariate in the 11 driving hours. Among all the observed driving shifts, 547 (29.6%) shifts overlapped with 1:00 AM to 5:00 AM, with an average late-night driving time of 2.54 hours per driving shift.
3 | BAYESIAN MODELING AND ESTIMATION OF MULTITYPE RECURRENT EVENT DATA

3.1 | Model specification

We jointly model the intensity functions of different types of SCEs and on-duty break events using a multitype Poisson process model with both fixed and time-varying coefficients following Liu and Guo. \(^9\) For the \(i\) th driving shift, \(i = 1, \ldots, n\), let \(X_i = (x_{i1}, \ldots, x_{iq})'\), \(Z_i = (z_{i1}, \ldots, z_{iq})'\), and \(W_i(t) = (w_{i1}(t), \ldots, w_{ik}(t))'\) represent the shift-level predictors with fixed effect, shift-level predictors with time-varying effect, and time-dependent covariates at time \(t\), respectively. Conditioning on the random effect \(b_{ij}\), we assume that the intensity function for the \(i\) th driving shift and the \(j\) th event type is given by

\[
\lambda_{ij}(t|X_i, W_i(t), Z_i, b_{ij}) = \exp\left\{ \alpha_j'X_i + \gamma_j'W_i(t) + \beta_{ij}(t) + \cdots + z_{iq}\beta_{qj}(t) + b_{ij} \right\},
\]

where \(\alpha_j\) and \(\gamma_j\) are two vectors of the linear regression coefficients for the time-independent and time-dependent covariates, respectively; \(\beta_{ij}(t)\) is a smooth function for the log baseline intensity of the \(j\) th event type; \(\beta_{ij}(t)\) are the time-varying coefficients of the covariates in \(Z_i\); \(b_{ij}\) is a shift-level random effect for \(j = 1, \ldots, J\). We further assume that \(b_i = (b_{i1}, \ldots, b_{ij})'\) follows a multivariate normal distribution with mean \((0, \ldots, 0)'\) and covariance \(\Sigma\). By incorporating the log-normal frailties with the dependence structure \(\Sigma\), the model is able to capture the variability within each type of the recurrent events and the association between two types of events.

The baseline intensity function and the time-varying coefficients, \(\beta_{ij}(t), l = 0, \ldots, q, j = 1, \ldots, J\), are unknown smooth functions, which are modeled by polynomial splines with equally spaced knots on the 11 driving hours. We adopt the B-splines with degree of \(v\), which are widely used in recurrent event models with nonparametric intensity functions.\(^9,10,12\)

Let \(h_i\) denote the number of knots and \(B_{ih}(t)\) represent the value of the \(h\) th B-spline at \(t\), where \(h = 1, \ldots, H_i\) and \(H_i = h_i + v\). Then, the smooth function \(\beta_{ij}(t)\) is expressed by a linear combination of the B-spline basis functions \(B_{ih}(t)\) given by \(\beta_{ij}(t) = \sum_{h=1}^{H_i} B_{ih}(t)\beta_{ijh}\), where \(B_{ih}(t) = (B_{i1h}(t), \ldots, B_{ijh}(t))'\) and \(\beta_{ijh} = (\beta_{i1h}, \ldots, \beta_{ijh})'\) is a vector of the corresponding regression coefficients.

For the \(i\) th driving shift, let \(t_{ijk}\) be the event time of the \(k\) th type \(j\) event, let \(n_{ij}\) denote the number of type \(j\) events occurring before the follow-up time \(c_i\), and \(D_i = \{X_i, Z_i, W_i, c_i\}\) stands for the observed shift-level data. Let \(t_{ij} = \{t_{ij1}, \ldots, t_{ijn_{ij}}\}\) and \(\Omega_j = \{\alpha_j, \gamma_j, \beta_{ij}, l = 0, \ldots, q\}\) represent the sets of the \(j\) th type-specific event time and parameters, respectively. Write \(t_i = \{t_{i1}, \ldots, t_{ij}\}\) and \(\Omega = \{\Omega_1, \ldots, \Omega_J\}\). The conditional likelihood given the random effects \(b_i\) for the \(i\) th driving shift is written as

\[
L_i(\Omega_i | b_i, t_i, D_i) = \prod_{j=1}^{J} L_{ij}(\Omega_j | b_{ij}, t_{ij}, D_i),
\]

where the type-specific conditional likelihood is defined as

\[
L_{ij}(\Omega_j | b_{ij}, t_{ij}, D_i) = \prod_{k=1}^{n_{ij}} \lambda_{ij}(t_{ijk}|X_i, W_i(t_{ijk}), Z_i, b_{ij}) \exp(-\Lambda_{ij}(c_i)),
\]

where \(\Lambda_{ij}(c_i) = \int_0^{c_i} \lambda_{ij}(t|X_i, W_i(t), Z_i, b_{ij})dt\) and \(\lambda_{ij}(t_{ijk}|X_i, W_i(t_{ijk}), Z_i, b_{ij})\) is defined in (1). We note that \(\Lambda_{ij}(c_i)\) does not have a closed form due to the time-varying coefficients, so we adopt a trapezoid method for obtaining an approximation of \(\Lambda_{ij}(c_i)\) following Liu and Guo.\(^9\)

3.2 | Prior and posterior

Controlling the smoothness of the spline function is crucial in the time-varying coefficient models. The smoothness is essentially controlled by the number of knots of basis splines and the order of penalty of smoothing. To allow for a flexible fit, the number of knots should be large enough, which may introduce an over-fitting problem. The penalized B-spline (P-spline), which uses a relatively large number of knots and a penalty of smoothing on the objective function to control the smoothness of the fit, has been widely studied in the literature.\(^21,22\) Following Liu and Guo\(^9\) and Lang and Brezger,\(^22\)
We specify the shrinkage priors for the coefficients of the B-splines given by
\[
\beta_{ij} | \tau_i \propto \exp \left\{ - \frac{\tau_i}{2} \beta_{ij}' (D_i'(D_i)^{-1}) \beta_{ij} \right\}, \tag{4}
\]
where \(D_i\) is a proper matrix such that \(D_i\beta_{ij} = (\Delta' \beta_{i+1,j}, \ldots, \Delta' \beta_{K_i,j})'\) with rank \(K_i - r\). \(\Delta\) denotes the difference between the coefficients of adjacent B-splines and \(r\) denote a finite order of difference. For instance, the first and second order differences of \(\beta_{ij}\) are given by \(\Delta \beta_{ij} = \beta_{ij} - \beta_{i-1,j}\) and \(\Delta^2 \beta_{ij} = \Delta \beta_{ij} = \beta_{ij} - 2\beta_{i-1,j} + \beta_{i-2,j}\), respectively. The roughness penalty parameter \(\tau_i\) is considered as random and it is estimated jointly with other unknown parameters. A Gamma prior is assigned to \(\tau_i\) such that \(\tau_i \sim \text{Gamma}(a_\tau, b_\tau)\), where \(a_\tau\) is the shape parameter and \(b_\tau\) is a small rate parameter of the Gamma distribution. However, the choice of \(b_\tau\) highly impacts the smoothness of the Bayesian P-spline.23
Thus, we reparameterize this parameter as \(b_\tau = \delta_i a_\tau\) and then assign a Gamma hyperprior to \(\delta_i\). Specifically, we specify \(\tau_i \sim \text{Gamma}(a_\tau, \delta_i a_\tau)\) and \(\delta_i \sim \text{Gamma}(a_\delta, b_\delta)\). In Section 5 below, we specify \(a_\tau = 1\), \(a_\delta = 0.01\) and \(b_\delta = 0.01\), which lead to a relatively non informative prior.

Since there is no prior information available for the analysis, a noninformative prior \(N(0, 10^3)\) is assigned to each of the time-independent coefficients \(\alpha_j\) and \(\gamma_j\). With \(\mathbf{\tilde{O}} = (\Omega, \Sigma, \tau_i, \delta_i, 1 \leq j \leq J, 0 \leq l \leq q)\) denoting all the parameters and hyperparameters in the model and priors, the joint distribution is written as
\[
p(\mathbf{\tilde{O}} | \mathbf{t}, \mathbf{D}_i, 1 \leq i \leq n) \propto \prod_{i=1}^{n} L_i(\Omega | \mathbf{b}_i, \mathbf{D}_i) p(\mathbf{b}_i | \Sigma) \times \prod_{j=1}^{J} \prod_{l=0}^{q} p(\beta_{ij} | \tau_i) p(\tau_i | \delta_i) p(\delta_i) \prod_{j=1}^{J} p(\alpha_j) p(\gamma_j) p(\Sigma). \tag{5}
\]
To achieve better convergence of MCMC sampling, the Cholesky decomposition is widely applied in modeling the covariance matrix of the random effects for longitudinal data24,25 and multi-type recurrent event data.9,12,13 Let \(\Sigma = \Phi \Phi'\), where \(\Phi\) is a lower triangular matrix with positive diagonal entries \(\phi_{ij}\) and unrestricted off-diagonal entries \(\phi_{jk}, j > k\). The random effects of the \(i\) th driving shift are reparameterized as \(\mathbf{b}_i = \Phi \mathbf{u}_i\), where \(\mathbf{u}_i \sim N(0, I_J)\) and \(I_J\) denotes a \(J \times J\) identity matrix. We assign the priors to the entries of \(\Phi\) such that \(\phi_{ij} \sim N^+(0, 1000)\) and \(\phi_{jk} \sim N(0, 1000)\), where \(N^+ = (0, \infty)\). A diagram of the multi-type Poisson process model with the prior specification is displayed in Figure 3.
To sample the parameters from the joint posterior, we follow Liu and Guo to implement a Metropolis-adjusted Langevin Algorithm designed for sampling the penalized spline parameters in the Bayesian multitype Poisson process model. This algorithm integrated with a joint sampling scheme helps achieve better convergence in the posterior sampling, with addressing the issue of the strong dependence of spline coefficients due to the random walk for the roughness penalty.

4 | BAYESIAN MODEL ASSESSMENTS

4.1 | Deviance information criterion

To assess the model fit of the data for the \( j \) th type of recurrent events, we specify the type-specific deviance as

\[
\text{Dev}_j(\theta_j) = -2 \sum_{i=1}^{n} \log L_{ij}(\theta_j | t_{ij}, D_i),
\]

where \( \theta_j = \{ \beta_j, \alpha_j, \gamma_j, b_{ij} \} \) are the model parameters and the random effects for the \( j \) th type of event, and the type-specific likelihood \( L_{ij}(\theta_j | t_{ij}, D_i) \) is given by (3). Following Reference 27, the type-specific DIC for the \( j \) th event type is defined as

\[
\text{DIC}_j = \text{Dev}_j(\theta_j) + 2p_{D[j]},
\]

where \( p_{D[j]} = \overline{\text{Dev}}_j(\theta_j) - \text{Dev}_j(\theta_j) \) is the effective number of parameters for the \( j \) th event type, and \( \overline{\theta}_j \) and \( \overline{\text{Dev}}_j(\theta_j) \) are the posterior means of \( \theta_j \) and \( \text{Dev}_j(\theta_j) \), respectively. Write \( \theta = \{ \theta_j, j = 1, \ldots, J \} \). The overall DIC for all types of recurrent events is further given by

\[
\text{DIC} = \text{Dev}(\overline{\theta}) + 2p_D = \sum_{j=1}^{J} \text{DIC}_j,
\]

where the overall deviance function and overall effective number of parameters can also be decomposed to the sum of the type-specific components such that \( \text{Dev}(\theta) = \sum_{j=1}^{J} \text{Dev}_j(\theta_j) \) and \( p_D = \sum_{j=1}^{J} p_{D[j]} \).

The type-specific \( \text{DIC}_j \) in (7) is attractive since it not only avoids analytical intractable integrals by treating the random effects as parameters in the likelihood function, but also provides a useful tool to compare Bayesian recurrent event models for a certain event type. The overall DIC, which is naturally decomposed into the type-specific \( \text{DIC}_j \)'s, measures the model fit of the data from all types of events together.

4.2 | CPO and LPML

Let \( t_{-i} \) and \( D_{-i} \) denote the observed data without the \( i \) th driving shift, and write \( \Gamma = \{ \Omega, \Sigma \} \). The CPO for the \( i \) th driving shift is defined as

\[
\text{CPO}_i = \int f(t_i | \Gamma, D_i)p(\Gamma | t_{-i}, D_{-i})d\Gamma,
\]

where

\[
f(t_i | \Gamma, D_i) = \int L_i(\Gamma | b_i, D_i)p(b_i | \Sigma)db_i,
\]

and \( p(\Gamma | t_{-i}, D_{-i}) \) denotes the marginal posterior distribution of \( \Gamma \) constructed by the data excluding the \( i \) th driving shift. Following the results from Chen et al. CPO can be rewritten as (CPO Identity I)

\[
\text{CPO}_i = \frac{1}{\int_{f(t_i | \Gamma, D_i)} \frac{1}{p(\Gamma | t, D)}d\Gamma}.
\]
The CPO Identity I leads to a popular Monte Carlo estimate of CPO using the MCMC samples from the posterior distribution\(^\text{30}\). Let \(\mathbf{\Gamma}_m\) stand for the \(m\)th MCMC sample of \(\mathbf{\Gamma}\) for \(m = 1, \ldots, M\). A Monte Carlo estimate of \(\text{CPO}^{-1}_i\) is given by

\[
\hat{\text{CPO}}^{-1}_i = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{f(t_i|\mathbf{\Gamma}_m, D_i)}. \tag{12}
\]

However, this Monte Carlo estimate involves the integral of the joint likelihood over the random effects \(b_i\)’s, which is computationally prohibitive under the complex multitype recurrent event models with time-varying effects and time-dependent covariates. Thus, we develop an easy-to-implement algorithm to approximate \(\text{CPO}_i\), which directly uses the MCMC samples generated from the joint posterior distribution\(^\text{5}\).

Let \(w_i(\mathbf{b}_i)\) be a normalized weight function such that \(\int w_i(\mathbf{b}_i) \, d\mathbf{b}_i = 1\) and let \(\{(\mathbf{\Gamma}_m, \mathbf{b}_im), m = 1, \ldots, M\}\) denote the MCMC samples of \((\mathbf{\Gamma}, \mathbf{b}_i)\) after burn-in. Following the CPO Identity II, \(^\text{15}\) a Monte Carlo estimate of the shift-level CPO is given by

\[
\hat{\text{CPO}}^{-1}_i = \frac{1}{M} \sum_{m=1}^{M} \frac{w_i(\mathbf{b}_im)}{f(t_i, \mathbf{b}_im|\mathbf{\Gamma}_m, D_i)}. \tag{13}
\]

The normalized weight function \(w_i(\mathbf{b}_i)\) can be constructed by a multivariate normal density via the Laplace approximation of the joint density given by

\[
f(t, \mathbf{b}|\mathbf{\Gamma}, D_i) = \prod_{j=1}^{J} \left\{ \prod_{k=1}^{n_j} \lambda_{ij}(t_{ij}k|D_i, b_{ij}) \exp(-\Lambda_{ij}(c_i)) \right\} p(\mathbf{b}_i). \tag{14}
\]

After some algebra, the weight \(w_i\) is set as

\[
w_i(\mathbf{b}_i) = \phi(\mathbf{b}_i; \tilde{\mu}, \tilde{\Sigma}), \tag{15}
\]

where

\[
\tilde{\mu} = \mathbf{b}_{i0}, \quad \tilde{\Sigma} = \left( -\frac{\partial^2 h(\mathbf{b}_{i0})}{\partial \mathbf{b_i} \partial \mathbf{b}_i^T} \right)^{-1}, \quad h(\mathbf{b}_i) = \sum_{j=1}^{J} n_j b_{ij} - \sum_{j=1}^{J} \tilde{\Lambda}_{ij}(c_i) \exp(b_{ij}) - \frac{1}{2} \mathbf{b}_i^T \tilde{\Sigma}^{-1} \mathbf{b}_i.
\]

\(\phi\) is the normal density function with the mean vector \(\tilde{\mu}\) and the covariance matrix \(\tilde{\Sigma}\), and \(\mathbf{b}_{i0}\) is the mode of \(h(\mathbf{b}_i)\) which can be calculated by numerically solving \(\frac{\partial h(\mathbf{b}_i)}{\partial \mathbf{b}_i} = 0\). The detailed derivation of the weight function \(w_i(\mathbf{b}_i)\) is given in Appendix. Therefore, all the components in the Monte Carlo estimate (13) are able to be directly computed by the MCMC samples from the joint posterior distribution. Furthermore, the LPML\(^\text{31}\) is defined as

\[
\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i).
\]

A larger LPML suggests a better model fit.

### 4.3 C-index for multitype recurrent event model

#### 4.3.1 Within-event C-index

Let \(N_j(t)\) denote the number of type \(j\) events occurring on the time interval \([0, t]\) of the \(i\)th driving shift. Conditioning on the random effects, the expectation of the number of type \(j\) events at time \(t\) for the \(i\)th driving shift is given by

\[
\mu_{ij}(t|b_{ij}) = E[N(t)|b_{ij}] = \int_0^t \lambda_{ij}(u|X_i, W_i(u), Z_i, b_{ij}) \, du, \tag{16}
\]
where the conditional intensity function \( \lambda_j(u|X_i, W_i(u), Z_i, b_{ij}) \) is given in (1). We rewrite \( \lambda_j(u|X_i, W_i(u), Z_i, b_{ij})du = \exp(b_{ij})\lambda_j(u|X_i, W_i(u), Z_i) \). The expected number of type \( j \) events at time \( t \) for the \( i \) th driving shift is given by

\[
\mu_j(t) = E_b[\mu_j(t|b_{ij})] = \int_0^t \int_0^t \exp(b_{ij})\lambda_j(u|X_i, W_i(u), Z_i)du \ p(b_i|\Sigma)db_i \\
= \int_0^t \lambda_j(u|X_i, W_i(u), Z_i)du \int_0^\infty \exp(b_{ij})p(b_{ij}|\sigma_j^2)db_{ij} \\
= \exp(\frac{\sigma_j^2}{2}) \int_0^t \lambda_j(u|X_i, W_i(u), Z_i)du,
\]

where \( \sigma_j^2 \) is the \( j \) th diagonal entry of the covariance matrix \( \Sigma \).

To measure the model’s discrimination ability for the type \( j \) event, we propose a within-event C-index based on the comparable pairs constructed by two driving shifts with different observed event rates of the type \( j \) event. The within-event C-index for event type \( j \) is defined as

\[
C_{W,j} = P\left( \frac{\mu_j(c_i)}{c_i} > \frac{\mu_j(c_{i'})}{c_{i'}} \bigg| \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij}}{c_{i'}} \right), \tag{18}
\]

where \( i \neq i', i = 1, \ldots, n, i' = 1, \ldots, n \), and the follow-up time \( c_i \) is independent to the occurrence of the type \( j \) events. A comparable pair is concordant if the comparisons between the expected event rates and between the observed event rates are consistent. Then, we estimate \( \hat{C}_{W,j} \) by the proportion of concordant pairs among all the comparable pairs, which is given by

\[
\hat{C}_{W,j} = \frac{\sum_{i=1}^n \sum_{i'=1}^n I \left\{ \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij}}{c_{i'}} \right\} I \left\{ \frac{\mu_j(c_i)}{c_i} > \frac{\mu_j(c_{i'})}{c_{i'}} \right\}}{\sum_{i=1}^n \sum_{i'=1}^n I \left\{ \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij}}{c_{i'}} \right\}}. \tag{19}
\]

### 4.3.2 Between-event C-index

For the multitype recurrent event model, comparable pairs can also be constructed between two types of events since the event rates are comparable across different event types. For two different types of recurrent event \( j \neq j' \), the between-event C-index is defined as

\[
C_{B,jj'} = P\left( \frac{\mu_j(c_i)}{c_i} > \frac{\mu_{j'}(c_{i'})}{c_{i'}} \bigg| \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij'}}{c_{i'}} \right), \tag{20}
\]

Similar to the within-event C-index, an estimate of the C-index between the types \( j \) and \( j' \) events is given by

\[
\hat{C}_{B,jj'} = \frac{\sum_{i=1}^n \sum_{i'=1}^n I \left\{ \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij'}}{c_{i'}} \right\} I \left\{ \frac{\mu_j(c_i)}{c_i} > \frac{\mu_{j'}(c_{i'})}{c_{i'}} \right\}}{\sum_{i=1}^n \sum_{i'=1}^n I \left\{ \frac{N(c_i)_{ij}}{c_i} > \frac{N(c_{i'})_{ij'}}{c_{i'}} \right\}}. \tag{21}
\]

where \( i = 1, \ldots, n \) and \( i' = 1, \ldots, n \). The between-event C-index can be interpreted as the proportion of concordant pairs among all the comparable pairs with event rates from two different event types. It evaluates the model’s discrimination ability between two different event types.

### 4.3.3 Overall C-index

Generally considering two event types, there are only two ways to construct the comparable pairs, either within one certain event type or between two different event types. Therefore, without loss of generality, we propose the overall C-index for a recurrent event model with a total of two types of events, which is given by
where \( \pi_{W,1}(\pi_{W,2}) \) and \( \pi_{B,12} \) denote the probability of a within-event comparable pair from the type 1(2) event and the probability of a between-event comparable pair constructed by the two types of events, respectively. The overall C-index with more than two types of events can be decomposed similarly with \( J \) within-event C-indices and \( \binom{J}{2} \) between-event C-indices. The overall C-index measures the model’s discrimination performance on all types of recurrent events simultaneously.

Within the Bayesian framework, each of the proposed C-indices is a function of the model parameters. Instead of a single value of a C-index within the frequentist framework, we can obtain the posterior distribution of the C-index by the posterior samples of the model parameters. Thus, the posterior values of the within-event C-index, the between-event C-index, and the overall C-index can be calculated by the MCMC samples from the joint posterior, denoted by \( \hat{C}_{W,j}^{(m)}, \hat{C}_{B,ij}^{(m)}, \) and \( \hat{C}_O^{(m)} \), respectively, where \( m = 1, \ldots, M \) and \( M \) is the iteration of MCMC chain. The means of \( \hat{C}_{W,j}^{(m)}, \hat{C}_{B,ij}^{(m)}, \) and \( \hat{C}_O^{(m)} \) are considered as the point estimates of the C-indices. In addition, the 95% credible interval of each C-index can also be easily obtained based on the posterior values.

## 5 SIMULATION STUDIES

### 5.1 Simulation study 1: DIC and LPML

We conduct a simulation study to evaluate the empirical performance of the proposed type-specific DIC, overall DIC and LPML on model comparison. We generate 100 datasets with multitype recurrent event data from multitype nonhomogeneous Poisson process models using the R package "reda." Specifically, in each simulation, we simulate 500 driving shifts with two types of recurrent event data and a fixed follow-up time \( c_t = 10 \). For each of the driving shifts, we firstly simulate the random effects \( b_i \) from a bivariate normal distribution \( N_2(0, \Sigma) \) where \( \Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \). Three shift-level covariates, \( X_i, Z_{i1}, \) and \( Z_{i2} \), are generated independently from Bernoulli(0.5), where \( X_i \) has time-independent effects and \( Z_{i1} \) and \( Z_{i2} \) have time-varying effects on the intensities. Conditional on the simulated random effects and the covariates, we then simulate the type \( j \) events from a nonhomogeneous Poisson process model with the type-specific conditional intensity function given by

\[
\lambda_j(t | X_i, Z_{i1}, Z_{i2}, b_{ij}) = \exp\{\alpha_j X_{i1} + \beta_0(t) + Z_{i1} \beta_{ij}(t) + Z_{i2} \beta_{2j}(t) + b_{ij}\},
\]

where \( \alpha_1 = \alpha_2 = 0.6 \) and each \( \beta_0(t) \) is defined by a linear combination of the cubic B-spline basis functions \( \beta_0(t) = \sum_{k=1}^{K+3} B_k(t) \beta_{jk} \) where \( l = 0, 1, 2, K = 6, \) and \( j = 1, 2. \) The knots are uniformly placed on the time axis. The specific settings of \( \beta_{jk}'s \) and the true curves of \( \beta_j(t)'s \) are given in Figure S1. We consider the following six scenarios to fit different models to each of the simulated datasets:

1. **TRUE**: Fit the model (22) with the true number \( (K = 6) \) of knots of the P-spline for the baseline intensity functions and the time-varying coefficients.
2. **MoreK10**: Fit the model (22) with 10 knots \( (K = 10) \) of the P-spline basis for the baseline intensity functions and the time-varying coefficients.
3. **LessK3**: Fit the model (22) with 3 knots \( (K = 3) \) of the P-spline basis for the baseline intensity functions and the time-varying coefficients.
4. **MissZ2**: Fit a misspecified model with the true number \( (K = 6) \) of knots, but without the predictor \( Z_{i2} \) with time-varying effects.
5. **MissX**: Fit a misspecified model with the true number \( (K = 6) \) of the knots, but without the predictor \( X_i \) with fixed effects.
6. **MissXZ2**: Fit a misspecified model with the true number \( (K = 6) \) of the knots, but without both \( X_i \) and \( Z_{i2} \).

For each of the simulated dataset, we take 3000 MCMC samples after 2000 burn-in iterations for the model in each of the scenarios. A sensitivity analysis in regards to the choice of the hyperparameter \( \alpha \), is conducted under the TRUE model.
FIGURE 4  Boxplots of the type-specific deviance information criterion (DIC) difference, overall DIC difference, and logarithm of the pseudo-marginal likelihood difference between TRUE and each of MoreK10, LessK3, MissZ2, MissX, and MissXZ2

No differences in coverage probability of the baseline intensity and time-varying coefficients are observed among TRUE models with different choices of $\alpha_i$ (Figure S2). The type-specific DIC, the overall DIC, and the LPML are calculated for each of the models in each simulation.

Figure 4 shows the boxplots of the type-specific DIC difference, the overall DIC difference, and the LPML difference between TRUE and each of the misspecified models. The summary statistics of DIC and LPML differences are given in Table S1. A negative DIC difference or a positive LPML difference suggests a better fit of TRUE in each of the comparisons. We see that except for MoreK10, the boxplots of the type-specific DIC difference and the overall DIC difference are almost below 0, and the boxplots for the LPML difference are almost above 0. This indicates that the proposed DIC and LPML are effective in distinguishing the better model in both type-specific and overall perspectives. In addition, LPML (LPML difference median = 0.776, IQR = (-0.576, 2.336), % of $> 0 = 67\%$) performs slightly better than the overall DIC (Overall DIC difference median = 1.102, IQR = (-5.787, 5.477), % of $< 0 = 44\%$) when comparing MoreK10 with TRUE.

5.2  Simulation study 2: C-index

We perform a simulation study to examine the empirical performance on assessing the discrimination ability of the proposed C-indices. Similar to Simulation Study 1, we generate 100 datasets based on the random effects $b \sim N_2(0, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 2 & 0.707 \\ 0.707 & 1 \end{pmatrix}$$

and the conditional intensity function given by

$$\lambda_i(t|x_{i1}, x_{i2}, z_{i1}, b_{ij}) = \exp\{a_1 x_{i1} + a_2 x_{i2} + \beta_0(t) + z_{i1} \beta_1(t) + b_{ij}\},$$

(23)
where $\alpha_{11} = 1.5, \alpha_{12} = 0.75, \alpha_{21} = 0.75, \alpha_{22} = 1.5$, and $X_{i1}$ and $X_{i2}$ are independently sampled from $N(0, 1)$. The baseline intensity function and time-varying coefficients $\beta_{ij}(t)$ have the same settings as those in Simulation Study 1. For each of the simulated data, we fit the following four models with the true number and location of knots:

1. **TRUE:** The true model (22) with the correct specification of the covariates.
2. **MissX1:** Fit a misspecified model without the covariate $X_{i1}$.
3. **MissX2:** Fit a misspecified model without the covariate $X_{i2}$.
4. **MissX1X2:** Fit a misspecified model without both covariates $X_{i1}$ and $X_{i2}$.

Same as Simulation Study 1, we take 3000 MCMC samples after 2000 burn-in iterations for each of the models. The posterior means of the within-event C-index, the between-event C-index, and the overall C-index are calculated for each of the models.

Figure 5 displays the boxplots of the posterior means of the three types of C-indices. The summary statistics of the C-index simulation results are given in Table S2. It shows that the proposed C-indices correctly indicate the discriminative abilities of the models, where TRUE obtains the highest discriminative ability and MissX1X2 has the lowest discriminative ability in all types of the C-indices. The within-event C-index also correctly reflects the effect size of the missing predictors. Specifically, we see that MissX1 has a lower type 1 within-event C-index (Median = 0.630, IQR = (0.616, 0.641)) compared with MissX2 (Median = 0.756, IQR = (0.744, 0.766)) since to the larger effect of $X_{i1}$ ($\alpha_{11} = 1.5$) on the intensity function of the type 1 event than $X_{i2}$ ($\alpha_{21} = 0.75$). Similarly, MissX1 has a higher type 2 within-event C-index (Median = 0.787, IQR = (0.777, 0.797)) than MissX2 (Median = 0.635, IQR = (0.623, 0.646)) since $X_{i1}$ ($\alpha_{12} = 0.75$) has a smaller effect on the type 2 event intensity function than $X_{i2}$ ($\alpha_{22} = 1.5$). Moreover, regarding to evaluating the TRUE model, the proposed

![FIGURE 5](image-url)  
Boxplots of the within-event C-index, between-event C-index, and overall C-index of TRUE, MissX1, MissX2, and MissX1X2
within-event C-index shows a lower discriminative ability on the type 1 event (Median = 0.791, IQR = (0.780, 0.801)) than on the type 2 event (Median = 0.827, IQR = (0.816, 0.836)). This is consistent with that the variance of the random effects of type 1 event is larger than that of the type 2 event.

6 | APPLICATION TO THE NDS DATA

We fit the Bayesian multitype Poisson process model discussed in Section 3 to the Commercial Truck Driver NDS data. The models with different B-spline settings are compared based on the assessment criteria developed in Section 4. The best fit model is first selected and further analysis under this model is then carried out in this section.

6.1 | Model comparison

Each of the driving shifts is considered as a multitype event Poisson process with type 1 to 4 events representing ULD, CRC, SB, and LB, respectively. Dummy variables are created for the categorical variables, including the restarting group, the driver type, the starting time group, and the sleeping group. Following the findings from previous studies on the same subject, we assume that each off-duty sleeping time group has different patterns of intensity functions of the SCE risk as well as the on-duty breaks incidence. Driving at 1:00 AM to 5:00 AM is considered as a time-dependent covariate in the model. The rest of the driving shift level factors are assumed to have fixed effects on the intensity functions for different event types.

The candidate models are constructed by setting different parameters for the roughness penalty of the B-splines, including the degree of basis splines ($\nu$), the number of knots ($k$) and the order of the difference penalty ($s$). Quadratic ($\nu = 2$) and cubic ($\nu = 3$) spline functions with a first ($s = 1$) or second ($s = 2$) order penalty are considered in the candidate models. In all of the posterior computations, we take 5000 burn-in samples, and 3000 MCMC samples with thinning of 10 iterations under each candidate model. Table 2 presents the values of the overall DIC and the LPML of the candidate models. We see from Table 2 that (i) the models with a smaller number of knots ($K = 5$) have larger overall DICs and smaller LPMLs than the models with 10 or 15 knots; (ii) for the models with a larger number of knots ($K = 10$ and $K = 15$), with the same $K$, the models with cubic splines ($\nu = 3$) and the first order penalty ($s = 1$) achieve larger overall DICs, and the models with more smoothed intensities ($\nu = 3, s = 2$) have smaller LPMLs. This is consistent with the empirical results that both LPML and overall DIC have a good performance when the number of knots is under specified while LPML prefers a simpler model when the number of knots is over specified for the P-splines. The slightly inconsistent model comparison results between DIC and LPML are also reported in the literature. In this data analysis, Model 4 ($K = 10, \nu = 3, s = 1$) has the second lowest overall DIC and the second highest LPML, indicating a good fit based on both overall DIC and LPML among all the candidate models. Thus, Model 4 is selected for further analysis. Moreover, the comparisons of the four type-specific DICs are displayed in Figure 6. The differences of the type-specific DICs between the candidate models are larger for the ULD and the SB events than the other two types of events. Specifically, the plots

| Model | $K$ | $\nu$ | $s$ | DIC   | LPML  |
|-------|-----|-------|-----|-------|-------|
| 1     | 5   | 3     | 1   | 18 762.1 | −10 177.1 |
| 2     | 5   | 3     | 2   | 18 780.0 | −10 222.5 |
| 3     | 5   | 2     | 1   | 18 790.1 | −10 274.0 |
| 4     | 10  | 3     | 1   | 18 734.1 | −10 162.1 |
| 5     | 10  | 3     | 2   | 18 769.5 | −10 129.2 |
| 6     | 10  | 2     | 1   | 18 741.8 | −10 254.7 |
| 7     | 15  | 3     | 1   | 18 733.8 | −10 285.5 |
| 8     | 15  | 3     | 2   | 18 770.0 | −10 185.3 |
| 9     | 15  | 2     | 1   | 18 739.1 | −10 217.8 |
FIGURE 6  Comparing the type-specific deviance information criterions among all of the candidate models

TABLE 3  Posterior estimates of the covariance matrix of the random effects under Model 4

| Mean (SD) 95% CI | ULD       | CRC       | SB         | LB         |
|-----------------|-----------|-----------|------------|------------|
|                 |           |           |            |            |
| ULD             | 2.124 (0.089) | 0.892 (0.024) | 0.208 (0.093) | −0.219 (0.190) |
|                 | [1.949, 2.295] | [0.840, 0.933] | [0.019, 0.386] | [−0.602, 0.139] |
| CRC             | 1.358 (0.059) | 0.021 (0.103) | −0.472 (0.170) |            |
|                 | [1.245, 1.476] | [−0.180, 0.219] |            |            |
| SB              | 0.781 (0.082) | 0.557 (0.231) |            |            |
|                 | [0.618, 0.936] | [0.033, 0.913] |            |            |
| LB              |            | 0.398 (0.090) |            |            |
|                 |            | [0.218, 0.578] |            |            |

Notes: The diagonal elements report the estimates of the standard deviations in the dependence structure, and the upper-diagonal elements present the estimates of the correlation between two types of events. Abbreviations: CI, Credible interval; SD, Posterior standard deviation.

in Figure 6 indicate that (i) the differences of type-specific DIC between the models are larger for ULD and SB compared with CRC and LB; (ii) Models 4 has relatively lower type-specific DIC for ULD, CRC, and SB events simultaneously.

6.2  Posterior estimates under the selected model

To investigate the association of the incidence rates between SCEs and on-duty breaks, Table 3 presents the posterior estimates of the parameters in the dependence structure of the joint model. A significant positive correlation is observed between ULDs and CRCs ($\rho_{12} = 0.892$, 95% CI = [0.840, 0.933]), which is consistent with the results in Liu and Guo. It is interesting to see a significant positive correlation between ULD events and SBs ($\rho_{13} = 0.208$, 95% CI = [0.019, 0.386]), which suggests that the driver’s decision to take a SB is related to the fatigue level experienced in driving. When
| Type | Coefficient | Mean  | SD    | 95% CI          |
|------|-------------|-------|-------|-----------------|
| ULD  | Commercial licence year | -0.228 | 0.089 | [-0.402, -0.054] |
|      | Driver type: long-haul | -0.074 | 0.249 | [-0.640, 0.361] |
|      | Number of nondriving work | -0.203 | 0.090 | [-0.376, -0.024] |
|      | Starting a shift in the morning vs daytime | -0.033 | 0.181 | [-0.377, 0.338] |
|      | Starting a shift at night vs daytime | -0.214 | 0.225 | [-0.666, 0.221] |
|      | Starting a shift in the morning vs night | 0.180 | 0.243 | [-0.330, 0.607] |
|      | Restarting driving vs no need to restart | 0.367 | 0.351 | [-0.325, 1.046] |
|      | Not restarting driving vs no need to restart | 0.121 | 0.186 | [-0.245, 0.503] |
|      | Not restarting driving vs restarting driving | -0.247 | 0.363 | [-0.917, 0.481] |
|      | Driving between 1:00 AM and 5:00 AM | -1.187 | 0.341 | [-1.884, -0.522] |

| CRC  | Commercial licence year | -0.099 | 0.056 | [-0.205, 0.017] |
|      | Driver type: long-haul | -0.044 | 0.137 | [-0.317, 0.222] |
|      | Number of non-driving work | -0.062 | 0.055 | [-0.169, 0.045] |
|      | Starting a shift in the morning vs daytime | -0.098 | 0.120 | [-0.336, 0.134] |
|      | Starting a shift at night vs daytime | -0.271 | 0.146 | [-0.554, -0.002] |
|      | Starting a shift in the morning vs night | 0.174 | 0.156 | [-0.144, 0.457] |
|      | Restarting driving vs no need to restart | 0.231 | 0.242 | [-0.244, 0.690] |
|      | Not restarting driving vs no need to restart | -0.041 | 0.121 | [-0.279, 0.198] |
|      | Not restarting driving vs restarting driving | -0.272 | 0.250 | [-0.751, 0.211] |
|      | Driving between 1:00 AM and 5:00 AM | -0.473 | 0.192 | [-0.855, -0.110] |

| SB   | Commercial licence year | -0.127 | 0.059 | [-0.245, -0.010] |
|      | Driver type: long-haul | -0.230 | 0.185 | [-0.596, 0.123] |
|      | Number of nondriving work | -0.233 | 0.063 | [-0.361, -0.112] |
|      | Starting a shift in the morning vs daytime | 0.350 | 0.123 | [0.107, 0.599] |
|      | Starting a shift at night vs daytime | -0.111 | 0.166 | [-0.434, 0.209] |
|      | Starting a shift in the morning vs night | 0.461 | 0.157 | [0.148, 0.766] |
|      | Restarting driving vs no need to restart | -0.371 | 0.288 | [-0.955, 0.185] |
|      | Not restarting driving vs no need to restart | -0.187 | 0.130 | [-0.442, 0.070] |
|      | Not restarting driving vs restarting driving | 0.187 | 0.298 | [-0.383, 0.789] |
|      | Driving between 1:00 AM and 5:00 AM | -0.242 | 0.202 | [-0.655, 0.154] |

| LB   | Commercial licence year | 0.030 | 0.047 | [-0.067, 0.120] |
|      | Driver type: long-haul | 1.383 | 0.348 | [0.716, 2.068] |
|      | Number of non-driving work | -0.357 | 0.065 | [-0.484, -0.226] |
|      | Starting a shift in the morning vs daytime | 0.410 | 0.118 | [0.181, 0.636] |
|      | Starting a shift at night vs daytime | -0.291 | 0.176 | [-0.645, 0.045] |
|      | Starting a shift in the morning vs night | 0.702 | 0.172 | [0.367, 1.046] |
|      | Restarting driving vs no need to restart | -0.465 | 0.292 | [-1.081, 0.060] |
|      | Not restarting driving vs no need to restart | 0.210 | 0.111 | [-0.006, 0.424] |
|      | Not restarting driving vs restarting driving | 0.675 | 0.296 | [0.140, 1.283] |
|      | Driving between 1:00 AM and 5:00 AM | 0.353 | 0.168 | [0.020, 0.672] |

Abbreviations: CI, Credible interval; SD, Posterior standard deviation.
experiencing more ULD events due to fatigue, the driver may have a higher chance of taking a SB to mitigate fatigue. In addition, negative correlation between CRCs and LBs is obtained ($\rho_{24} = -0.472, 95\% CI = [-0.803, -0.128]$). This indicates that drivers may benefit from taking a long on-duty break to improve driving safety.

Table 4 presents the posterior estimates of the fixed effects of both time-independent and time-dependent covariates. Among the driving shifts in the “not restarting group,” the drivers were more likely to take LB compared to the “restarting driving” group. For the intensity of LB, the posterior estimate of the coefficient for “not restarting driving” compared with “restarting driving” is 0.675 (95% CI = [0.140, 1.283]). This result indicates that drivers may take more long on-duty breaks to compensate for the fatigue accumulated over a week-long continuous working period. During the early morning period (1:00 AM-5:00 AM), we see lower risks of ULD (−1.187, 95% CI = [−1.884, −0.522]) and CRC (−0.473, 95% CI = [−0.855, −0.110]) events compared to the other time of day. The light traffic condition could be the primary reason for the lower SCE risks during the early morning. Driving between 1:00 AM and 5:00 AM was positively associated with on-duty LB (0.353, 95% CI = [0.020, 0.672]), which indicates that drivers may choose to have additional rest when driving in the late night. The starting time of a driving shift significantly impacted the short and LB intensities. Specifically, the driving shifts starting in the morning had a higher chance of taking both short and LB compared to those starting in the daytime (SB: 0.350, 95% CI = [0.107, 0.599]; LB: 0.410, 95% CI = [0.181, 0.636]) and at night (SB: 0.461, 95% CI = [0.148, 0.766]; LB: 0.702, 95% CI = [0.367, 1.046]). The number of nondriving tasks was associated with lower risk of ULD events (−0.203, 95% CI = [−0.376, −0.024]) as well as both lower SB (−0.233, 95% CI = [−0.361, −0.112]) and LB (−0.357, 95% CI = [−0.484, −0.226]) intensities. Thus, the nondriving tasks may be considered as a break from driving, which positively affects driving performance. In addition, the drivers with longer years of commercial license reported a lower risk of ULD (−0.228, 95% CI = [−0.402, −0.054]) and a lower chance of taking a SB (−0.127, 95% CI = [−0.245, −0.010]).

The estimated smoothing curves of the baseline intensities and the time-varying coefficients are presented in Figure 7. Similar to the results in Liu and Guo, we observe an increase of the intensity rate of ULD events in each of the insufficient

**Figure 7** Time-varying effect of the off-duty sleeping groups on the intensity function of unintentional lane deviations, crash-relevant conflict, short breaks, and long breaks events
sleeping group and the abundant sleeping group. However, there are no obvious time-varying effects of the sleeping groups on the intensity functions of the on-duty break events.

As shown in Figure 8, we calculate the proposed C-indices for the multi-type recurrent event model to evaluate the selected model’s adequacy on the four types of recurrent events. In general, Model 4 fits the data moderately well with an overall C-index of 0.592 (95% CI = [0.580, 0.603]). Model 4 has a relatively high discrimination ability on LB with the within-event C-index 0.694 (95% CI = [0.684, 0.701]). The within-event C-indices for ULD, CRC, and SBs are 0.589 (95% CI = [0.572, 0.603]), 0.560 (95% CI = [0.544, 0.574]), and 0.584 (95% CI = [0.568, 0.599]), respectively, which are lower than that for the LB events. These results are consistent with the the posterior estimates of the parameters in Model 4. The model has more significant fixed effects and a smaller variance of random effect for LB events, which results in a relatively higher discrimination ability on the LB events.

As displayed in Figure 9, we further compare the C-indices between Model 4 and a null model having the off-duty sleeping groups as the only predictor with time-varying effects. Within each of the three sleeping groups, Model 4 gains an increase of the within-event C-index for both short and LB by including the fix effects of both time-invariant and time-dependent covariates, since many of these predictors are significant to the intensity of SB and LB as shown in Table 4. The within-event C-indices for the two SCE events of the two models are similar.

7 | DISCUSSION

In this paper, we have developed several model assessment criteria for evaluating and comparing the adequacy of the Bayesian multitype recurrent event models. Our proposed approaches are attractive in assessing the goodness-of-fit of the models in different aspects: (i) the easily computed type-specific DICs assess the model fit for each type of the recurrent events, and (ii) the overall DIC and the LPML capture the model fit for all types of the recurrent events simultaneously.
Unlike the DIC, the type-specific decomposition of the LPML is not straightforward but possible by following the LPML decomposition under the joint models for longitudinal and survival data. However, such decomposition requires substantial theoretical derivations and new developments of computational algorithms.

The proposed C-indices are able to evaluate the discrimination abilities of the Bayesian multi-type recurrent event models in the different angles—that is, within-event C-index, between-event C-index, and overall C-index. Compared with the existing C-index for recurrent event models, the proposed method is still easy to implement under the complex multi-type recurrent model with a large number of random effects, time-varying coefficients and time-dependent predictors since the comparable pairs are constructed based on the event rates from two event processes. We note that the independence between the termination of the event process and the occurrence of events is required, which is proper in the NDS application since the end of a driving task is usually prescheduled rather than based on SCEs or driving behaviors occurring in a driving shift.

This study is among the first to jointly model drivers’ rest break behaviors and safety outcomes with consideration of several key risk factors, such as driving between 1:00 AM and 5:00 AM, week-long cumulative driving hours, and driving shift starting time. The results indicate that driving between 1:00 AM and 5:00 AM is associated with significantly lower ULD and SCE intensity. The 1:00 AM to 5:00 AM period is a critical rest time due to the circadian rhythm and it is generally suggested that sleeping during this time is important for reducing fatigue. However, many truck drivers prefer driving at night to avoid traffic congestion as well as being constrained by a delivery schedule. A truck driver can adjust their circadian rhythm to mitigate the fatigue associated with night time driving. The low traffic volume during the 1:00 AM to 5:00 AM period represents less vehicle interaction and thus a relatively safe driving environment, which could lead to the relatively low ULD and CRC during this time. A comprehensive evaluation of the benefit and risk is needed to provide a safety strategy for driving at this critical period.
The results also imply drivers tend to take various countermeasures to mitigate fatigue. When they could not take a 34-hour break after 60 hours over 7 consecutive days or 70 hours over 8 consecutive days as recommended by the HOS rule, the drivers were more likely to take LB while driving. Similarly, the intensity of LB was higher during the 1:00 AM to 5:00 AM period. When taking more work-related breaks, drivers tended to have fewer rest breaks, both long and short. These results imply the self-adaptation of drivers in managing their schedule to improve driving safety. A systematic fatigue management approach using a combination of various methods will mitigate fatigue while driving and benefit driving safety for commercial truck drivers.

ACKNOWLEDGMENTS
We thank Richard Hanowski and Susan Soccolich for their help with truck driver safety and the Hour of Service regulations. We would also like to thank the editor, the associate editor, and the two referees for their constructive comments and suggestions, which have led to a much improved version of the paper.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available from the Virginia Tech Transportation Institute. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the corresponding author with the permission of the Virginia Tech Transportation Institute.

ORCID
Ming-Hui Chen  https://orcid.org/0000-0003-1935-2447
Feng Guo  https://orcid.org/0000-0002-2572-481X

REFERENCES
1. NHTSA. National motor vehicle crash causation survey: report to Congress (DOT HS 811 059); 2013. 1492772607. doi: citeulike-article-id:13424710.
2. FMCSA. Hours of service of drivers, federal register; Vol 85, 2020:105, FMCSA–2018–0248.
3. Dingus TA, Guo F, Lee S, et al. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc Natl Acad Sci. 2016;113(10):2636-2641.
4. Hanowski RJ, Hickman J, Fumero MC, Olson RL, Dingus TA. The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. Accid Anal Prevent. 2007;39(6):1140-1145.
5. Sweeney MM, Ellingstad VS, Mayer DL, Eastwood MD, Weinstein EB, Loeb BS. The Need for Sleep: Discriminating Between Fatigue-Related and Nonfatigue-Related Truck Accidents. Vol 39. Sage CA: SAGE Publications; 1995:1122-1126.
6. Chen GX, Fang Y, Guo F, Hanowski RJ. The influence of daily sleep patterns of commercial truck drivers on driving performance. Accid Anal Prevent. 2016;91:55-63. doi:10.1016/j.aap.2016.02.027
7. Blanco M, Hanowski RJ, Olson RL, et al. The impact of driving, non-driving work, and rest breaks on driving performance in commercial motor vehicle operations; 2011.
8. Chen C, Xie Y. Modeling the safety impacts of driving hours and rest breaks on truck drivers considering time-dependent covariates. J Safety Res. 2014;51:57-63. doi:10.1016/j.jsr.2014.09.006
9. Liu Y, Guo F. A Bayesian time-varying coefficient model for multitype recurrent events. J Comput Graph Stat. 2020;29(2):383-395. doi:10.1080/10618600.2019.1686988
10. Liu Y, Guo F, Hanowski RJ. Assessing the impact of sleep time on truck driver performance using a recurrent event model. Stat Med. 2019;38(21):4096-4111. doi:10.1002/sim.8287
11. Li Z, Chinchilli VM, Wang M. A Bayesian joint model of recurrent events and a terminal event. Biometr J. 2019;61(1):187-202.
12. Lin LA, Luo S, Chen BE, Davis BR. Bayesian analysis of multi-type recurrent events and dependent termination with nonparametric covariate functions. Stat Methods Med Res. 2017;26(6):2869-2884.
13. Lin LA, Luo S, Davis BR. Bayesian regression model for recurrent event data with event-varying covariate effects and event effect. J Appl Stat. 2018;45(7):1260-1276.
14. Sinha D, Maiti T, Ibrahim JG, Ouyang B. Current methods for recurrent events data with dependent termination: a Bayesian perspective. J Am Stat Assoc. 2008;103(482):866-878.
15. Zhang D, Chen MH, Ibrahim JG, Boye ME, Shen W. Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. J Comput Graph Stat. 2017;26(1):121-133.
16. Antolini L, Boracchi P, Bignanzi E. A time-dependent discrimination index for survival data. Stat Med. 2005;24(24):3927-3944.
17. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247(18):2543-2546.
18. Mauguen A, Collette S, Pignon JP, Rondeau V. Concordance measures in shared frailty models: application to clustered data in cancer prognosis. Stat Med. 2013;32(27):4803-4820.
19. Kim S, Schaubel DE, McCullough KP. AC-index for recurrent event data: Application to hospitalizations among dialysis patients. Biometrics. 2018;74(2):734-743.
20. Guo F. Statistical methods for naturalistic driving studies. *Annu Rev Stat Appl*. 2019;6(1):309-328. doi:10.1146/annurev-statistics-030718-105153
21. Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci*. 1996;11(2):89-121.
22. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Stat*. 2004;13(1):183-212.
23. Jullion A, Lambert P. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Comput Stat Data Anal*. 2007;51(5):2542-2558.
24. Chen BE, Cook RJ, Lawless JF, Zhan M. Statistical methods for multivariate interval-censored recurrent events. *Stat Med*. 2005;24(5):671-691.
25. Chen Z, Dunson DB. Random effects selection in linear mixed models. *Biometrics*. 2003;59(4):762-769.
26. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Stat*. 2004;13(1):183-212.
27. Jullion A, Lambert P. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Comput Stat Data Anal*. 2007;51(5):2542-2558.
28. Chen Z, Dunson DB. Random effects selection in linear mixed models. *Biometrics*. 2003;59(4):762-769.
29. Geisser S, Eddy WF. A predictive approach to model selection. *J Am Stat Assoc*. 1979;74(365):153-160.
30. Gelfand AE, Dey DK. Bayesian model choice: asymptotics and exact calculations. *J Royal Stat Soc Ser B (Methodol)*. 1994;56(3):501-514.
31. Ibrahim JG, Chen MH, Sinha D. *Bayesian survival analysis*. Wiley StatsRef: Statistics Reference Online; 2014.
32. Wang W, Fu H, Yan J. reda: recurrent event data analysis. R package version 0.4 2017:1.
33. Ge M, Chen MH. Bayesian inference of the fully specified subdistribution model for survival data with competing risks. *Lifetime Data Anal*. 2012;18(3):339-363.
34. Zhang F, Chen MH, Cong XJ, Chen Q. Assessing importance of biomarkers: a Bayesian joint modelling approach of longitudinal and survival data with semi-competing risks. *Stat Modell*. 2020;21(1-2):30-55.1471082X20933363.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zhang Y, Chen M-H, Guo F. Bayesian criterion-based assessments of recurrent event models with applications to commercial truck driver behavior studies. *Statistics in Medicine*. 2022;41(23):4607-4628. doi: 10.1002/sim.9528

**APPENDIX A. DERIVATION OF THE NORMALIZED WEIGHT FUNCTION**

As introduced in Section 4.2, the normalized weight function $w_i$ is chosen by a multivariate normal density constructed by the Laplace approximation of the joint density (14). We write that $\Lambda_{ij}(c_i) = \bar{\Lambda}_{ij}(c_i) \exp(b_{ij})$ and $\pi_{ij}(t) = \exp(\mathbf{X}'_i \mathbf{a}_j + \mathbf{W}_i(t)' \mathbf{y}_j + \mathbf{b}_{ij}' + \mathbf{Z}_i \mathbf{\beta}_j)$. Then, the joint density (14) can be written by

$$f(t_i, b_i | \Gamma, D_i) = \prod_{j=1}^{J} \left( \prod_{k=1}^{n_j} (\pi_{ij}(t_{ijk}) \exp(b_{ij})) \exp(-\bar{\Lambda}_{ij}(c_i) \exp(b_{ij})) \right)$$

$$\times \frac{1}{(2\pi)^{\frac{J}{2}} | \mathbf{\Sigma} | ^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{b}_i' \mathbf{\Sigma}^{-1} \mathbf{b}_i \right)$$

$$= \prod_{j=1}^{J} \left( \prod_{k=1}^{n_j} \pi_{ijk} \right)$$

$$\times \frac{1}{(2\pi)^{\frac{J}{2}} | \mathbf{\Sigma} | ^{\frac{1}{2}}} \exp\left( \sum_{j=1}^{J} n_j b_{ij} - \sum_{j=1}^{J} \bar{\Lambda}_{ij}(c_i) \exp(b_{ij}) - \frac{1}{2} \mathbf{b}_i' \mathbf{\Sigma}^{-1} \mathbf{b}_i \right)$$

$$= C_i \frac{1}{(2\pi)^{\frac{J}{2}} | \mathbf{\Sigma} | ^{\frac{1}{2}}} \exp(h(b_i)),$$
where

\[ C_i = \prod_{j=1}^{J} \left( \prod_{k=1}^{n_i} \pi_j(t_{ijk}) \right) \quad \text{and} \quad h(b_i) = \sum_{j=1}^{J} n_j b_{ij} - \sum_{j=1}^{J} \bar{\lambda}_{ij}(c_i) \exp(b_{ij}) - \frac{1}{2} b_i^T \Sigma^{-1} b_i. \]

Let \( b_{i0} \) denote the mode of \( h(b_i) \) by solving \( \frac{\partial h(b_i)}{\partial b_i} = 0 \). In practice, the equation can be solved numerically. By the Laplace approximation, we can get

\[
 f(t_i, b_i | \Gamma, D_i) \approx \frac{C_i}{(2\pi)^{\frac{J}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left( h(b_{i0}) + \frac{1}{2} (b_i - b_{i0})^T \frac{\partial^2 h(b_{i0})}{\partial b_i \partial b_i^T} (b_i - b_{i0}) \right)
\]

\[
 = \frac{C_i \exp(h(b_{i0}))}{|\Sigma|^{\frac{1}{2}}} \times \frac{1}{(2\pi)^{\frac{J}{2}}} \exp \left( \frac{1}{2} (b_i - b_{i0})^T \frac{\partial^2 h(b_{i0})}{\partial b_i \partial b_i^T} (b_i - b_{i0}) \right),
\]

and the elements in the Hessian matrix \( \frac{\partial^2 h(b_i)}{\partial b_i \partial b_i^T} \) are given by

\[
 \frac{\partial^2 h}{\partial b_{ij}^2} = -\bar{\lambda}_{ij}(c_i) \exp(b_{ij}) - \sigma_{jj} \quad \text{and} \quad \frac{\partial^2 h}{\partial b_i \partial b_k} = -\sigma_{kj},
\]

where \( \sigma_{kj} \) is the entry of row \( k \) and column \( j \) of \( \Sigma^{-1} \). Therefore, we set up the \( w_i(b_i) \) as

\[ w_i(b_i) = \phi(b_i; \bar{\mu}, \bar{\Sigma}). \]

Let \( \phi \) denote the normal density function with mean \( \bar{\mu} \) and covariance matrix \( \bar{\Sigma} \), where \( \bar{\mu} = b_{i0} \) and \( \bar{\Sigma} = \left( -\frac{\partial^2 h(b_{i0})}{\partial b_i \partial b_i^T} \right)^{-1} \).