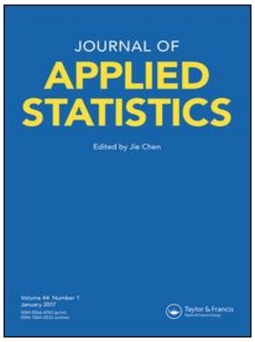

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

# A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers

Qing Li, Feng Guo, Inyoung Kim, Sheila G. Klauer & Bruce G. Simons-Morton

JOURNAL OF APPLIED STATISTICS, 2017
http://dx.doi.org/10.1080/02664763.2017.1288202

# A Bayesian finite mixture change-point model for assessing the risk of novice teenage drivers

Qing Li[a], Feng Guo[b,c], Inyoung Kim[b], Sheila G. Klauer[c] and Bruce G. Simons-Morton[d]

[a]Department of Statistics, University of Wisconsin–Madison, Madison, WI, USA; [b]Department of Statistics, Virginia Tech, Blacksburg, VA, USA; [c]Virginia Tech Transportation Institute, Blacksburg, VA, USA; [d]Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), Bethesda, MD, USA

**ABSTRACT**

The driving risk during the initial period after licensure for novice teenage drivers is typically the highest but decreases rapidly right after. The change-point of driving risk is a critical parameter for evaluating teenage driving risk, which also varies substantially among drivers. This paper presents latent class recurrent-event change-point models for detecting the change-points. The proposed model is applied to the Naturalist Teenage Driving Study, which continuously recorded the driving data of 42 novice teenage drivers for 18 months using advanced in-vehicle instrumentation. We propose a hierarchical BFMM to estimate the change-points by clusters of drivers with similar risk profiles. The model is based on a non-homogeneous Poisson process with piecewise-constant intensity functions. Latent variables which identify the membership of the subjects are used to detect potential clusters among subjects. Application to the Naturalistic Teenage Driving Study identifies three distinct clusters with change-points at 52.30, 108.99 and 150.20 hours of driving after first licensure, respectively. The overall intensity rate and the pattern of change also differ substantially among clusters. The results of this research provide more insight in teenagers' driving behaviour and will be critical to improve young drivers' safety education and parent management programs, as well as provide crucial reference for the GDL regulations to encourage safer driving.

**ARTICLE HISTORY**
Received 17 September 2016
Accepted 22 January 2017

**KEYWORDS**
Change-point models; Latent class models; Naturalistic driving study; Recurrent events; Teenage driving risk

## 1. Introduction

Motor vehicle crashes are the leading cause of mortality for teenagers and young adults [8]. Teenage driver risk has been a focus of traffic safety research. Statistics from the National Center for Statistics and Analysis showed that young drivers between 15 and 20 years old had more fatal crashes than other age groups [52]. Compared to experienced drivers, novice teenage drivers are susceptible to driving conditions and driving behaviours. Williams [73] showed that the driving situation such as night or day driving, alcohol use, and the presence of passengers would affect the teenage driving risk. Kim *et al.* [31] stated

**CONTACT** Feng Guo ✉ feng.guo@vt.edu

that the presence of a passenger and night driving decreased the driving risk, while having risky friends increased the risk. Klauer *et al.* [33] showed that the risk of secondary driving task engagement was considerably higher for teenage drivers compared to experienced drivers.

The initial high-risk period is crucial for novice teenage drivers' safety education program development and the graduated driver licensing (GDL) regulations. Studies have shown that the driving risk during initial period after licensure is the highest and followed by a rapid decrease [39,45,63,73], indicating a change-point of driving risk during the initial period after licensure. Based on crash and fatality database, Mayhew *et al.* [45] examined the crash rate of novice teenager drivers by month and found that the driving risk decreased significantly after the first few months. Using naturalistic driving study approach, the driving risk were shown to decrease significantly at six months after licensure [26,39,40,63].

Teenage driving risk varies substantially among individuals [32,51,59,71]. Guo *et al.* [26] showed three clusters with distinct risk levels existed among teenage drivers using naturalistic driving data. Deery *et al.* [11] identified young driver subtypes related to personality, driving experience, and environmental factors from two studies. Ouimet *et al.* [55] demonstrated that the teenage drivers with higher level of cortisol when stressed tended to show low-risk level and faster decrease in risk compared to the drivers with low cortisol level. The high variability in driver risk level and risk profile over time should be considered when determining the change-point of novice teenage driving risk.

Most existing research aggregated data into fixed calendar-time intervals for analysis, such as by three-month [26]. However, driving experience is more directly related to actual driving time than calendar time. The change-point in driving time provides an important measure on how much experience is required before relatively safer driving occurs [63]. The GDL or the learner's permit laws for teenagers commonly require specific driving time. For example, GDL in Virginia requires 45 hours of supervised driving including 15 hours at night for teenagers before obtaining a licence [19]. Limited research has been conducted to evaluate the change-point of driving risk in terms of cumulative driving time. Therefore, there is a need to study the change-point with respect to driving time.

The Naturalistic Driving Study (NDS), an innovative driving data collection method, provides detailed information on driver performance, environment and other safety-critical factors by installing advanced instruments and sensors in vehicles [22–24,55]. The participants of an NDS drive as in everyday life without special instructions or the presence of experimenters. The NDS provides continuous driving data without modifying the environment, which is especially useful for evaluating the novice teenage driving risk.

The Naturalistic Teenage Driving Study (NTDS), sponsored by the National Institute of Child Health and Development, was conducted between 2006 and 2009 in Virginia [39]. The study recruited 42 newly licenced teenage drivers within two weeks after licensure and followed them up for 18 months. The vehicles of the participants were instrumented with sophisticated data acquisition systems developed by the Virginia Tech Transportation Institute. The data collection system continuously collected data from five video cameras, GPS, kinematic sensors, as well as vehicle network. Crashes and near-crashes (CNC) were identified from the recorded data. The NTDS provides a unique opportunity to evaluate the driving risk change-points for novice teenage drivers using objectively observed exposure and safety critical data.

Piecewise-constant hazard functions are widely used in change-point models: Matthews and Farewell [44] and Loader [42] derived the likelihood ratio test for the hazard rate with a change-point against a constant hazard. Yao [74] proposed a maximum likelihood estimator (MLE) for the change-point. Müller and Wang [49] provided a review on different change-point estimators. Karasoy and Kadilar [29] analyzed change-point models using Bayesian methods.

Typically, the teenage drivers would experience multiple CNC events during the study period. Therefore, our focus is on recurrent-event change-point models. A commonly used recurrent-event modelling framework is based on Poisson processes [10, Chapter 2]. The non-homogeneous Poisson process (NHPP) is a Poisson process when the intensity function varies across time [60, p. 32]. The time where the intensity function of an NHPP changes quickly is a change-point.

Most of the research on recurrent-event change-point models assume piecewise-constant intensity, rate or hazard functions, which is highly robust [37]. Lawless and Zhan [37] analyzed recurrent-event data using piecewise-constant rate functions by mixed-Poisson-Process method and robust estimation, where the exact occurrence time of events were not observed. Based on the assumption that the event counts were NHPP with piecewise-constant intensity functions, West and Odgen [72] and Aschar *et al.* [3] estimated the change-points for one individual with multiple events; Frobish and Ebrahimi [14] developed a maximum likelihood estimator (MLE) and a Nelson–Aalen estimator for the change-point; Li *et al.* [40] developed two recurrent-event change-point models to detect the time of change in driving risks by maximizing the likelihood. The models in Frobish and Ebrahimi [14] and Li *et al.* [40] assumed identical change-points among subjects.

Majority of the existing models assume the change-points and intensity function are identical for all subjects. However, studies have shown that risk profiles vary by drivers [26]. Varying change-points by individuals can accommodate the heterogeneity among drivers but could lead to over-parametrization issues. It has been demonstrated that clusters existed among teenagers with different patterns of risk change [26,32,51,59,71].

The latent class analysis (LCA) can be used to detect subtypes and classify subjects according to their maximum class membership probability [38]. The LCA assumes unobserved subclasses among observed data and typically with same model form but different parameters. Goodman [18] proposed MLEs for the model parameters, which made the LCA more practically useful. Hagenaars and McCutcheon [27] developed a general framework for LCA. Other applications of LCA includes density estimation, random-effects modelling, and scaling [46].

A commonly used probabilist model of LCA is the finite mixture model (FMM), which can achieve model-based clustering when the overall population are observed with missing subgroup identity [70]. The density function of observations is a weighted sum of a finite number of known parametric distributions. The weight, or the mixing proportion, is non-negative and the summation is one. Assuming that the observations are from a mixture distribution, the optimal number of components, the mixing proportions, the component parameters given the number of components, and the model adequacy can be addressed formally [47]. FMMs can also be used to track changes across time in the data structure [27]. FMMs rely on the assumption of the data distribution while other clustering algorithm used certain distance measure to find similarities among clusters, therefore, FMMs are more flexible [15].

In this study, we propose a Bayesian latent class change-point model to detect the change-points of novice teenage driving risk in the context of recurrent-event using naturalistic driving data. The latent model is based on Bayesian finite mixture model (BFMM) [12]. The BFMM uses latent variables to represent the mixture structure in a hierarchical manner, and is appropriate and flexible when the number of subgroups is unknown [5,57]. Diebolt and Robert [12] proposed a Gibbs sampling approach to sample from the posterior distribution of a BFMM. Celeux *et al.* [7] provided alternatives for Bayesian estimators when multi-modes existed in the posterior. McLachlan and Peel [47] discussed the methodology, applications, and major issues such as identifiability problems and evaluating the number of components of BFMM. Methods to analyze the BFMM when the number of components was unknown were presented in [54,56,66].

In this research, the drivers are assumed to be from several latent classes, with the number of classes unknown. Estimating the number of classes in an FMM is a special model selection problem: Kass and Raftery [30] and Ishwaran *et al.* [28] developed a weighted Bayes factor method. Green [20] and Gruet *et al.* [21] proposed reversible jump Markov chain Monte Carlo (RJMCMC); Mengersen and Robert [48] and Sahu and Cheng [61] proposed K-L divergence or entropy distance; Stephens [66] and Cappè *et al.* [6] used a birth-and-death process; Nobile [54] sampled from the posterior of the number of components via Markov chain Monte Carlo (MCMC); information-based criteria are also used such as Bayesian information criterion (BIC) [58,62] and deviance information criterion (DIC) [65].

For estimating the number of clusters in an FMM, the methods based on Bayes factor or entropy distance emphasize on testing perspective; the computational challenge of RJMCMC is relatively high and the validation of moves between parameter spaces with different dimensions can be challenging [43]. Under certain conditions, the method using a birth-and-death process has similar properties as RJMCMC according to the comparison between two methods in [6]. Model selection via DIC is more straightforward than selection via Bayes factor and DIC is easier to calculate from the MCMC samples than AIC and BIC [4], though DIC might result in over-fitted models [1].

The CNC event count of each teenager is assumed to follow an NHPP with a piecewise-constant intensity function and each subject is assumed to have one unknown change-point, because Li *et al.* [40] showed that the model with one change-point was optimal for NTDS. The proposed Bayesian latent class change-point models assume subjects in the same latent class share identical intensity rates and change-point. The number of clusters, the cluster parameters and the subject identity are estimated. Assuming that the change-point of each teenager comes from a finite mixture, the mixing proportions of each subject are assigned a Dirichlet distribution prior in a hierarchical Bayesian framework. We would expect several clusters with distinct subgroup features.

The rest of the paper is organized as follows. Section 2 provides details of NTDS and conducts an exploratory data analysis on the NTDS. Section 3 develops the Bayesian method to classify the teenage drivers. Section 4 applies the method to the simulated data and checks the sensitivity of the method in different scenarios. Section 5 applies the model to the NTDS data. Section 6 presents the conclusion and discusses the future work.

## 2. Naturalistic teenage driving methods

Traditional analyses on teenage drivers driving behaviour and factors related to driving risk were based on data reported by police or subjects and crash database, which are prone to information bias [13]. However, the accurate information is especially critical to evaluate the risk pattern of novice teenage drivers whose risk change rapidly over a short-time period with considerable variation among drivers. Naturalistic driving studies provide detailed and accurate data to evaluate driving risk patten and risk factors.

The NTDS was conduced between 2006 and 2009, during which the driving behaviour of 22 female and 20 male teenagers from Virginia were recorded continuously, started within two weeks from first licensure [39]. The primary vehicles of the participants were installed with sophisticated devices including a computer, sensors, video cameras, and a GPS. The data acquisition systems continuously recorded video and kinematic data from ignition on to ignition off. Surveys and tests were conducted before or during the study to collect demographic information, personality, and family income, etc. The participants typically adjusted to the instrumentation in a short time [13].

CNC were identified via a multi-step process. The near-crash is ‘any circumstance that requires a rapid, evasive maneuver by the subject vehicle, or any other vehicle, pedestrian, cyclist, or animal to avoid a crash [39, p. 1474].’ Both CNC are critical to safety and are typically combined for small size naturalistic driving studies [22,25,33,34,55,64].

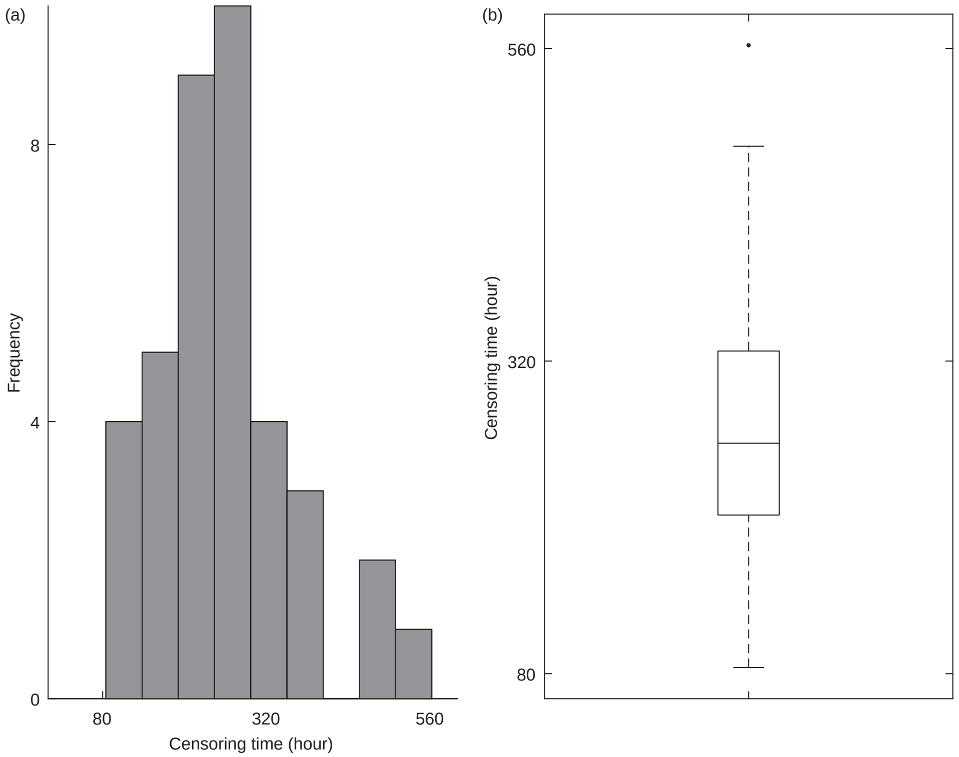

**Figure 1.** The censoring times (cumulative driving time) of the teenagers with CNC events in NTDS: (a) histogram, and (b) boxplot.

**Table 1.** The cumulative average driving time at the end of several months per teenager for NTDS.

| Month | 4 | 6 | 8 | 10 | 11 | 14 |
|---|---|---|---|---|---|---|
| Average driving time (hour) | 55.07 | 82.29 | 109.32 | 139.16 | 153.80 | 193.94 |

A driver might encounter multiple CNC events during the study. Among the 42 teenagers, 38 experienced a total of 279 CNC events (242 near-crashes), while four had no events. The average driving distance of the participant teenagers is 366.6 miles per month. The cumulative driving time at each event can be calculated from the data.

The 38 drivers with CNC events are included in the analysis because the drivers without events cannot provide information for the change-point. The cumulative driving time of a driver is used as the censoring time for the driver. For drivers with at least one CNC event, Figure 1 shows the histogram and boxplot of censoring time which varies from 84.70 to 562.47 hours, with a mean of 263.07 and standard deviation (SD) of 102.91 hours. As can be seen, the total driving time varies substantially among drivers during the same calendar period. Therefore, driving time is a more relevant measure of driving experience than calendar time. Table 1 shows the cumulative average driving time at the end of several months per teenager.

## 3. The hierarchical BFMM for change-point detection

We propose a Bayesian latent class change-point model to detect the change-point of novice teenage driving risk. Li *et al.* [40] developed change-point detection models assuming that all the drivers share the same change-points. However, the patterns of risk change vary among the teenagers and subgroups exist [26]. The models with identical change-points assumption cannot address the heterogeneity among drivers. Correspondingly, we develop a hierarchical model based on BFMM where the drivers fall in several clusters.

### *3.1. The FMM for change-point detection*

We assume that the CNC events are generated from Poisson processes. Let $\{N(t), t \geq 0\}$ be the number of events at or before time $t$, $N(t)$ is a counting process. The process history $H(t)$ is defined to be $H(t) = \{N(x) : x \in [0, t), t \geq 0\}$. The intensity function of the counting process is

$$\rho(t|H(t)) = \lim_{\nabla \to 0} P(N[t, t + \nabla) = 1|H(t))/\nabla,$$

where $\nabla$ is a short interval and $N[t, t + \nabla)$ is the number of events during the time interval $[t, t + \nabla)$ [10, Chapter 2, p. 28]. Assuming that two events cannot occur at the same time, a counting process is a Poisson process when the intensity function $\rho(t|H(t))$ is a non-negative integrable function $\lambda(t)$. The cumulative intensity function is $\Lambda(t) = \int_0^t \lambda(x)\,\mathrm{d}x, t > 0$.

Assuming there is at least one CNC event per teenager, we denote the total number of events for the $j$th driver to be $n_j, j = 1, \ldots, m$, where $m$ is the total number of drivers. The ordered times of these events are $t_{j1}, t_{j2}, \ldots, t_{jn_j}$. The total number of events for all drivers during the study period is $N = \sum_{j=1}^{m} n_j$. The total driving time for the $j$th subject is $C_j$ during the study, which is used as the censoring time.

$N_j(t)$ is defined to be the number of events for each driver to time $t$, and is assumed to be an NHPP with cumulative intensity $\Lambda_j(t)$. $N_j(t)$ follows a Poisson distribution with parameter $\Lambda_j(t)$: $N_j(t) \sim \text{Poisson}(\Lambda_j(t))$. The intensity function $\lambda(t)$ of an NHPP can increase or decrease abruptly and the shift point in time is the change-point. Lawless and Zhan [37] showed that a piecewise-constant function was a good alternative when the form of the intensity function was unknown. Accordingly, we assume the intensity function $\lambda_j(t)$ of the NHPP is piecewise constant with a single unknown change-point $\tau_j \in (0, C_j)$, $j = 1, 2, \ldots, m$. By denoting $\lambda_{jb}$ and $\lambda_{ja}$ as the intensity rates before and after the change-point for the $j$th driver, the intensity function is

$$\lambda_j(t) = \lambda_{jb} I(0 \le t \le \tau_j) + \lambda_{ja} I(t > \tau_j),$$

where $I(t)$ is the indicator function. By integrating the above equation from time 0 to $t$, the cumulative intensity function $\Lambda_j(t)$ is $\lambda_{jb} t$ if $t \le \tau_j$, and is $\lambda_{jb}\tau_j + \lambda_{ja}(t - \tau_j)$ if $t > \tau_j$. That is

$$\Lambda_j(t) = \lambda_{jb} t I(0 \le t \le \tau_j) + [\lambda_{jb}\tau_j + \lambda_{ja}(t - \tau_j)] I(t > \tau_j). \tag{1}$$

Drivers fall in $K$ clusters with unique change-point values $\mu_1, \mu_2, \ldots, \mu_K$. The intensity rates are $(\lambda_{b1}, \lambda_{a1}), \ldots, (\lambda_{bK}, \lambda_{aK})$, where $\lambda_{bk}$ and $\lambda_{ak}$ denote the intensity rates before and after the change-point of the $k$th cluster, respectively. The drivers in the same cluster share identical change-point and intensity rates.

Therefore, the change-point $\tau_j$ comes from a finite mixture with $K$ components:

$$f(\tau_j | \boldsymbol{\pi}_j, \boldsymbol{\mu}) = \sum_{k=1}^{K} \pi_{jk} \delta_{\mu_k}(\tau_j),$$

where the mixing proportion $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \ldots, \pi_{jK}), 0 \le \pi_{jK} \le 1, \sum_{k=1}^{K} \pi_{jk} = 1; \boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K); \delta_{\mu_k}(\tau_j)$ is a degenerate measure with probability 1 when $\mu_k = \tau_j$ and 0 otherwise.

Let $z_j$ be the latent membership of the $j$th driver, $z_j \in \{1, 2, \ldots, K\}$. Given that $z_j = k$, indicating that the $j$th driver is in the $k$th cluster, $\tau_j = \mu_k, \lambda_{jb} = \lambda_{bk}, \lambda_{ja} = \lambda_{ak}$. This latent variable indicates which cluster the driver is from. $z_j$ follows $\text{GeneralizedBernoulli}(\boldsymbol{\pi}_j)$ with the point mass function $f(z_j | \boldsymbol{\pi}_j, K) = \prod_{k=1}^{K} \pi_{jk}^{\delta_k(z_j)} = \sum_{k=1}^{K} \pi_{jk}\delta_k(z_j)$ [50, p. 35].

The mixing proportion $\boldsymbol{\pi}_j | \alpha_0, K$ follows $Dir(\alpha_0/K \cdot \mathbf{1}_K)$, where $\text{Dir}(\cdot)$ is a Dirichlet distribution [36, Chapter 49], $\alpha_0/K$ is the concentration parameter of the Dirichlet distribution, and $\mathbf{1}_K$ is a vector of ones with length $K$. The symmetric Dirichlet distribution is chosen because of no prior information on mixing proportions. $K > 1$ is the number of clusters. The support of a Dirichlet distribution is an open $(K-1)$-dimensional simplex defined by $\{x_k \in (0,1), \sum_{k=1}^{K} x_k = 1\}$. The Dirichlet distribution is the conjugate prior of a multinomial distribution. Define indicator variable $i_{jk} = 1$ if $z_j = k$ and $i_{jk} = 0$ otherwise. Hence $\mathbf{i}_j = (i_{j1}, i_{j2}, \cdots, i_{jK})$ follows a multinomial distribution $\text{multinomial}(1, \boldsymbol{\pi}_j)$.

We assign $G_0(\theta)$ as the prior of the change-point value for the $k$th cluster $\mu_k$, where $G_0(\theta)$ is a continuous uniform distribution $\text{Unif}(0, \theta)$ and $\theta$ is a fixed upper bound. $\theta$ can be chosen based on the censoring times and prior knowledge of the possible values of the change-points. This method assumes that $\mu_k$ is independent from the intensity rates $\lambda_{bk}$ and $\lambda_{ak}$, and also that $\lambda_{bk}$ and $\lambda_{ak}$ are conditionally independent given $\mu_k$. Considering Gamma priors for $\lambda_{bk}$ and $\lambda_{ak}$ given $\mu_k$, $\lambda_{bk}|\mu_k \sim \text{Gamma}(a_{1k}, b_{1k})$, $\lambda_{ak}|\mu_k \sim \text{Gamma}(a_{2k}, b_{2k})$, where $\text{Gamma}(a, b)$ is a Gamma distribution with mean $a/b$ and variance $a/b^2$. The prior of the scaled concentration parameter for Dirichlet distribution $\alpha_0$ is $f(\alpha_0)$, which can be a gamma distribution. The joint distribution is

$$f(\lambda_{bk}, \lambda_{ak}, \mu_k) \propto f(\lambda_{bk}, \lambda_{ak}|\mu_k)f(\mu_k) \propto \lambda_{bk}^{a_{1k}-1} exp(-b_{1k}\lambda_{bk})\lambda_{ak}^{a_{2k}-1} exp(-b_{2k}\lambda_{ak})f(\mu_k),$$

$\lambda_{bk} > 0, \lambda_{ak} > 0$.

By denoting $\boldsymbol{\lambda}_j = (\lambda_{jb}, \lambda_{ja})$ and $\boldsymbol{X}_j = (t_{j1}, t_{j2}, \cdots, t_{jn_j})$, the likelihood for the $j$th driver [69] can be expressed as

$$L_j(\boldsymbol{\lambda}_j, \tau_j|\boldsymbol{X}_j) = \exp[-\Lambda(C_j)] \prod_{i=1}^{n_j} \lambda_j(t_{ji}).$$

By denoting $\boldsymbol{\lambda}_b = (\lambda_{b1}, \lambda_{b2}, \ldots, \lambda_{bK})$, $\boldsymbol{\lambda}_a = (\lambda_{a1}, \lambda_{a2}, \ldots, \lambda_{aK})$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)$, the complete likelihood for the $j$th driver is then:

$$\begin{aligned} L_j(\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\mu}, \boldsymbol{\pi}_j|z_j, \boldsymbol{X}_j, K) &= \left\{ \exp[-\Lambda(C_j)] \prod_{i=1}^{n_j} \lambda_j(t_{ji}) \right\} f(z_j|\boldsymbol{\pi}_j, K) \\ &= \sum_{k=1}^{K} \exp[-\lambda_{bk}\mu_k - (C_j - \mu_k)\lambda_{ak}]\lambda_{bk}^{N_j^{(b)}} \lambda_{ak}^{N_j^{(a)}} \pi_{jk}\delta_k(z_j) \\ &= \prod_{k=1}^{K} \{\exp[-\lambda_{bk}\mu_k - (C_j - \mu_k)\lambda_{ak}]\lambda_{bk}^{N_j^{(b)}} \lambda_{ak}^{N_j^{(a)}} \pi_{jk}\}^{\delta_k(z_j)}, \end{aligned}$$

where $N_j^{(b)}$ is the number of events for the $j$th driver before change-point $\tau_j$ and $N_j^{(a)}$ after.

Let $S_k$ be the index set of drivers in the $k$th cluster, the number of drivers in the $k$th cluster be $m_k$, and the event times of the drivers in the $k$th cluster be $\boldsymbol{X}_{(k)}$. The complete likelihood for the drivers in the $k$th cluster is

$$\begin{aligned} &L_{(k)}(\lambda_{bk}, \lambda_{bk}, \mu_k, \boldsymbol{\pi}_{(k)}|\boldsymbol{X}_{(k)}) \\ &\quad = \exp\left\{ -(\lambda_{bk} - \lambda_{ak})m_k\mu_k - \lambda_{ak} \sum_{j \in S_k} C_j \right\} \lambda_{bk}^{\sum_{j \in S_k} N_j^{(b)}} \lambda_{ak}^{\sum_{j \in S_k} N_j^{(a)}} \prod_{j \in S_k} \pi_{jk}. \end{aligned}$$

Let $\boldsymbol{\Pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_m)$, $\boldsymbol{z} = (z_1, z_2, \ldots, z_m)$, and $\boldsymbol{X}$ be the event times of all drivers, the joint posterior distribution is

$$f(\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\mu}, \boldsymbol{\Pi}, \alpha_0|\boldsymbol{z}, \boldsymbol{X}) \propto \left\{ \prod_{j=1}^{m} L_j(\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\mu}, \boldsymbol{\pi}_j|z_j, \boldsymbol{X}) \right\} f(\boldsymbol{\lambda}_b)f(\boldsymbol{\lambda}_a)f(\boldsymbol{\mu})f(\boldsymbol{\Pi})f(\alpha_0). \quad (2)$$

Therefore, the full conditional distribution of $\boldsymbol{\pi}_j$ is

$$f(\boldsymbol{\pi}_j|\lambda_{jb}, \lambda_{ja}, K, \tau_j, \alpha_0, \boldsymbol{X}_j) = \text{Dir}(\alpha_0/K + i_{j1}, \alpha_0/K + i_{j2}, \ldots, \alpha_0/K + i_{jK}). \quad (3)$$

The full conditional distribution of $\mu_k$ is

$$f(\mu_k|\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)}) \propto G_0(\theta)L_{(k)}(\cdot). \tag{4}$$

In the Bayesian Markov chain Monte Carlo (MCMC) algorithm, we sample from the full conditional distribution of $z_j$:

$$f(z_j|\boldsymbol{\pi}_j, \boldsymbol{\mu}, \lambda_{bj}, \lambda_{aj}, \alpha_0, \boldsymbol{X}_j) \propto \sum_{k=1}^{K} L_j(\tau_j = \mu_k|z_j = k, X_j)\delta_k(z_j). \tag{5}$$

Based on Bayes' theorem, the conditional distributions given data of the $k$th cluster $\boldsymbol{X}_{(k)}$ are as follows:

$$\begin{aligned} f(\lambda_{bk}|\mu_k, \lambda_{ak}, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)}) &\propto f(\lambda_{bk}|\mu_k)L_{(k)}(\cdot) = \frac{\int_0^\infty f(\lambda_{bk}, \lambda_{ak}, \boldsymbol{\tau}_{(k)})\,\mathrm{d}\lambda_{ak}}{f(\mu_k)}L_{(k)}(\cdot) \\ &\sim \text{Gamma}\left(a_{1k} + \sum_{j\in S_k} N_j^{(b)}, b_{1k} + \sum_{j\in S_k}\tau_j\right), \qquad (6) \\ f(\lambda_{ak}|\mu_k, \lambda_{bk}, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)}) &\propto L_{(k)}(\cdot)\lambda_{ak}^{a_{2k}-1}e^{-b_{2k}\lambda_{ak}} \\ &\sim \text{Gamma}\left(a_{2k} + \sum_{j\in S_k} N_j^{(a)}, b_{2k} + \sum_{j\in S_k}(C_j - \tau_j)\right). \qquad (7) \end{aligned}$$

The full conditional distribution of the scaled concentration parameter $\alpha_0$ is

$$f(\alpha_0|\boldsymbol{\Pi}, K, \boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\mu}) \propto f(\alpha_0)f(\boldsymbol{\Pi}|\alpha_0, K) \propto f(\alpha_0)\prod_{j=1}^{m} f(\boldsymbol{\pi}_j|\alpha_0, K). \tag{8}$$

Mixture models cannot distinguish different components from the likelihood because of the label switching issues [7]. Once a sample from the posterior distribution is generated, it can be ordered for estimation [67].

### 3.2. Bayesian sampling from the full conditional distributions

The collapsed Gibbs sampler [41] combined with rejection sampling [53] is used to sample from the posterior distribution. The change-point values of the clusters are reordered in each iteration based on Stephens [67]. The initial values for the MCMC are as follows: $\boldsymbol{\mu}^{(0)}$ and $\alpha_0^{(0)}$ are randomly generated from their priors; the initial values of $\lambda_{bk}$ and $\lambda_{ak}$ are the average incident rates of each cluster before and after the change-point; the subjects are randomly assigned to the clusters. The algorithm is as follows. For $t = 1, 2, \ldots, B_t$ ($B_t$ is the number of MCMC iterations),

*Step 1* For $j = 1, 2, \ldots, m$, generate $\boldsymbol{\pi}_j^{(t)}$ from $f(\boldsymbol{\pi}_j|\lambda_{jb}, \lambda_{ja}, K, \tau_j, \alpha_0, \boldsymbol{X}_j)$ in Equation (3);

*Step 2* For $k = 1, 2, \ldots, K$, sample $\mu_k^{(t)}$ from $f(\mu_k|\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)})$ in Equation (4) using rejection sampling;

*Step 3* Arrange $\mu_k$'s in increasing order;
*Step 4* For $j = 1, 2, \ldots, m$, generate $z_j^{(t)}$ from $f(z_j|\boldsymbol{\pi}_j, \boldsymbol{\mu}, \lambda_{bj}, \lambda_{aj}, \alpha_0, \boldsymbol{X}_j)$ in Equation (5), and update $\tau_j^{(t)}$ and $\mathbf{i}_j^{(t)}$ based on $z_j^{(t)}$ and $\boldsymbol{\mu}^{(t)}$;
*Step 5* For $k = 1, 2, \ldots, K$, generate $\lambda_{bk}^{(t)}$ from $f(\lambda_{bk}|\mu_k, \lambda_{ak}, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)})$ in Equation (6); generate $\lambda_{ak}^{(t)}$ from $f(\lambda_{ak}|\mu_k, \lambda_{bk}, \boldsymbol{\pi}, \alpha_0, \boldsymbol{X}_{(k)})$ in Equation (7);
*Step 6* Generate $\alpha_0^{(t)}$ from $f(\alpha_0|\boldsymbol{\Pi}, K, \boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\mu}))$ in Equation (8) using rejection sampling.

Five chains with widely dispersed initial values are run until the Gelman–Rubin statistics are close to 1.0 for all scalar summaries [17]. The number of clusters is unknown. The DICs [65] of the models with different number of clusters are compared to determine the number of subgroups. According to Gelman *et al.* [16, p. 188], $DIC = 2\bar{D}(\boldsymbol{X}, \boldsymbol{\theta}) - D(\boldsymbol{X}, \bar{\boldsymbol{\theta}})$, where $\boldsymbol{X}$ is the data, $\boldsymbol{\theta} = (\boldsymbol{\lambda}_b, \boldsymbol{\lambda}_a, \boldsymbol{\tau})$, the deviance $D(\boldsymbol{X}, \boldsymbol{\theta}) = -2\log L(\boldsymbol{\theta}|\boldsymbol{X})$. $L(\boldsymbol{\theta}|\boldsymbol{X}) = \prod_{j=1}^{m} L_j(\lambda_b, \lambda_a, \tau_j|\boldsymbol{X}_j)$ is the likelihood of all drivers. The average of the $D(\boldsymbol{X}, \boldsymbol{\theta})$ over the samples of $\boldsymbol{\theta}$ after burn-in is calculated as $\bar{D}(\boldsymbol{X}, \boldsymbol{\theta})$. $D(\boldsymbol{X}, \bar{\boldsymbol{\theta}})$ is the deviance at the posterior mean. The parameter estimates are the posterior mean after burn-in. Each subject is assigned to the cluster with the maximum class membership probability.

## 4. Simulation studies

We conduct simulation studies to examine the model performance under different configurations. Data generation from an NHPP with piecewise-constant intensity functions is based on the inter-event times' distribution [35]. Given the previous event times $Y_0 = y_0 = 0, Y_1 = y_1, \ldots, Y_{i-1} = y_{i-1}$, the $i$th inter-event time $X_i = Y_i - Y_{i-1}$ for each subject has the cumulative density function (CDF):

$$\begin{aligned} F_i(x) &= \Pr[X_i \leq x | Y_p = y_p, p = 1, 2, \ldots, i-1] \\ &= 1 - \exp[\Lambda(y_{i-1}) - \Lambda(y_{i-1} + x)], \end{aligned} \tag{9}$$

where $\Lambda(\cdot)$ is specified in Equation (1) in Section 3.1.

The event time $y_{i+1}$ is the summation of the $(i+1)$th inter-event time with CDF $F_i$ and the $i$th event time. The algorithm to generate data from an NHPP is as follows:

*Step 1* $i = 0$, set $y_i = 0$;
*Step 2* Sample from $F_{i+1}$ to obtain $x_{i+1}$;
*Step 3* $y_{i+1} = y_{(i)} + x_{i+1}$;
*Step 4* $i = i+1$, return to Step (2).

Repeat the above process until $y_{i+1}$ is larger than the censoring time for each individual. The event times are $y_1, y_2, \ldots, y_i, i = 1, 2, \ldots, n_j$ for the $j$th subject.

### 4.1. Simulation configurations

Nine settings with different change-points, intensity rates, mixing proportions, sample sizes and number of clusters are considered. The nine settings are listed in Table 2.

**Table 2.** Nine settings for the simulation.

| Setting | Description |
|---|---|
| 1 | There are three clusters ($K = 3$). The total number of drivers $m = 40$. Each driver has the same probability to be from different clusters: $\pi_{j1} = \pi_{j2} = \pi_{j3} = 1/3$. The change-point values of the three clusters are: $\mu_1 = 110, \mu_2 = 220, \mu_3 = 330$. The intensity rates before and after the change-points are: $(\lambda_{b1}, \lambda_{a1}) = (100, 200), (\lambda_{b2}, \lambda_{a2}) = (250, 150), (\lambda_{b3}, \lambda_{a3}) = (300, 200)$. |
| 2 | It is the same as Setting 1 except that the first two clusters have identical intensity rates: $\lambda_{b1} = \lambda_{b2}, \lambda_{a1} = \lambda_{a2}$. |
| 3 | It is the same as Setting 1 except that the first two clusters have identical change-points: $\mu_1 = \mu_2$. |
| 4 | It is the same as Setting 1 except that there are 80 subjects: $m = 80$. |
| 5 | It is the same as Setting 1 except that the mixing proportions are not equal: $(\pi_{j1}, \pi_{j2}, \pi_{j3}) = (10\%, 45\%, 45\%)$. |
| 6 | The parameter values are close to the estimates of NTDS application in Section 5, $\pi_{j1} = \pi_{j2} = \pi_{j3} = \frac{1}{3}$, and $m = 40$. |
| 7 | There are two clusters: $K = 2$, $\mu_1 = 110$, $\mu_2 = 330$, $(\lambda_{b1}, \lambda_{a1}) = (100, 200)$, $(\lambda_{b2}, \lambda_{a2}) = (300, 200)$, $m = 40$, $\pi_{j1} = \pi_{j2} = \frac{1}{2}$. |
| 8 | It is the same as Setting 1 except that the censoring time of each driver $C_j$ is generated uniformly from $\tau_j + 10$ to 550. |
| 9 | It is the same as Setting 6 except that the censoring time of each driver $C_j$ is generated uniformly from $\tau_j + 10$ to 550. |

**Table 3.** Simulation results for Settings 1–3. The number of drivers $m = 40$. The censoring time $C_j \sim \text{Unif}(450, 500)$. The number of clusters $K$ is 3. Each driver has the same probability to be from different clusters: $\pi_{j1} = \pi_{j2} = \pi_{j3} = \frac{1}{3}$. The change-point values are $\mu_k$, the intensity rates before and after the change-points are $(\lambda_{bk}, \lambda_{ak})$, $k = 1,2,3$. In Settings 2–3, the first two clusters share identical change-points or intensity rates.

| Setting | Parameter | True value | Average of estimates | RMSE | \|Bias(%)\| | Coverage probability (%) |
|---|---|---|---|---|---|---|
| 1 | $\mu_1$ | 110 | 110.31 | 6.82 | 0.28 | 96.5 |
| | $\mu_2$ | 220 | 220.76 | 9.46 | 0.34 | 97.0 |
| | $\mu_3$ | 330 | 329.74 | 14.06 | 0.08 | 95.0 |
| | $\lambda_{b1}$ | 100 | 100.93 | 17.26 | 0.93 | 91.5 |
| | $\lambda_{a1}$ | 200 | 198.91 | 7.19 | 0.54 | 93.5 |
| | $\lambda_{b2}$ | 250 | 248.22 | 16.54 | 0.71 | 91.0 |
| | $\lambda_{a2}$ | 150 | 151.36 | 10.92 | 0.91 | 93.0 |
| | $\lambda_{b3}$ | 300 | 300.48 | 9.86 | 0.16 | 92.5 |
| | $\lambda_{a3}$ | 200 | 198.47 | 11.51 | 0.76 | 94.0 |
| 2 | $\mu_1$ | 110 | 114.31 | 17.90 | 3.92 | 92.5 |
| | $\mu_2$ | 220 | 220.90 | 11.98 | 0.41 | 95.0 |
| | $\mu_3$ | 330 | 328.92 | 10.80 | 0.33 | 95.5 |
| | $\lambda_{b1}$ | 250 | 249.18 | 15.44 | 0.33 | 93.5 |
| | $\lambda_{a1}$ | 150 | 150.45 | 7.52 | 0.30 | 92.0 |
| | $\lambda_{b2}$ | 250 | 249.67 | 12.79 | 0.13 | 94.5 |
| | $\lambda_{a2}$ | 150 | 149.67 | 9.91 | 0.22 | 90.5 |
| | $\lambda_{b3}$ | 300 | 299.24 | 9.67 | 0.25 | 93.5 |
| | $\lambda_{a3}$ | 200 | 199.37 | 11.59 | 0.31 | 95.0 |
| 3 | $\mu_1$ | 110 | 107.68 | 6.28 | 2.11 | 92.5 |
| | $\mu_2$ | 110 | 112.48 | 9.55 | 2.25 | 92.5 |
| | $\mu_3$ | 330 | 329.84 | 9.48 | 0.05 | 95.0 |
| | $\lambda_{b1}$ | 100 | 99.73 | 6.03 | 0.27 | 96.5 |
| | $\lambda_{a1}$ | 200 | 199.08 | 5.05 | 0.46 | 92.5 |
| | $\lambda_{b2}$ | 250 | 247.39 | 10.42 | 1.04 | 94.0 |
| | $\lambda_{a2}$ | 150 | 149.81 | 4.90 | 0.13 | 90.5 |
| | $\lambda_{b3}$ | 300 | 300.61 | 5.84 | 0.20 | 96.0 |
| | $\lambda_{a3}$ | 200 | 199.92 | 7.04 | 0.04 | 97.0 |

Specifically, we compare the results from Setting 1 with other settings to check how different scenarios would affect the results. The total number of drivers $m$ for each data set is 40 except for Setting 4. The censoring times $C_j$ are uniform between 450 and 500 hours for Settings 1–7 and between $\tau_j + 10$ and 550 hours for Settings 8-9.

The priors and initial values for the MCMC are as follows: the prior of the change-points $G_0(\theta)$ is a continuous uniform distribution Unif$(0, 400)$; $f(\lambda_{bk}) \propto 1/\lambda_{bk}$ and $f(\lambda_{ak}) \propto 1/\lambda_{ak}$ are the priors for the intensity rates $\lambda_{bk}$ and $\lambda_{ak}$; $\alpha_0$ is 2.2, which is obtained from screening simulation; the initial values of the change-point $\mu_k$ is from Unif$(0, 400)$; the subjects are randomly assigned to the clusters; the initial values of $\lambda_{bk}$ and $\lambda_{ak}$ are the average incident rates of each cluster before and after the change-points.

Under each parameter setting, $B = 200$ data sets are generated. The method in Section 3 with the number of clusters from one to seven is applied to each data set and DICs are calculated to determine the number of clusters. The MCMC converges well based

**Table 4.** Simulation results for Settings 4–7. The number of drivers $m$ is 80 for Setting 4 and 40 for Settings 5-7. $C_j \sim$ Unif$(450, 500)$.

| Setting | Parameter | True value | Average of estimates | RMSE | \|Bias(%)\| | Coverage probability (%) |
|---|---|---|---|---|---|---|
| 4 | $\mu_1$ | 110 | 109.98 | 3.20 | 0.02 | 95.5 |
| | $\mu_2$ | 220 | 220.15 | 4.41 | 0.07 | 93.0 |
| | $\mu_3$ | 330 | 329.57 | 5.13 | 0.13 | 93.0 |
| | $\lambda_{b1}$ | 100 | 99.96 | 6.27 | 0.04 | 94.0 |
| | $\lambda_{a1}$ | 200 | 199.68 | 4.58 | 0.16 | 94.5 |
| | $\lambda_{b2}$ | 250 | 250.12 | 7.13 | 0.05 | 94.0 |
| | $\lambda_{a2}$ | 150 | 150.35 | 5.09 | 0.23 | 97.0 |
| | $\lambda_{b3}$ | 300 | 299.92 | 6.40 | 0.03 | 94.5 |
| | $\lambda_{a3}$ | 200 | 200.35 | 8.24 | 0.18 | 93.0 |
| 5 | $\mu_1$ | 110 | 108.50 | 16.74 | 1.37 | 96.5 |
| | $\mu_2$ | 220 | 215.75 | 5.76 | 1.93 | 93.0 |
| | $\mu_3$ | 330 | 328.36 | 13.79 | 0.50 | 95.5 |
| | $\lambda_{b1}$ | 100 | 111.28 | 27.62 | 11.28 | 90.5 |
| | $\lambda_{a1}$ | 200 | 199.17 | 13.04 | 0.41 | 93.5 |
| | $\lambda_{b2}$ | 250 | 254.43 | 10.13 | 1.77 | 93.0 |
| | $\lambda_{a2}$ | 150 | 150.88 | 12.00 | 0.59 | 91.5 |
| | $\lambda_{b3}$ | 300 | 299.80 | 8.96 | 0.07 | 93.0 |
| | $\lambda_{a3}$ | 200 | 198.39 | 10.12 | 0.81 | 95.5 |
| 6 | $\mu_1$ | 50 | 59.52 | 13.84 | 19.05 | 89.0 |
| | $\mu_2$ | 110 | 108.27 | 11.46 | 1.58 | 89.0 |
| | $\mu_3$ | 150 | 157.51 | 14.62 | 5.01 | 89.0 |
| | $\lambda_{b1}$ | 48 | 41.02 | 15.62 | 14.54 | 76.5 |
| | $\lambda_{a1}$ | 54 | 42.67 | 21.59 | 20.99 | 70.0 |
| | $\lambda_{b2}$ | 27 | 28.81 | 10.02 | 6.71 | 86.5 |
| | $\lambda_{a2}$ | 13 | 18.71 | 15.40 | 43.89 | 85.0 |
| | $\lambda_{b3}$ | 20 | 26.98 | 13.38 | 34.91 | 86.5 |
| | $\lambda_{a3}$ | 11 | 17.62 | 16.31 | 60.19 | 87.0 |
| 7 | $\mu_1$ | 110 | 110.23 | 3.87 | 0.21 | 95.0 |
| | $\mu_2$ | 330 | 329.29 | 5.66 | 0.21 | 95.0 |
| | $\lambda_{b1}$ | 100 | 100.15 | 6.40 | 0.15 | 96.5 |
| | $\lambda_{a1}$ | 200 | 200.32 | 4.97 | 0.16 | 95.5 |
| | $\lambda_{b2}$ | 300 | 300.12 | 6.32 | 0.04 | 97.5 |
| | $\lambda_{a2}$ | 200 | 200.71 | 7.64 | 0.35 | 98.5 |

Notes: The number of clusters $K$ is 3 for Settings 4–6 and 2 for Setting 7. The change-point values are $\mu_k$, the intensity rates before and after the change-points are $(\lambda_{bk}, \lambda_{ak})$, $k = 1, \cdots, K$. The mixing proportions are not equal for Setting 5: $(\pi_{j1}, \pi_{j2}, \pi_{j3}) = (10\%, 45\%, 45\%)$. The parameter values are close to the estimates of NTDS application in Section 5 for Setting 6.

on the Gelman–Rubin approach. We run $B_t = 15,000$ iterations for each MCMC chain, discard the first 5000 iterations, and thin the sample after burn-in by using every fifth observation. The percentages of the data sets with correctly identified number of clusters and the parameter estimates given correct number of clusters under each setting are recorded in Table 6. We average the estimates of the $B$ data sets, compute the root-mean-square error (RMSE) for a parameter $\tau$ by $\sqrt{(1/B)\sum_{k=1}^{B}(\hat{\tau}-\tau)^2}$ and |bias (%)| by $(1/B)\sum_{k=1}^{B}|\hat{\tau}-\tau|/\tau \times 100\%$, estimate the coverage probability by the 95% credible interval, and average the percentage of correctly grouped subjects. For the credible interval, we use the equal-tailed intervals for the change-points and the highest posterior density intervals for other parameters because the posterior of the change-points may not be unimodal.

### 4.2. Simulation results

The percentages of the data sets with correctly estimated number of clusters (%) and the average percentage of correctly grouped subjects (%) are listed in Table 6. Tables 3–5 show the estimation results for the nine settings. For Setting 1 where the parameters are widely dispersed, the RMSE is reasonable, the absolute percentage bias (%) is within 0.1%, and

**Table 5.** Simulation results for Settings 8–9.

| Setting | Parameter | True value | Average of estimates | RMSE | \|Bias (%)\| | Coverage probability (%) |
|---|---|---|---|---|---|---|
| 8 | $\mu_1$ | 110 | 109.46 | 6.90 | 0.49 | 94.0 |
| | $\mu_2$ | 220 | 218.78 | 9.85 | 0.55 | 96.5 |
| | $\mu_3$ | 330 | 327.92 | 15.17 | 0.63 | 94.0 |
| | $\lambda_{b1}$ | 100 | 103.32 | 16.45 | 3.32 | 90.0 |
| | $\lambda_{a1}$ | 200 | 200.18 | 9.61 | 0.09 | 92.5 |
| | $\lambda_{b2}$ | 250 | 253.44 | 16.80 | 1.38 | 94.0 |
| | $\lambda_{a2}$ | 150 | 152.06 | 10.93 | 1.37 | 94.0 |
| | $\lambda_{b3}$ | 300 | 301.50 | 9.95 | 0.50 | 93.5 |
| | $\lambda_{a3}$ | 200 | 199.03 | 12.87 | 0.48 | 93.0 |
| 9 | $\mu_1$ | 50 | 41.03 | 12.14 | 17.95 | 87.0 |
| | $\mu_2$ | 110 | 112.09 | 12.06 | 1.90 | 89.5 |
| | $\mu_3$ | 150 | 157.36 | 15.72 | 4.90 | 89.5 |
| | $\lambda_{b1}$ | 48 | 43.39 | 15.88 | 9.61 | 75.0 |
| | $\lambda_{a1}$ | 54 | 44.35 | 18.21 | 17.86 | 78.0 |
| | $\lambda_{b2}$ | 27 | 25.16 | 9.39 | 6.83 | 86.0 |
| | $\lambda_{a2}$ | 13 | 9.77 | 14.38 | 24.86 | 83.5 |
| | $\lambda_{b3}$ | 20 | 25.45 | 12.90 | 27.25 | 88.5 |
| | $\lambda_{a3}$ | 11 | 16.99 | 16.04 | 54.41 | 84.5 |

Notes: The number of drivers $m$ is 40 for Settings 8–9. The censoring time of each driver $C_j$ is generated uniformly from $\tau_j +$ 10 to 550. The number of clusters $K$ is 3. The change-point values are $\mu_k$, the intensity rates before and after the change-points are $(\lambda_{bk}, \lambda_{ak}), k = 1, \ldots, K$. The mixing proportions are the same: $(\pi_{j1} = \pi_{j2} = \pi_{j3}) = 1/3$. The parameter values are close to the estimates of NTDS application in Section 5 for Setting 9.

**Table 6.** Two percentages for all the simulation settings.

| Setting | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Correctly estimated number of clusters (%) | 93.0 | 90.5 | 88.5 | 93.5 | 91.0 | 84.0 | 94.0 | 89.5 | 82.0 |
| Average percentage of correctly grouped subjects (%) | 96.23 | 97.45 | 98.51 | 97.21 | 94.35 | 80.28 | 99.83 | 95.76 | 80.48 |

the coverage probabilities are close to 95.0 %. For Settings 2–3 where two clusters share the same change-point value or intensity rates, the estimations are accurate, however, the percentages of the data sets with correctly estimated number of clusters are smaller compared with Setting 1. For Setting 4 when the sample size $m$ is larger than Setting 1, the RMSE is much smaller. For Setting 5, both the RMSE and |bias (%)| are larger for the first cluster, which has the smallest mixing proportion. For Setting 6, the change-points and intensity rates are less dispersed and smaller, the |bias (%)| are much larger and the coverage probability are lower than that of previous settings. The results are satisfactory for Setting 7 with two clusters. For Setting 8 when the censoring time has larger variation, the RMSE and |bias (%)| are a little larger than those in Setting 1. For Setting 9, the results are close to that in Setting 6.

Overall, when the difference in change-points or intensity rates among clusters are larger, or when the change-points and intensity rates are more dispersed and larger, the estimates will be more accurate. When the mixing proportion of the cluster is smaller, the estimate will be less stable. Larger sample size results in smaller variation in the estimators. The estimation and clustering tend to be a little less accurate under larger variation in censoring time, but trivial especially when the change-points and intensity rates are less dispersed and smaller. The model gives fine results under different scenarios, and is not sensitive to priors and initial values.

## 5. Application to the NTDS

The model is applied to the NTDS data. The 38 teenagers with CNC events are considered.

### 5.1. NTDS results

The prior of $\mu$'s is a continuous uniform distribution Unif$(0, 263)$. We choose 263 as the upper bound of the change-points because the average of the censoring time for the 38 teenagers is 263.07 hours. Based on preliminary NTDS application results, the model is not sensitive to other hyper-parameters. To obtain faster convergence, other prior settings are: $\lambda_{bk} \sim \text{Gamma}(30, 1)$, $\lambda_{ak} \sim \text{Gamma}(10, 1), k = 1, 2, \ldots, K, \alpha_0 \sim \text{Gamma}(2.6, 1)$.

The initial values of $\mu_2$ till $\mu_K$ are sampled uniformly from 0 to 263. $\mu_1^{(0)}$ is set to be smaller than the minimum of the censoring times because the change-point of each subject must be smaller than his or her censoring time to guarantee the MCMC convergence. $\lambda_{bk}^{(0)}$, $\lambda_{ak}^{(0)}$ and $\mu_1^{(0)}$ are 30, 10 and 65, respectively, which are close to the estimated intensity rates and change-point of the identical intensity model with one change-point in [40]. Each subject is assigned randomly to the clusters. The initial value of $\alpha_0$ is 2.6.

The MCMC converges well based on the Gelman–Rubin approach. For each fixed number of clusters, we generate an MCMC chain of $B_t = 70,000$ iterations, discard the first 10,000 iterations, and use every sixth observation after burn-in.

The DICs are calculated for the number of clusters from one to seven. The model with three clusters is chosen for NTDS data. The DIC is 2493.6 when the number of clusters is one, 2464.3 for two clusters and 2378.9 for three clusters. When the number of clusters is larger than three, though the DIC can be smaller, the differences between the intensity estimates before change-point and after can be smaller than one CNC events per teenager per 1000 hours. We ignore such subtle difference in practice. The model with three clusters also has practical interpretation, and explain the variation in the driving risk pattern of a relatively small sample well.

Figure 2 shows the cumulative event plot vs. the estimated cumulative intensity functions of the three clusters. The events happened most intensively in the first cluster. The event rate of the drivers in the first cluster stays high through the study, while tends to decrease for other drivers. From the cumulative event plot in Figure 2(a), we can clearly see three different subgroups. Figure 2(b) shows the estimated cumulative intensity functions of the three clusters. The intensity rates before and after the change-point for the first cluster are close to each other, so the change-point is not obvious from the figure. The classification is reasonable and the estimation explains the risk pattern variation among clusters. Table 7 shows the posterior means, SDs, and 95% credible intervals for the model with three clusters. For the credible interval, we use the equal-tailed intervals for the change-points and the highest posterior density intervals for other parameters because the posterior of the change-points may not be unimodal. There are 14, 13 and 11 teenagers in the three clusters correspondingly. For the first cluster, the change-point is 52.30 hours and the intensity before the change-point is 48.94 CNC events per teenager per 1000 hours and 54.22 after the change-point. Comparing with the average cumulative driving time per month per teenager in Table 1, the change of the first cluster happened during the fourth month. For

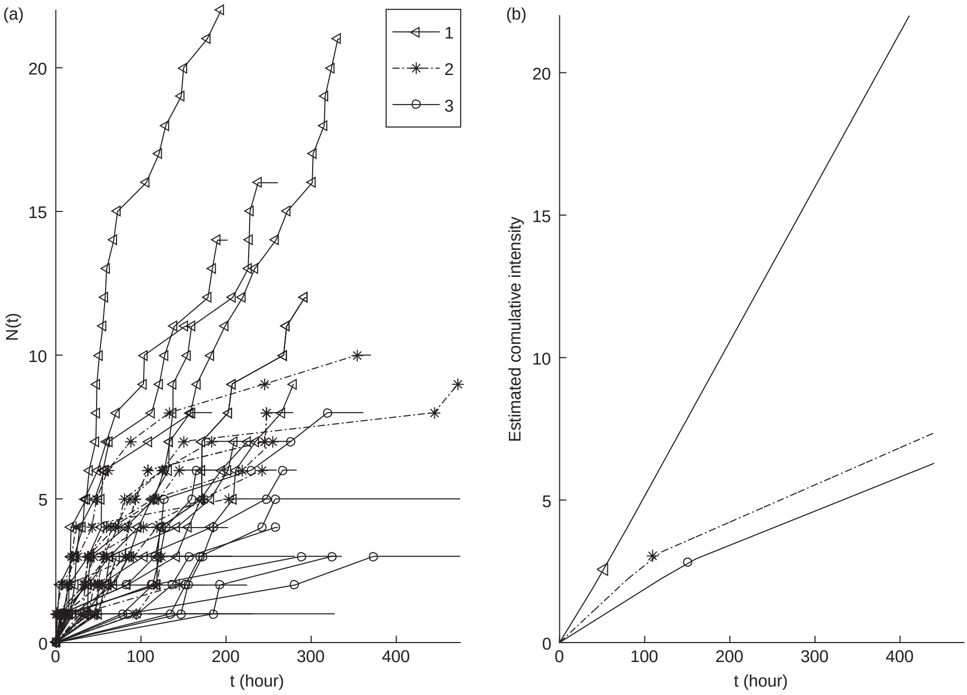

**Figure 2.** The model with three clusters: (a) The cumulative event plot of NTDS. The triangle, asterisk, and circle mark the event times of the three clusters. (b) The estimated cumulative intensity functions of the three clusters. The triangle, asterisk, and circle show the estimated change-points of each cluster.

**Table 7.** Posterior estimates of the model with three clusters. The change-point values are $\mu_k$, the intensity rates before and after the change-points are $(\lambda_{bk}, \lambda_{ak})$, $k = 1,2,3$. $\alpha_0$ is the scaled concentration parameter of the Dirichlet distribution.

| Parameters | Posterior mean | Posterior SD | 95% credible interval |
|---|---|---|---|
| $\mu_1$ | 52.30 | 12.12 | (29.38, 67.80) |
| $\mu_2$ | 108.99 | 14.04 | (59.06, 113.70) |
| $\mu_3$ | 150.20 | 17.23 | (114.24, 177.53) |
| $\lambda_{b1}$ | 48.94 | 10.93 | (28.37, 67.99) |
| $\lambda_{a1}$ | 54.22 | 5.60 | (44.35,64.37) |
| $\lambda_{b2}$ | 27.89 | 9.86 | (9.72, 44.42) |
| $\lambda_{a2}$ | 13.05 | 3.68 | (6.56,19.55) |
| $\lambda_{b3}$ | 18.76 | 7.05 | (7.17, 31.47) |
| $\lambda_{a3}$ | 11.98 | 4.10 | (5.18,19.56) |
| $\alpha_0$ | 2.58 | 1.25 | (0.59, 4.94) |

the second cluster, the change-point is 108.99 hours and the intensity before the change-point is 27.89 and 13.05 after. The change of the second cluster happened during the eighth month. For the third cluster, the change-point is 150.20 hours and the intensity before the change-point is 18.76 and 11.98 after. The change of the third cluster happened during the 11th month. There are 156 CNC events out of 256 in the first cluster (61.9%) during 30.5% of the total driving time, and 42 events in the third cluster (16.4%) during 36.1% of the total driving time.

For the model with three clusters, the ratio of female vs. male are close to one in each cluster. Based on the Gelman–Rubin statistics, the convergence is fine. Figure 3 shows the posterior distributions of the parameters with kernel fitting. Most of the posterior distributions are unimodal and right skewed. For each driver, there are three estimated mixing proportions. Figure 4 shows the boxplot and histogram of the mixing proportion estimates

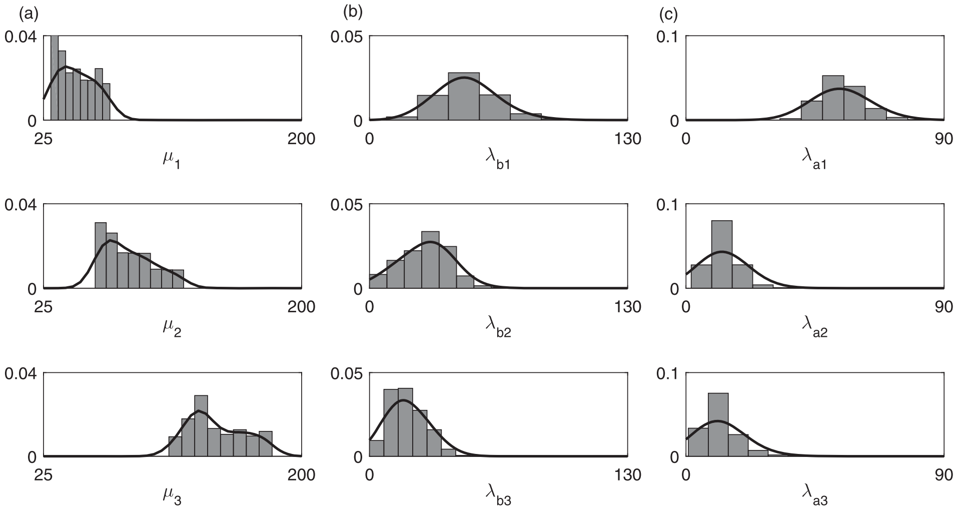

**Figure 3.** Posterior distributions of the parameters for the three clusters with kernel fitting: (a) the change-point, (b) the intensity rate before the change-point, and (c) the intensity rate after the change-point.

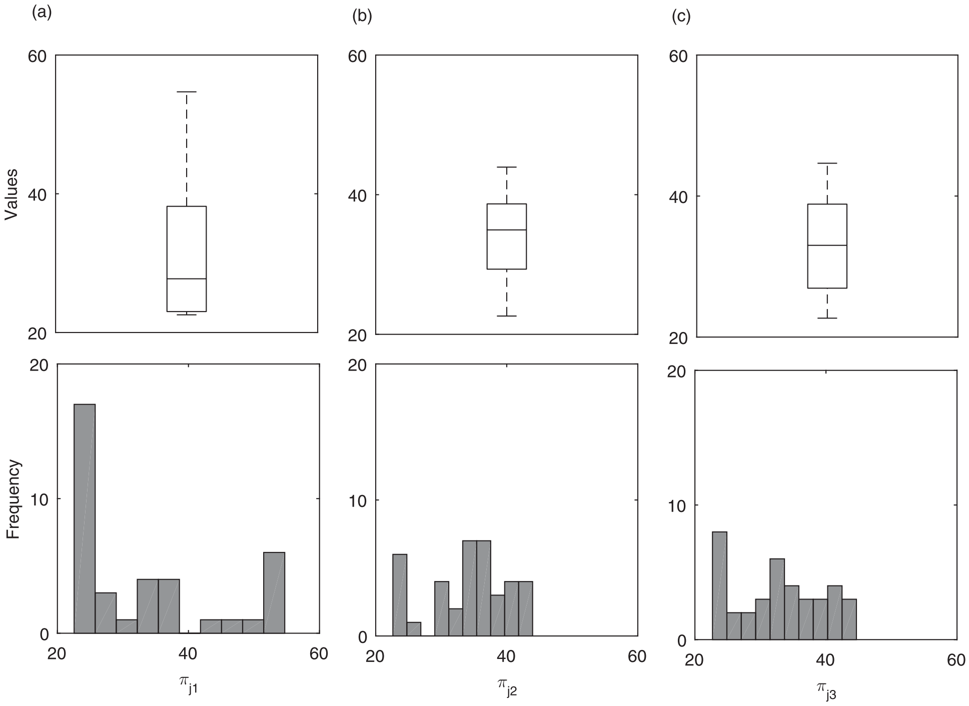

**Figure 4.** Boxplot and histogram of the mixing proportions for the 38 teenagers in NTDS: (a) the first cluster, (b) the second cluster, and (c) the third cluster.

of all the drivers. Each driver has three mixing proportions corresponding to the three clusters. Figure 5 shows the scatter plot of the mixing proportion median vs. maximum of each driver in NTDS. The straight line is the line when median equals maximum. For most of

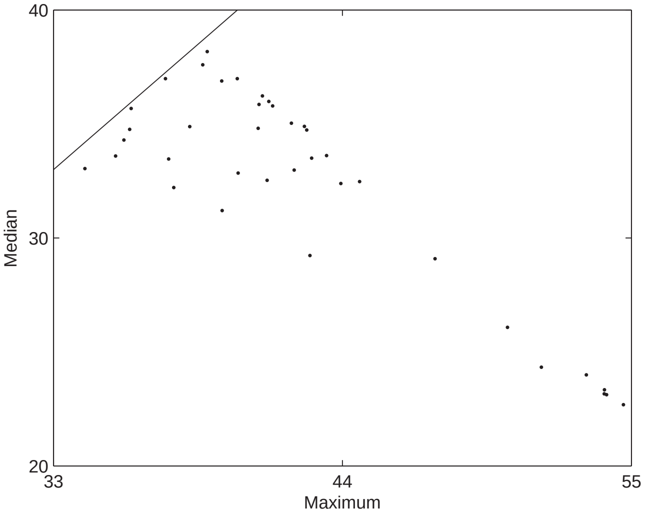

**Figure 5.** Scatter plot of the mixing proportion median vs. maximum for each driver in NTDS. The straight line is the line when median equals maximum for the driver.

**Table 8.** The contingency table of clustering results between Guo *et al.* [26] and the BFMM in this research.

| | K-means | | | |
|---|---|---|---|---|
| BFMM | Low-risk | Moderate-risk | High-risk | Total |
| Cluster 1 | 1 | 3 | 10 | 14 |
| 2 | 0 | 10 | 3 | 13 |
| 3 | 8 | 3 | 0 | 11 |
| Total | 9 | 6 | 13 | 38 |

the drivers, the maximum mixing proportion is much larger than the median, which makes the membership of the driver identifiable.

Overall, for the intensity rates before and after the change-point, it increases for the first cluster while decreases for the other clusters. The change-point of the first cluster is the smallest and the average intensity rate is the highest. The second cluster has the largest decrease in intensity rates after the change-point. The changes happened during the fourth, eighth, and eleventh month accordingly for the three clusters.

### 5.2. *Comparison with previous research*

We compare the results in this paper with Guo *et al.* [26]. Guo *et al.* [26] grouped the teenagers of NTDS into three clusters based on overall CNC rates by employing a K-means clustering method, and estimated the CNC rates over time by mixed effect Poisson models. This paper clusters the teenagers and detect the change in one model. Guo *et al.* [26] showed the change-points roughly by three-month intervals, while the method in the paper estimates the exact values of the change-points in actual driving time. Guo *et al.* [26] removed the two subjects with the highest CNC rates in the initial analysis because of the sensitivity of K-means clustering to outliers, while the presence of outliers do not affect the results in this paper. Table 8 shows the contingency table of the NTDS clustering results in [26] vs. in this research. The teenagers in the first cluster mainly fall in the high-risk group, and the teenagers in the second cluster mainly fall in the moderate-risk group, while the teenagers in the third cluster mainly fall in the low-risk group. Twenty-eight teenagers out of 38 are grouped similarly by the two methods. The average intensity rate for the first cluster by BFMM is the highest among the three clusters and the change happened at an earlier stage. It is reasonable that this cluster is corresponding to the high-risk group, and vice versa. Both methods show that the moderate-risk group has the largest decline in risk over time, mainly because of learning through experience. The change-points are also consistent between the two methods. The consistency between these two methods strengthens the validity of both methods while the method in this article provides more details of each cluster.

## 6. Conclusions

The driving risk of novice teenage drivers decreased significantly in the early period after licensure [26,45,73]. How much experience is required before relatively safe driving can be quantified by the change-points in crash rates. This study shows that the actual driving time of teenagers varies substantially during the same calendar period, and identifying the risk change-point in terms of driving hours is crucial.

Subgroups might exist among the teenagers with different patterns of risk change. Classifying the teenagers based on the risk pattern accommodates the heterogeneity among drivers and provides more insight in teenagers' driving behaviour. This paper is based upon a hierarchical BFMM. The simulation studies indicate that the model performs well under different scenarios and parameter settings.

We apply the proposed model to the NTDS data. The results show that the change-point model with three clusters provides the best model fitting. For the first cluster, the change-point is 52.30 hours and the average intensity rate is the highest among the three clusters. For the second cluster, the change-point is 108.99 hours. For the third cluster, the change-point is 150.20 hours. The intensity rates increase for the first cluster and decrease for the other clusters after the change-points. The changes happened during the fourth, eighth, and eleventh month accordingly for the three clusters. The results of this study suggest that to reduce the crashes risk of novice teenage driver populations, it is imperative to identify and develop appropriate education or safety counter measure tailored for the high-risk group. The results demonstrate the heterogeneity among drivers and provide key parameters for GDL and teenage drivers' safety training programs. These results are based on limited number of participants from a relative rural environment. Cautions should be used when generalizing the results to other populations.

There are several possible extensions of current study. We have assumed one change-point per cluster. Allowing more than one change-point or varying number of change-points across clusters would accommodate more complex data in practice. In addition, other forms of the intensity function can also be considered to increase the model flexibility.

We use model comparison approach to determine the best number of clusters. Teh [68] pointed out that an alternative was the Bayesian non-parametric approach. The non-parametric method to automatically detect the number of clusters with a Dirichlet Process prior [2] will be considered in the future. For future research, we would also like to incorporate covariates such as cortisol response [55], the presence of passengers [73], certain personality factors [9], total distance driven, and changes in external conditions.

The methodology has many applications across disciplines including the study of medical treatment effects, the hazard pattern of certain diseases, stock market analysis, customer behaviour change, and quality control.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported in part by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

## References

[1] T. Ando, *Bayesian Model Selection and Statistical Modeling (Statistics: A Series of Textbooks and Monographs)*, Chapman & Hall/CRC, Boca Raton, FL, 2010.

[2] C.E. Antoniak, *Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems*, Ann. Statist. 2 (1974), pp. 1152–1174.

[3] J.A. Aschar, S. Loibel, and M. Andrade, *Interfailure data with constant hazard function in the presence of change-points*, REVSTAT 5 (2007), pp. 209–226.

[4] A. Berg, R. Meyer, and J. Yu, *Deviance information criterion for comparing stochastic volatility models*, J. Bus. Econom. Statist. 22 (2004), pp. 107–120.

[5] J. Besag, P. Green, D. Higdon, and K. Mengersen, *Bayesian computation and stochastic systems*, Statist. Sci. 10 (1995), pp. 3–66.

[6] O. Cappé, C. Robert, and T. Rydén, *Reversible jump MCMC converging to birth-and-death MCMC and more general continuous time samplers*, J. Royal Statist. Soc. Ser. B 65 (2002), pp. 679–700.

[7] G. Celeux, M. Hurn, and C.P. Robert, *Computational and inferential difficulties with mixture posterior distributions*, J. Amer. Statist. Assoc. 95 (2000), pp. 957–970.

[8] Centers for Disease Control and Prevention. *Leading causes of death reports 1999–2005*, 2008, Available at http://webappa.cdc.gov/sasweb/ncipc/leadcaus10.html.

[9] S. Clarke and T.I. Robertson, *A meta-analytic review of the big five personality factors and accident involvement in occupational and non-occupational settings*, J. Occup. Organ. Psychol. 78 (2005), pp. 355–376.

[10] R.J. Cook and J.F. Lawless, *The Statistical Analysis of Recurrent Events*, Statistics for Biology and Health, Springer, New York, NY, 2007.

[11] H.A. Deery and B.N. Fildes, *Young novice driver subtypes: Relationship to high-risk behavior, traffic accident record, and simulator driving performance*, Hum. Fact.: J. Hum. Fact. Ergon. Soc. 41 (1999), pp. 628–643.

[12] J. Diebolt and C.P. Robert, *Estimation of finite mixture distributions through Bayesian sampling*, J. R. Statist. Soc. Ser. B 56 (1994), pp. 363–375.

[13] T.A. Dingus, S.G. Klauer, V.L. Neale, A. Petersen, S.E. Lee, J. Sudweeks, M.A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z.R. Doerzaph, J. Jermeland, and R.R. Knipling, *The 100-car naturalistic driving study: Phase ii – results of the 100-car field experiment*, Tech. Rep. DOT-HS-810-593, National Highway Traffic Safety Administration, Washington, DC, 2006.

[14] D. Frobish and N. Ebrahimi, *Parametric estimation of change-points for actual event data in recurrent events models*, Comput. Statist. Data Anal. 53 (2009), pp. 671–682.

[15] S. Frühwirth-Schnatter, *Finite mixture and Markov switching models*, Springer Series in Statistics, Springer, New York, NY, 2006.

[16] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, 2nd ed., Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[17] A. Gelman and D.B. Rubin, *Inference from iterative simulation using multiple sequences*, Stat. Sci. 7 (1992), pp. 457–511.

[18] L.A. Goodman, *Exploratory latent structure analysis using both identifiable and unidentifiable models*, Biometrika 61 (1974), pp. 215–231.

[19] Governor's Highway Safety Association. *State laws & funding: Virginia*, 2015, Available at http://www.ghsa.org/html/stateinfo/bystate/va.html.

[20] P. Green, *Reversible jump MCMC computation and Bayesian model determination*, Biometrika 82 (1995), pp. 711–732.

[21] M. Gruet, A. Philippe, and C. Robert, *MCMC control spreadsheets for exponential mixture estimation*, J. Comput. Graph. Stat. 8 (1999), pp. 298–317.

[22] F. Guo and Y.J. Fang, *Individual driver risk assessment using naturalistic driving data*, Accid. Anal. Prev. 61 (2013), pp. 3–9.

[23] F. Guo, Y.J. Fang, and J.F. Antin, *Evaluation of older driver fitness-to-drive metrics and driving risk using naturalistic driving study data*, Tech. Rep. 15-UM-036, Federal Highway Administration, 2015.

[24] F. Guo, Y.J. Fang, and J.F. Antin, *Older driver fitness-to-drive evaluation using naturalistic driving data*, J. Safety Res. 54 (2015), pp. 49–54.

[25] F. Guo, S.G. Klauer, J.M. Hankey, and T.A. Dingus, *Using near-crashes as a crash surrogate for naturalistic driving studies*, J. Transport. Res. Board 2147 (2010), pp. 66–74.

[26] F. Guo, B.G. Simons-Morton, S.E. Klauer, M.C. Ouimet, T.A. Dingus, and S.E. Lee, *Variability in crash and near-crash risk among novice teenage drivers: a naturalistic study*, J. Pediatrics 163 (2013), pp. 1670–1676.

[27] J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, Cambridge University Press, Cambridge, 2002.

[28] H. Ishwaran, L.F. James, and J.Y. Sun, *Bayesian model selection in finite mixtures by marginal density decompositions*, J. Amer. Statist. Assoc. 96 (2001), pp. 1316–1332.

[29] D.S. Karasoy and C. Kadilar, *A new bayes estimate of change-point in the hazard function*, Comput. Statist. Data Anal. 51 (2007), pp. 2993–3001.

[30] R.E. Kass and A.E. Raftery, *Bayes factors and model uncertainty*, Tech. Rep., University of Washington, Department of Statistics, Seattle, WA, 1993.

[31] S. Kim, Z. Chen, Z. Zhang, B.G. Simons-Morton, and P.S. Albert, *Bayesian hierarchical Poisson regression models: an application to a driving study with kinematic events*, J. Amer. Statist. Assoc. 108 (2013), pp. 494–503.

[32] S.G. Klauer, T.A. Dingus, V.L. Neale, J. Sudweeks, and D.J. Ramsey, *Comparing real-world behaviors of drivers with high vs. low rates of crashes and near-crashes*, Tech. Rep. DOT-HS-811-091, National Highway Traffic Safety Administration, Washington, DC, 2009.

[33] S.G. Klauer, F. Guo, B.G. Simons-Morton, M.C. Ouimet, S.E. Lee, and T.A. Dingus, *Distracted driving and risk of road crashes among novice and experienced drivers*, New Engl. J. Med. 370 (2014), pp. 54–59.

[34] S.G. Klauer, F. Guo, J.D. Sudweeks, and T.A. Dingus, *An analysis of driver inattention using a case-crossover approach on 100-car data*, Tech. Rep. DOT-HS-811-334, National Highway Traffic Safety Administration, Washington, DC, 2010.

[35] R.W. Klein and S.D. Roberts, *A time-varying Poisson arrival process generator*, Simulation 43 (1984), pp. 193–195.

[36] S. Kotz, N. Balakrishnan, and N.L. Johnson, *Continuous Multivariate Distributions. Volume 1*, John Wiley, New York, NY, 2000.

[37] J. Lawless and M. Zhan, *Analysis of interval-grouped recurrent-event data using piecewise constant rate functions*, Canad. J. Stat. 26 (1998), pp. 549–565.

[38] P. Lazarsfeld and N. Henry, *Latent Structure Analysis*, John Wiley, New York, NY, 1968.

[39] S.E. Lee, B.G. Simons-Morton, S.E. Klauer, M.C. Ouimet, and T.A. Dingus, *Naturalistic assessment of novice teenage crash experience*, Accid. Anal. Prev. 43 (2011), pp. 1472–1479.

[40] Q. Li, F. Guo, S. Klauer, and B. Simons-Monton, *Recurrent-event change-point models for novice teenage driving risk evaluation*, Comput. Statist. Data Anal., submitted, 2016.

[41] J.S. Liu, *The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem*, J. Amer. Statist. Assoc. 89 (2003), pp. 958–966.

[42] C.R. Loader, *Inference for a hazard rate change-point*, Biometrika 78 (1991), pp. 749–757.

[43] J.M. Marin, K. Mengersen, and C.P. Robert, *Bayesian modelling and inference on mixtures of distributions*, in *Bayesian Thinking: Modeling and computation*, Handbook of Statistics, Vol. 25, C. Rao and D. Dey, eds., Elsevier/North-Holland, Amsterdam, 2005, pp. 459–507.

[44] D. Matthews and V. Farewell, *On testing for a constant hazard against a change-point alternative*, Biometrics 38 (1982), pp. 463–468.

[45] D.R. Mayhew, H.M. Simpson, and A. Pak, *Changes in collision rates among novice drivers during the first months of driving*, Accid. Anal. Prev. 35 (2003), pp. 683–691.

[46] A.C. McCutcheon, *Latent Class Analysis*, Sage Publications, Newbury Park, CA, 1987.

[47] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York, 2000.

[48] K. Mengersen and C. Robert, *Testing for mixtures: A Bayesian entropic approach (with discussion)*, Bayesian Statistics, Vol. 5, Oxford University Press, Oxford, 1996.

[49] H.G. Müller and J.L. Wang, *Change-point models for hazard functions*, in *Change-Point Problems*, IMS Lecture Notes Monogr. Ser., Vol. 23, E. Carlstein, H.G. Muller, and D. Siegmund, eds., Institute of Mathematical Statistics, Hayward, CA, 1994, pp. 224–241.

[50] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, Boston, MA, 2012.

[51] C. Musselwhite, *Attitudes towards vehicle driving behavior: Categorizing and contextualizing risk*, Accid. Anal. Prev. 38 (2006), pp. 324–334.

[52] National Highway Traffic Safety Administration: *Traffic safety facts 2000: Young drivers*, Tech. Rep. DOT-HS-809-336, National Center for Statistics & Analysis, Washington, DC, 2000.

[53] R.M. Neal, *Slice sampling*, Ann. Statist. 31 (2003), pp. 705–767.

[54] A. Nobile, *Bayesian finite mixtures: A note on prior specification and posterior computation*, Tech. Rep. 05-3, University of Glasgow, Department of Statistics, Scotland, UK, 2007.

[55] M.C. Ouimet, T.G. Brown, F. Guo, S.G. Klauer, B.G. Simons-Morton, Y.J. Fang, S.E. Lee, C. Gianoulakis, and T.A. Dingus, *Higher crash and near-crash rates in teenaged drivers with lower cortisol response: An 18-month longitudinal, naturalistic study*, JAMA Pediatrics 168 (2014), pp. 517–522.

[56] S. Richardson and P.J. Green, *On Bayesian analysis of mixtures with an unknown number of components (with discussion)*, J. Roy. Statist. Soc. Ser. B 59 (1997), pp. 731–792.

[57] C.P. Robert, *The Bayesian Choice*, 2nd ed., Springer, New York, NY, 2001.

[58] K. Roeder and L. Wasserman, *Practical Bayesian density estimation using mixtures of normals*, J. Amer. Statist. Assoc. 92 (1997), pp. 894–902.

[59] G. Rolls, R.D. Hall, R. Ingham, and M. McDonald, *Accident Risk and Behavioral Patterns of Younger Drivers*, AA Foundation for Road Safety Research, Hampshire, UK, 1991.

[60] S.M. Ross, *Simulation*, Academic Press, San Diego, CA, 2006.

[61] S.K. Sahu and R.C.H. Cheng, *A fast distance-based approach for determining the number of components in mixtures*, Canad. J. Statist. 31 (2003), pp. 3–22.

[62] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist. 6 (1978), pp. 461–464.

[63] B.G. Simons-Morton, M.C. Ouimet, Z. Zhang, S.E. Lee, S.E. Klauer, J. Wang, P.S. Albert, and T.A. Dingus, *Crash and risky driving involvement among novice adolescent drivers and their parents*, Amer. J. Pub. Health 101 (2011), pp. 2362–2367.

[64] B.G. Simons-Morton, M.C. Ouimet, Z. Zhang, S.E. Lee, S.E. Klauer, J. Wang, R. Chen, P.S. Albert, and T.A. Dingus, *The effect of passengers and risk-taking friends on risky driving and crashes/near crashes among novice teenagers*, J. Adolesc. Health. 49 (2011), pp. 587–593.

[65] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. Van Der Linde, *Bayesian measures of model complexity and fit*, J. Roy. Statist. Soc. Ser. B 64 (2002), pp. 583–639.

[66] M. Stephens, *Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods*, Ann. Statist. 28 (2000), pp. 40–74.

[67] M. Stephens, *Dealing with label switching in mixture models*, J. R. Stat. Soc. Ser. B Stat. Methodol. 62 (2000), pp. 795–809.

[68] Y.W. Teh, *Encyclopedia of Machine Learning*, Springer, New York.

[69] W. Thompson, *Point Process Models with Applications to Safety and Reliability*, Chapman and Hall, New York, NY, 1988.

[70] D.M. Titterington, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixture Distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley, Chichester, 1985.

[71] P. Ulleberg, *Personality subtypes of young drivers. relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign*, Transp. Res. F: Traffic Psychol. Behav. 4 (2001), pp. 279–297.

[72] R. West and R. Odgen, *Continuous-time estimation of a change-point in a Poisson process*, JSCS 56 (1997), pp. 293–302.

[73] A.F. Williams, *Teenage drivers: Patterns of risk*, J. Safety Res. 34 (2003), pp. 5–15.

[74] Y.C. Yao, *Maximum likelihood estimation in hazard rate models with a change-point*, Commun. Statist. A. Theory Methods 15 (1986), pp. 2455–2466.