Taylor & Francis
Taylor & Francis Group

Journal of Transportation Safety & Security

ISSN: 1943-9962 (Print) 1943-9970 (Online) Journal homepage: https://www.tandfonline.com/loi/utss20

Applying Bayesian data mining to measure the effect of vehicular defects on crash severity

Subasish Das, Anandi Dutta & Srinivas Reddy Geedipally

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
https://doi.org/10.1080/19439962.2019.1658674

Applying Bayesian data mining to measure the effect of vehicular defects on crash severity

Subasish Das[a], Anandi Dutta[b], and Srinivas Reddy Geedipally[c]

[a]Texas A&M University Transportation Institute, College Station, Texas, USA; [b]Department of Computer Science and Engineering, Ohio State University, Columbus, Ohio, USA; [c]Texas A&M Transportation Institute, Arlington, Texas, USA

ABSTRACT

The National Motor Vehicle Crash Causation Survey (NMVCCS), conducted from 2005 to 2007, showed that an estimated 44,000 crashes occurred due to vehicular defects— 2% of the NMVCCS crashes. Vehicle defects have an adverse effect upon overall roadway safety as they can increase the likelihood of traffic crashes, thus increasing the frequency of crash-related injuries and fatalities. Even though Louisiana requires a biennial vehicular safety inspection, recent traffic crash statistics have shown a higher than average percentage of vehicle defect–related crash fatalities in Louisiana (3% of all traffic fatalities). This fact called for an in-depth analysis of the vehicle defect–related crashes in Louisiana. The current study used 7 years (2010–2016) of traffic crash data from Louisiana to investigate the association between crash severity and vehicle-defect types by applying a Bayesian data mining approach. The findings showed that vehicle age is associated with severe injury crashes. Worn tires and defective brakes are the over-represented vehicle-defect categories. The significant association patterns can be used by different stakeholders to enhance roadway safety and reduce vehicular defect associated crashes.

KEYWORDS
vehicle defect; vehicle age; vehicle type; crash severity; empirical Bayes data mining

1. Introduction

As vehicles age, their performance and safety can deteriorate and potentially result in dangerous operating conditions. Despite this risk, there have been a limited number of studies that have conducted an in-depth analysis of crashes caused by vehicular defects. In the U.S., 34 out of 50 states do not require annual or biennial vehicle safety inspections, and many of the states that do currently require safety inspections are reconsidering the impact of vehicle inspection policies. The purpose of the state-regulated vehicular inspections is to ensure that registered vehicles are roadworthy. These inspection programs identify vehicle defects associated with safety

CONTACT Subasish Das s-das@tti.tamu.edu Texas A&M Transportation Institute, College Station, TX 77843-0001, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/utss.

risks and provide the vehicle owner with proper guidance in making the defected vehicles roadworthy. However, a direct relationship between state inspection policies and the extent of vehicle defect–related crashes have not yet been established through research studies.

The National Motor Vehicle Crash Causation Survey (NMVCCS), conducted from 2005 to 2007, found that 2% of all crashes – estimated to be 44,000 – were caused by vehicular defects (USDOT, 2015). There is a significant research gap in understanding the effect of vehicle defects on crash outcomes. This study developed a database for 7 years (2010–2016) of Louisiana traffic crash data by using information about the condition of the vehicles involved to filter crashes with vehicle defect as the primary contributing factor. From 2010 to 2016, Louisiana experienced an increase of 1.5% of vehicle defect–related crashes; in this time period, 24,185 drivers were involved in vehicle defect–associated crashes.

The objective of this study is to determine the relationship between vehicle age, vehicle defect, vehicle type, and crash severity types. The general hypothesis is that the safety issues of vehicles are exacerbated over time due to increasing rates of defects. This study aims to examine this hypothesis with the inclusion of other surrogate measures (e.g., vehicle type, type of defects, and vehicle type). The Empirical Bayes Geometric Mean (EBGM) method was applied to determine the relationship between these variables. This paper demonstrates the value of using rule-based analysis methods that can identify associated risk subgroups in the dataset without imposing assumptions on the subgroups or method. The rules define the key contributing factor groups by, considering not only the interactions between the factors, but also their ranges. The proposed method showed promising results from this type of data set. As many states are reexamining the benefits of annual or biennial vehicular inspections, the findings of this study are imperative to help inform potential policy and legislative changes.

This paper is organized as follows. First, there is a review of the literature concerning the studies on vehicle defect–associated crashes. Next, the data preparation and descriptive statistics of the key variables are explained. The following section provides a brief overview of EBGM. The paper concludes with a discussion of results, conclusions, and recommendations for future research.

2. Earlier works and research context

There are a limited number of studies that have examined the relationship between vehicle defects and associated safety implications. Lloyd, Wallbank, Broughton, and Cuerden (2017) estimated the potential impact of vehicle secondary safety regulations in reducing vehicle defect–related crashes. In Brazil, the estimated fatality reduction would be in between 12,500 and

34,200 fatalities during 2015–2030. Moodley and Allopi (2008) performed a survey of parked vehicles to investigate whether the vehicles were adequately safe for driving (USDOT, 2015). This observational study found that 24% of these vehicles had tire defects, whereas 11% of them had light defects. This information may be important to determine if vehicle defects are contributing factors to crashes and, if so, which defects are the most prominent. In past studies, defects in the brake system or tires have been cited as the most frequent cause of crashes (Moodley & Allopi, 2008; USDOT, 1975). Wolf (1968) conducted a study in 1962 and found that 6% of truck crashes occurred due to mechanical failure, and vehicle defects caused at least 18% of the crashes. Other researchers have found, through crash reconstruction, that vehicle defects contributed to 9% of the crashes (Kinsey, 1976).

More recently, several studies have aimed to determine the potential vehicle-related factors that can contribute to crashes. Studies by Conroy et al. (2008) and Hoque and Hasan (2006) indicate that mechanical defects are the most common defect responsible for crashes. Cuerden et al. (2011) estimated that vehicle defects are likely to be a contributing factor in around 3% of crashes in the U.K. The findings also showed that reducing the frequency of testing of newer vehicles is likely to have adverse road safety consequences. Schoor and Niekerk (2001) attempted to establish the contribution of mechanical failures in traffic crashes. Data obtained from accident response units (ARU) indicated that tire and brake defects were the most significant contributors to mechanical failures resulting in crashes.

Similarly, Solah et al. (2017) found that the defects in the brake system, tires, lighting, body/paint work, and upper body part are the most common components that contributed to road crashes. This finding aligns with other studies such as one conducted by DEKRA Automobil GmbH (2015) in Germany which found that brake system defects, running gear defects, and tire defects are the key factors leading up to road crashes. DEKRA Automobil GmbH (2015) suggested that older vehicles are more likely to cause crashes due to lack of services and maintenance, and therefore, a periodic vehicle inspection is critical to road safety improvement. Many studies also concluded that the likelihood of crashes increased as the age of the vehicle being driven increased (Akloweg, Hayshi, & Kato, 2011; Al-Ghaweel, Mursi, Jack, & Joel, 2009; Islam & Kanitpong, 2008; Treat et al., 1979; Vaughan, 1992). This conclusion was later supported in the National Highway Traffic Safety Administration (NHTSA) report. Failure of tires/wheels and braking system are frequently cited as critical factors of increased crash risk (National Motor Vehicle Crash Causation Survey, 2018). Vehicle defects are also one of the 15 safety problems identified by NHTSA to be examined during the vehicular inspection (Green, Agent, Pigman, & Ross, 2018). Barry, Ginpil, and O'Neill (1999) found that an airbag system effectively promotes the likelihood of the vehicle occupants' survival in case of a crash. This implies that in the event of a crash, occupants in vehicles with defective airbag system would have less protection and a higher chance of trauma or critical injuries.

Despite ongoing improvements in vehicle safety, the rate of personal injuries and economic loss caused by traffic crashes remain unacceptably high. Therefore, there is a need for more research in new directions to combat the personal and economic toll that crashes take. Data mining and rules-based studies have been widely used by transportation safety researchers (Baireddy, Zhou, & Jalayer, 2018; Das & Sun, 2015; Das & Sun, 2016; Das, Avelar, Dixon, & Sun, 2018; Das, Brimley, Lindheimer, & Pant, 2017, Das, Minjares-Kyle, Avelar, Dixon, & Bommanayakanahalli, 2017; Das, Dutta, & Sun, 2019; Das, Dutta, Avelar et al., 2018; Das, Dutta, Jalayer, Bibeka, & Wu, 2018; Das, Kong, & Tsapakis, 2019; Das, Mudgal, Dutta, & Geedipally, 2018; Das, Sun, Wang, & Leboeuf, 2015; Factor, Yair, & Mahalel, 2010; Jalayer, Pour-Rouholamin & Zhou, 2018; Jalayer, Zhou, & Das, 2018; Weng & Li, 2019). This study applied a unique method to identify the association between crash injury and vehicle defect–associated characteristics.

3. Method

3.1. Data integration

This study presented an analysis of vehicle defect–involved crashes. The research team used 7 years (2010–2016) of police-reported crash data in Louisiana. The data are separated into several tables; the three major tables (crash table, vehicle table, and geometric table) contain a substantial amount of information. Louisiana crash data contain a variable that provides details about vehicular defects, and this information was used to prepare a database for this study that includes the vehicle defect–related attributes. The final data set contains data for 24,185 crashes in which vehicle defect is the primary contributing factor. The key objective of this study is to investigate the relationship between crash severity and vehicle defect–related characteristics. During the preliminary analysis, three vehicle-related variables were identified as the key target variables: vehicle type, vehicle defect, and vehicle age.

3.2. Descriptive statistics

Figure 1 illustrates the number of crashes attributed to various defects from the year 2010 to 2016. Based on the results, defects associated with brakes and tires are the leading factors that result in crashes; these findings align with previous studies (Solah et al., 2017; Vaughan, 1992). The frequency of

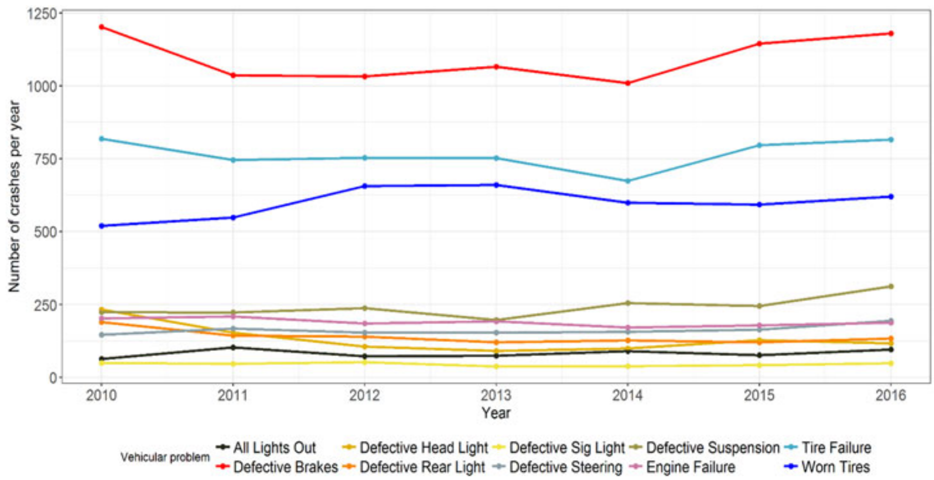

Figure 1. Vehicle defect associated crashes in Louisiana.

defective brake–related crashes showed a downward trend from 2010 to 2014, but after 2014 the frequency has trended upwards. A similar pattern is shown for tire failure–related crashes. Additionally, worn tire-related crashes show an increasing trend starting from 2010. Other significant defects include defective suspension, which causes around 250 crashes every year, as well as engine failure. The other vehicular defects (all lights out, defective lights, and defective steering) combined represent around 20% of all vehicular defect–associated crashes.

Two other databases have been used to make the comparison in contexts: (1) Fatality Analysis Reporting System (FARS) and (2) NHTSA Vehicle Complaint Database or VCD (FARS, 2018; National Highway Traffic Safety Administration, 2018). The NHTSA maintains FARS database, which is a census of all fatal traffic crashes in the U.S. that "provides uniformly coded, national data on police-reported fatalities, and contains information on crashes in which vehicle component failure was noted but is limited to crashes involving fatalities" (FARS, 2018). The research team will use 2012–2016 FARS data for this comparison. The VCD is based on consumer provided complaints collected by NHTSA. As of June 15, 2018, this database incorporates 1,416,390 complaint reports in a structured form with 33 variables. Around 7.26% of these reports involve some level of injury or fatalities. The complaints file in the VCD contains all safety-related defect complaints received by NHTSA since January 1, 1995. The number of reports associated with some level of injury or deaths is 67,201 (around 4.9% of total complaint entries). However, approximately 26% of those complaint reports are not associated with traffic crashes. These vehicular defect factors account for around 89% of the vehicular defect–related fatal crashes.

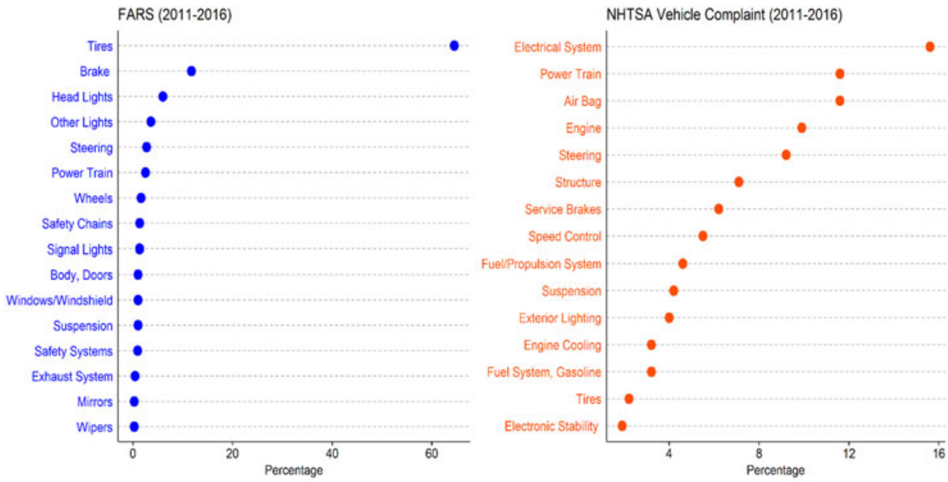

Figure 2. Comparison of major vehicular defects between Fatality Analysis Reporting System (FARS) and National Highway Traffic Safety Administration (NHTSA) database.

There are some significant differences between the FARS dataset and the VCD. The top five vehicle defects in the FARS data set are tires, brakes, headlights, other lights, and steering (see Figure 2). However, the top five vehicular defects in the VCD are electric system, powertrain, airbag, engine, and steering. All of these factors incorporate around 61% of the vehicular defect related fatal crashes. The powertrain is the second most significant factor in the VCD, though it only ranks as the sixth factor in the FARS data set. The top five vehicular defects in Louisiana database are defective brakes, tire failure, worn tires, defective suspension, and engine failure. The top five factors of Louisiana data represent 81% of all vehicular defect–related crashes, which is in line with the FARS findings.

The injury classification system (known as KABCO) divides crash severity into five major groups: (1) fatal injury (K), (2) incapacitating injury (A), (3) nonincapacitating injury (B), (4) minor injury (C), and (5) noninjury or property damage only (PDO) or O. Figures 2 to 4 present the severity distribution of crashes in Louisiana by vehicle type, vehicle defect, and vehicle age, respectively. Figure 3 shows that the proportions of KABC crashes associated with motorcycles are significantly higher than other vehicle types. Around 84% of these crashes involve some sort of injuries. For sport utility vehicles (SUV) , the KABC crashes represent 25% of all crashes. Among all types of vehicles, bus and truck show over-representations in PDO crashes. Figure 4 shows the distribution of crash severities by vehicular defects. Light-related vehicular defects (all lights out, defective headlight, and defective rear light) contribute to around 20% of KABC crashes. Tire defect related crashes (worn tires, and tire failures) show the highest proportions of injury-related crashes. The findings are in line with the previous findings (DEKRA Automobil GmbH,

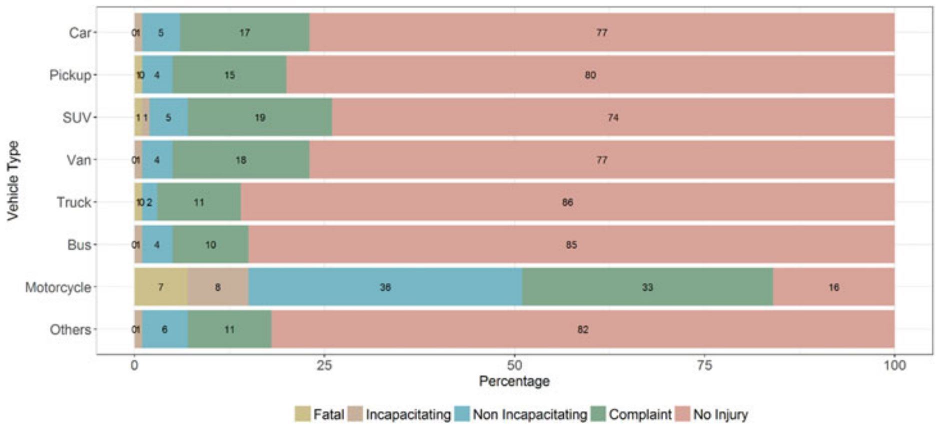

Figure 3. Severity distribution by vehicle type.

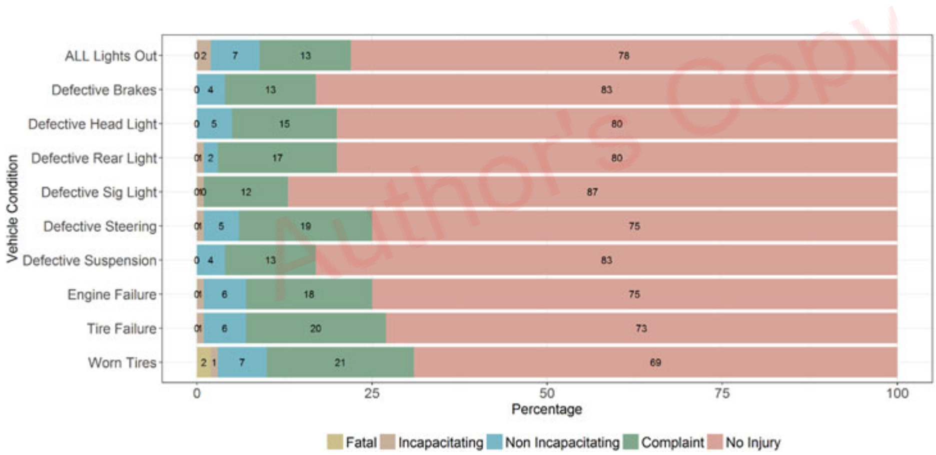

Figure 4. Severity distribution by vehicle condition.

2015; Schoor & Niekerk, 2001; Solah et al., 2017). One possible explanation for vehicle defects is that drivers are less aware of tire condition and lacking knowledge of vehicle performance (Kane, Liberman, DiViesti, & Click, 2018; Malaysian Institute of Road Safety Research, 2018). Engine failure and defective steering are also predominant in KABC crashes. The current study focused only on vehicle defects and associated crash outcomes. However, roadway, environment, and human factors play a significant role in the occurrences of crashes. It is expected that there are potential interactions between vehicle defects and other factors. However, many studies are designed to answer the research question by considering other variables as constant throughout the intervention period. For example, the *Highway Safety Manual* acknowledges the importance of inclusion of human factors in safety studies but considers only geometric variables in developing the safety performance functions (AASHTO, 2010).

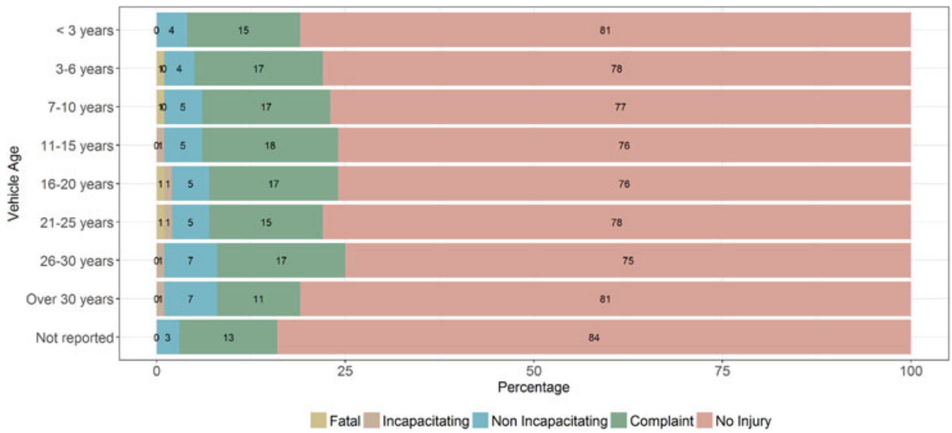

Figure 5. Severity distribution by vehicle age.

Figure 5 shows the proportions of crash severities for different ages of vehicles. The general concept is that the safety aspects of the vehicles decline over the years. The figure shows that the proportions of KABC crashes increase as the age of the vehicle increases (from fewer than 3 years to 16–20 years). However, this increase is not consistent with vehicles older than 20 years. One possible reason for this is that there is not enough data because vehicles aged over 20 years only represent 7% of all vehicle defect–related crashes. The findings are consistent with past studies (Al-Ghaweel et al., 2009; Akloweg et al., 2011; Islam & Kanitpong, 2008; Vaughan, 1992). It is generally anticipated that drivers are more likely to use vehicles with low ages for long trips which are more likely to result in crashes due to fatigue or drowsiness. The current study is not suitable to examine this hypothesis. However, the Strategic Highway Research Program 2 (SHRP 2) naturalistic driving study (NDS) data can be used to examine this hypothesis.

3.3. EBGM method

Contingency tables are usually analyzed by relative reporting ratio *RR* which is given by $^{N}/_{E}$, where N is the real frequency or count of the entries and E is the expected count assuming that rows and columns are independent. However, *RR* tends to exaggerate the significance of a count when the actual count is small. The association between vehicle characteristics (vehicle type, vehicle problem, and vehicle age) and crash severity (for example, incapacitating injury) were analyzed using the EBGM method, which is a rules mining method. Rules mining is a widely popular machine learning method, which is used for discovering interestingness from an unsupervised data set (dataset with no predefined response variable). The generated rules in rules mining methods describe the range of values on one or more risk factors that are associated with other attributes. Thus, rules provide a natural semantics to define the associated risk groups from an unsupervised data set. Another important advantage is that rules are scale independent in which data do not need any sort of scaling. Conventional statistical models are based on assumptions. Deviation from the assumption would make the results biased. For data mining procedure, no prior assumption is required. Das, Mudgal, Dutta, et al. (2018) used the EBGM method in a previous study to identify 'vehicle model with major defect' groups from the large and complex data set of NHTSA VCD. This mining tool helps to identify how significant the frequency of an outcome is given that it has a high occurrence. For example, it can help identify the significance of an occurrence of a severity type due to a certain defect in the vehicle given its model and model year. However, the result of this analysis will create direction to further investigate the impact on crash severity from the vehicle model, age, and the occurrence of the given corresponding defects. The analysis presented in this paper is performed using open source R package openEBGM, which estimates the significance of unusually large cell counts in large, sparse, contingency tables (Ihrie, Ahmed, & Poncet, 2017; R Core Team, 2013). This method can be used to find unusually high reporting rates of vehicle defects associated with the vehicle model.

EBGM can be considered as a Bayesian approach to *RR* because *EBGM* adds Bayesian shrinkage corrections to *RR* such that high values (resulting from low *E* and low *N*) are no longer identified as significant because small *N* could occur randomly. This property makes the current data a perfect fit for further exploration. The *EB* approach shrinks large *RR* toward 1 when *N* is small. The shrinkage is smaller for large counts and becomes negligible for very large counts. Therefore, *the EB* approach provides more stable results as compared to the conventional *RR* measurement.

3.3.1. Theory

This section presents a brief discussion of the theoretical concept of EBGM, which is mostly based on DuMouchel (1999). Interested readers can consult the conceptual paper developed by DuMouchel (1999) for more information. The EBGM approach not only provides interpretability of associated factors but also accommodates appropriate sampling variation. Consider a contingency table with number of cells N_{ij} that follows a Poisson distribution with unknown mean μ_{ij}. Here i and j represent row and column numbers respectively. N_{ij} is defined as the original frequencies. To evaluate interested cell counts N_{ij}, a statistic to rank cells according to their significance, is introduced as expected counts E_{ij}, which can be interpreted as a baseline or null hypothesis frequency. These two are used to estimate the hyperparameters of the prior distribution as well. The measure k is introduced as a grouping or stratification parameter to utilize within-group correlation measures. Cells with much less N_{ij} compared with the counts E_{ij} are also incorporated in the analysis. E_{ij} can be defined as:

$$E_{ij} = \sum_k \frac{N_{i.k}N_{.jk}}{N_{..k}} \tag{1}$$

Where,

$$N_{ij} = N_{ij}. = \sum_k N_{ijk};\ \ N_{i.k} = \sum_j N_{ijk};\ \ N_{.jk} = \sum_i N_{ijk}; N_{.k} = \sum_i \sum_j N_{ijk}$$

N_{ijk} denotes reported count for row i (vehicle characteristics) and column j (crash severity) of stratified combination k. E_{ij} is the expected counts assuming that the rows and columns variables are independent conditional on stratification variable.

Consider $\lambda_{ij} = \frac{\mu_{ij}}{E_{ij}}$ as the decision statistics to evaluate unusually large frequencies for each cell in the contingency table. The statistic λ is drawn from a mixture of two gamma densities with having (α_1, α_2) and (β_1, β_2) as the shape parameters and rate parameters and θ as proportion constant of these two densities. Prior probability density for λ can be deribed as:

$$\pi(\lambda; \alpha_1, \beta_1, \alpha_2, \beta_2,\ P) = \theta \times g(\lambda; \alpha_1, \beta_1) + (1-\theta) \times g(\lambda; \alpha_2, \beta_2)$$

where g is gamma distribution with mean of α/β and variance of α/β^2. The free parameters $\alpha_1, \beta_1, \alpha_2, \beta_2,$ *and* θ are needed to be estimated by maximum likelihood method in corresponding parameter space.

The posterior distribution of each λ is seen following a mixture of two gamma distributions with modified parameters. The marginal distribution of N is a combination of two negative binomial distributions, denoted as $Pr(N = n) = \theta f(n; \alpha_1, \beta_1, E) + (1-\theta) f(n; \alpha_2, \beta_2, E)$.

Where,

$$f(n; \alpha; \beta, E) = \left(1 + \frac{\beta}{E}\right)^{-(n+\alpha)} \frac{\Gamma(\alpha + n)n!}{\Gamma(\alpha)}$$

Defining Q_n as the posterior probability given $N = n$. Based on Bayes rule, $Q_n = \frac{\theta f(n; \alpha_1, \beta_1, E)}{Pr(N=n)}$.

Thus, to evaluate $EBlog2_{ij}$, the posterior expectation of $log_2(\lambda_{ij})$ as the Bayesian version of RR for ranking cells, can be derived as:

$$E\left[log_2(\lambda)|N = n\right] = \frac{\{Q_n\left[\psi(\alpha_1 + n) - log(\beta_1 + E)\right] + (1-Q_n)[\psi(\alpha_2 + n) - log(\beta_2 + E)]\}}{log(2)} \tag{2}$$

Where, ψ is digamma function, the derivative of $log\left[\Gamma(.)\right]$

For large values of E and N/E, $\psi(\mathrm{n})$ approaches $log(n)$, then $EBlog2_{ij}$ or $EB\left[log_2(\lambda)|N = n\right]$ approaches $log(\alpha + N) - \ log(\beta + N),\ or\ log(N/E)$ or $log(RR)$. However, when E and N/E are not large, $EBlog2_{ij}$ usually shrinks toward a lower value. To make $EBlog2_{ij}$ of the same scale as RR and obtain a value that is easily interpretable, DuMouchel (1999) computes $EBGM_{ij}$, the geometric mean of $EBlog2_{ij}$, which is given by the equation $EBGM_{ij} = \ 2^{EBlog2_{ij}}$. According to the EB method, choice of prior parameters depends on data characteristics. The values of $\alpha_1, \beta_1, \alpha_2, \beta_2,\ and\ \theta$ are obtained by maximizing the likelihood of the parameters, which is the likelihood is the marginal distribution of $\mathrm{N_{ij}}$.

4. Results

This study used an open source R package openEBGM, which considers the EB data mining approach developed by DuMouchel (1999). This method provides information about the significance of frequency of a given combination of effect and response in the contingency table. It works well even with very large but sparse tables. The final data set of this study has 24,285 entries of vehicle defect–associated vehicle-level data. Variable 1 consists of three vehicle-related properties: vehicle type, vehicle defect, and vehicle age. For example, motorcycle worn tires 11–15 years represents Variable 1. Variable 2 is the injury severity of the drivers. The number of unique combination group (Variable 1–Variable 2) is 3,657. Conventional statistical procedures are not suitable enough in identifying significant rules from this data set due to its form of a large contingency table.

The preliminary step is to populate actual counts of each vehicle characteristics and crash injury combination, expected counts (E) under the row/column independence assumption, relative reporting ratio (RR), and proportional reporting ratio (PRR). The original counts (N) and expected counts (E) are used to estimate the hyperparameters of the prior distribution. A large contingency table will have many cells that will result in computational complexities for the optimization routines needed for estimation. The common practice of hyperparameters estimation is to minimize the negative log-likelihood function. The optimized hyperparameters generated for this study are $(\alpha_1, \beta_1, \alpha_2, \beta_2,\ \theta) = (3.220 \text{ x } 10^{-07}, 0.952, 7.024, 3.120, 0.351)$. The optimization was performed by using expectation/conditional maximization (ECM) algorithm. A score with a value larger than one indicates the generated rule is found at an unusually high rate. Table 1 lists the top 20 rules with high PRRs.

The EBGM score represents the antilog of the mean of the ln-transformed posterior distribution. This value is a measure of central tendency of the posterior distribution. The EBGM scores indicate an adjusted estimate for the relative reporting ratio. For example, SUV Tire Failure 11-

Table 1. Top 10 combination groups with high Empirical Bayes Geometric Mean (EBGM) values.

Vehicle Characteristics	Crash Severity	*n*	E	PRR	EBGM
SUV tire failure 11–15 years	Minor	117	66.39	1.79	1.62[a]
Motorcycle worn tires 11–15 years	Nonincapacitating	10	0.74	13.62	1.57
Motorcycle defective brakes 7–10 years	Nonincapacitating	10	1.18	8.51	1.52
SUV tire failure 16–20 years	Nonincapacitating	18	6.51	2.79	1.52
SUV tire failure 7–10 years	Nonincapacitating	29	14.45	2.03	1.48
Car worn tires 7–10 years	Fatal	12	3.19	4.05	1.46
SUV engine failure 11–15 years	Minor	28	14.51	1.94	1.44
SUV tire failure 11–15 years	Nonincapacitating	34	19.63	1.75	1.40
Pickup worn tires 11–15 years	Fatal	9	2.00	4.76	1.39
SUV tire failure 7–10 years	Minor	74	48.87	1.52	1.39
Car worn tires 11–15 years	Fatal	10	2.85	3.72	1.38
Car worn tires 16–20 years	Minor	53	34.03	1.57	1.38
Motorcycle tire failure 7–10 years	Nonincapacitating	7	0.64	10.97	1.38
Car defective suspension 11–15 years	Nonincapacitating	18	8.88	2.04	1.36
Car tire failure 11–15 years	Minor	137	96.24	1.44	1.36
Motorcycle worn tires 7–10 years	Fatal	6	0.16	40.30	1.36
SUV defective steering 11–15 years	Minor	19	9.67	1.97	1.36
SUV worn tires 11–15 years	Minor	76	51.87	1.47	1.35
Car defective rear light 16–20 years	Minor	14	6.67	2.10	1.32
Motorcycle worn tires < 3 years	Nonincapacitating	6	0.54	11.11	1.32

Note. PRR = Proportional Reporting Ratio; SUV = Sport Utility Vehicle
[a]Sorted by EBGM scores.

Table 2. Fatality rules with high Empirical Bayes Geometric Mean (EBGM) scores and quantiles.

Vehicle Characteristics	Crash Severity	EBGM	Quantile 5%	Quantile 95%
Car worn tires 7–10 years	Fatal	1.46	1.05	2.00
Pickup worn tires 11–15 years	Fatal	1.39	0.97	1.93
Car worn tires 11–15 years	Fatal	1.38	0.97	1.90
Motorcycle worn tires 7–10 years	Fatal	1.36	0.93	1.93
Pickup worn tires 16–20 years	Fatal	1.29	0.88	1.84
Pickup worn tires 7–10 years	Fatal	1.29	0.89	1.82
Car worn tires 16–20 years	Fatal	1.28	0.87	1.82
Motorcycle worn tires 3–6 years	Fatal	1.23	0.82	1.78
Pickup worn tires 21–25 years	Fatal	1.2	0.8	1.74
Pickup worn tires 3–6 years	Fatal	1.17	0.78	1.69

15 years → Minor rule has an EB score of 1.62. This score indicates that this rule occurred in the data 1.62 times more frequently than expected under the assumption of no association between defect associated characteristics and crash severity. For a two-sided 90% credibility interval for λ_{ij} (given N_{ij}), the 5% and 95% quantiles of the posterior distributions were used. Table 2 lists the top ten rules (by keeping fatal as the fixed response) with high EBGM and quantiles. Tire defect is a major factor for fatal crashes involved with vehicular defects. Another interesting finding is that the vehicle older than 7 years in the generated rules. Motorcycles older than 3 years are also found in the rules.

To take a closer look at the incapacitating injury type, another table is prepared using the top 10 rules (see Table 3). The rule with the highest EBGM score is SUV Worn Tires 11–15 years → Incapacitating Injury. The EBGM score of this rule indicates that it occurred in the data 1.22 times more frequently than expected under the assumption of no association

Table 3. Incapacitating injury rules with high Empirical Bayes Geometric Mean (EBGM) scores and quantiles.

Vehicle Characteristics	Crash Severity	EBGM	Quantile 5%	Quantile 95%
SUV worn tires 11–15 years	Incapacitating	1.22	0.83	1.73
Motorcycle defective brakes 7–10 years	Incapacitating	1.15	0.76	1.69
Pickup worn tires 21–25 years	Incapacitating	1.13	0.75	1.66
SUV tire failure 16–20 years	Incapacitating	1.11	0.73	1.62
Car defective rear light 21–25 years	Incapacitating	1.09	0.71	1.62
Car defective steering 21–25 years	Incapacitating	1.09	0.71	1.61
Motorcycle defective brakes 3–6 years	Incapacitating	1.09	0.71	1.62
Motorcycle defective brakes NA	Incapacitating	1.09	0.71	1.62
Motorcycle tire failure 7–10 years	Incapacitating	1.09	0.71	1.62
Motorcycle worn tires 3–6 years	Incapacitating	1.09	0.71	1.62

between defect-associated characteristics and crash severity. Fifty percent of the rules involve motorcycles (aging from 3–10 years). For motorcycle crashes, defective brakes are over-represented in the generated rules, whereas tire-related issues are predominant in SUV and pickup crashes. Vehicles older than 20 years are more likely to have defective rear light and defective steering.

The application of EBGM is effective in comparison to other rules mining technique (e.g., association rules mining). In association rules mining, there is a need to determine the threshold of support and confidence values to acheive more intuitive results. The EBGM technique does not require any threshold determination, and it provides more intuitive scores with confidence intervals.

5. Conclusion

Approximately 2% of traffic crashes occur due to defective vehicle–related events; and approximately 3% of traffic fatalities in Louisiana are associated with vehicle defect–related issues. Additionally, it is commonly hypothesized that vehicle safety diminishes over time. Despite this, there has not been much previous research conducted to evaluate the impact of vehicle defects on traffic crash occurrences. This study used 7 years of Louisiana vehicle defect–related crash data to determine the association between crash severity and vehicle defect–associated factors. The study compared the proportion of vehicular defects in the vehicle fleet (specific to vehicle type and age) with the proportion found in the crashed vehicle population in Louisiana. The EBGM scoring method identified some key groups that require further attention.

This study shows the uniqueness of EBGM in determining the significant associations between the factors. The findings from this study are the following:

- The top vehicular defects resulted in crashes in Louisiana are defective brakes, tire failure, worn tires, defective suspension, and engine failure.

These factors represent approximately 81% of all vehicular defect–related crashes. Tires and brakes are the most significant defects in Louisiana database, which is in line with the proportion distribution of FARS data set.

- Motorcycle crashes are over-represented in KABC crash proportions. Tire, defective steering, and engine failure are also over-represented in KABC crash proportions.
- Vehicle age is associated with the severe injury types.
- The EBGM scoring clearly shows that vehicle older than 7 years is strongly associated with vehicle defect–related crashes.
- The EBGM scoring methods also show that tire, defective brake, engine failure, and defective steering are the key defects associated with different vehicle types and ages.
- The top ten rules for fatal and incapacitating injury crashes identify several risk groups: worn tires in different vehicle types, vehicles older than 7 years, motorcycles older than 3 years, a motorcycle with defective brakes, and cars older than 20 years with either defective rear light or defective steering.

The current study is not without limitations. First, crash data of Louisiana do not provide detailed precrash information. The database was prepared based on the police-reported vehicle condition variable, and researchers captured the targeted information based on this specific variable. A further investigation of the crash reports would be necessary to identify more robust data on vehicle defects. Second, the important technical issue in the rule-based analysis is the control of redundancy of the rules. Removal of redundant rules requires manual investigation that would be difficult for a large set of generated rules. Third, this procedure provides estimated risk measures associated with the rules. The expected crashes and similar information can not directly be extracted from this process.

In summary, it can be said that the rule-based approach will be useful in many traffic safety studies. The generated rules that define a subgroup of associated factors can lead to developing different strategies. Besides the safety improvement, it can also help with the prioritization of the safety targets.

Acknowledgements

The authors appreciate the help of Dr. Xiaoduan Sun for the data and suggestions. The authors would like to thank two anonymous reviewers for their valuable comments and suggestions.

Funding

This project has not been funded by any research grant.

ORCID

Subasish Das http://orcid.org/0000-0002-1671-2753

References

American Association of State Highway and Transportation Officials (AASHTO). (2010). *Highway safety manual* (1st ed.)., Washington, DC: American Association of State Highway and Transportation Officials.

Al-Ghaweel, I., Mursi, S. A., Jack, J. P., & Joel, I. (2009). Factors affecting road traffic accidents in Benghazi. *Journal of Family & Community Medicine*, *16*(1), 7–9.

Akloweg, Y., Hayshi, Y., & Kato, H. (2011). The effect of used cars on African road traffic accidents: A case study of Addis Ababa, Ethiopia. *International Journal of Urban Sciences*, *15*(1), 61–69. doi:10.1080/12265934.2011.580153

Baireddy, R., Zhou, H., & Jalayer, M. (2018). Multiple correspondence analysis of pedestrian crashes in rural Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(38), 116–127. doi:10.1177/0361198118777088

Barry, S., Ginpil, S., & O'Neill, T. J. (1999). The effectiveness of air bags. *Accident Analysis and Prevention*, *31*(6), 781–787.

Conroy, C., Tominaga, G. T., Erwin, S., Pacyna, S., Velky, T., Kennedy, F., … Coimbra, R. (2008). The influence of vehicle damage on injury severity of drivers in head-on motor vehicle crashes. *Accident Analysis & Prevention*, *40*(4), 1589–1594. doi:10.1016/j.aap.2008.04.006

Cuerden, R., Edwards, M., & Pittman, M. (2011). Effect of vehicle defects in road accidents. Report: PPR565, Transport Research Laboratory (TRL), UK.

Das, S., & Sun, X. (2015). Factor association with multiple correspondence analysis in vehicle–pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, *2519*(1), 95–103. doi:10.3141/2519-11

Das, S., Sun, X., Wang, F., & Leboeuf, C. (2015). Estimating likelihood of future crashes for crash-prone drivers. *Journal of Traffic and Transportation Engineering (English Edition)*, *2*(3), 145–157. doi:10.1016/j.jtte.2015.03.003

Das, S., & Sun, X. (2016). Association knowledge for fatal run-off-road crashes by multiple correspondence analysis. *IATSS Research*, *39*(2), 146–155. doi:10.1016/j.iatssr.2015.07.001

Das, S., Brimley, B. K., Lindheimer, T., & Pant, A. (2017). *Safety impacts of reduced visibility in inclement weather*. Final Report ATLAS-2017-19, University of Michigan, MI.

Das, S., Minjares-Kyle, L., Avelar, R. E., Dixon, K. K., & Bommanayakanahalli, B. (2017). *Improper passing related crashes on rural roadways: Using association rules negative binomial miner*. Presented at the Transportation Research Board 96th Annual Meeting: Transportation Research Board.

Das, S., Dutta, A., Jalayer, M., Bibeka, A., & Wu, L. (2018). Factors influencing the patterns of wrong way driving crashes on freeway exit ramps and median crossovers: Exploration using 'Eclat' association rules to promote safety. *International Journal of Transportation Science and Technology*, *7*(2), 114–123. doi:10.1016/j.ijtst.2018.02.001

Das, S., Dutta, A., Avelar, R., Dixon, K., Sun, X., & Jalayer, M. (2018). Supervised association rules mining on pedestrian crashes in urban areas: Identifying patterns for appropriate countermeasures. *International Journal of Urban Sciences*, *23*(1), 1–19. doi: 10.1080/12265934.2018.1431146

Das, S., Mudgal, A., Dutta, A., & Geedipally, S. (2018). Vehicle consumer complaint reports involving severe incidents: Mining large contingency tables. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(32), 72. doi:10.1177/0361198118788464

Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention*, *111*, 43–55. doi:10.1016/j.aap.2017.11.016

Das, S., Kong, X., & Tsapakis, I. (2019). Hit and run crash analysis using association rules mining. *Journal of Transportation Safety & Security*. doi:10.1080/19439962.2019.1611682

Das, S., Dutta, A., & Sun, X. (2019). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety & Security*. doi:10.1080/19439962.2019.1572681

DEKRA Automobil GmbH. Road Safety Report (2015). A future based on experience. DEKRA, Stuttgart, Germany. http://www.dekra-vision-zero.com/downloads/road_safety_report_2015.pdf.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician*, *53*(3), 177–190. doi:10.2307/2686093

Factor, R., Yair, G., & Mahalel, D. (2010). Who by accident? The social morphology of car accidents. *Risk Analysis*, *30*(9), 1411. doi:10.1111/j.1539-6924.2010.01423.x

Fatality Analysis and Reporting System (FARS). (2018). National Highway Traffic Safety Administration. ftp://ftp.nhtsa.dot.gov/fars/.

Green, E. R., Agent, K. R., Pigman, J. G., & Ross, P. A. (2018). Analysis of traffic crash data in Kentucky (2012–2016). Lexington, KY: Kentucky Transportation Center, University of Kentucky. uknowledge.uky.edu/ktc_researchreports/1585.

Hoque, M. S., & Hasan, M. R. (2006). *Vehicle factors in road accidents: The context of developing countries*. Paper presented at International Conference on Road Safety in Developing Countries, Dhaka, Bangladesh.

Ihrie, J., Ahmed, I., & Poncet, A. (2017). EBGM scores for mining large contingency tables. Version 0.1.0

Islam, M. B., & Kanitpong, K. (2008). Identification of factors in road accidents through in-depth accident analysis. *IATSS Research*, *32*(2), 58–67. doi:10.1016/S0386-1112(14)60209-0

Jalayer, M., Zhou, H., & Das, S. (2018). *Exploratory analysis of run-off-road crash patterns. Data analytics for smart cities*. Washington, DC: CRC Press.

Jalayer, M., Pour-Rouholamin, M., & Zhou, H. (2018). Wrong-way driving crashes: A multiple correspondence approach to identify contributing factors. *Traffic Injury Prevention*, *19*(1), 35–41. doi:10.1080/15389588.2017.1347260

Kane, S., Liberman, E., DiViesti, T., & Click, F. (2018) Toyota sudden unintended acceleration. www.rightinginjustice.com/media/2010/02/SRS-Report-on-Toyota-Sudden-Unintended-Acceleration.pdf.

Kinsey, G. (1976). Contribution of unroadworthy vehicles to accidents. 2nd Conference, National Institute of Transport and Road Research, Pietersburg, South Africa.

Lloyd, L., Wallbank, C., Broughton, J., & Cuerden, R. (2017). Estimating the potential impact of vehicle secondary safety regulations and consumer testing programs on road casualties in emerging markets. *Journal of Transportation Safety & Security*, *9*, 147–149. doi:10.1080/19439962.2016.1228091

Malaysian Institute of Road Safety Research. (2018). Automotive consumerism: A study of car user's practices & behaviour in Klang Valley, Malaysia. *Selangor Darul Ehsan, Malaysia*.

Moodley, S., & Allopi, D. (2008). *An analytical study of vehicle defects and their contribution to road accidents*. In Proceedings of the 27th Southern Africa Transport Conference, Pretoria, South Africa.

National Highway Traffic Safety Administration. (2018). Office of defects investigation. www-odi.nhtsa.dot.gov/downloads/.

National Motor Vehicle Crash Causation Survey. (2018). DOT HS 811 059. crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811059.

R Core Team (2013). R: A language and environment for statistical computing, version 3.4.1. (2017). Vienna, Austria: The R Foundation for Statistical Computing. http://www.R-project.org/

Schoor, O. V., & Niekerk, J. L. (2001). Mechanical failures as a contributing cause to motor vehicle accidents-South Africa. *Accident Analysis & Prevention*, *33*(6), 713–721.

Solah, M. S., Hamzah, A., Ariffin, A. H., Paiman, N. F., Hamid, I. A., … Osman, M. R. (2017). Private vehicle roadworthiness in Malaysia from the vehicle inspection perspective. *Journal of the Society of Automotive Engineers Malaysia*, *1*(3), 262–271.

Treat, J., Tumbas, N., McDonald, S., Shinar, D., Hume, R., … Castellan, N. (1979). Tri-level study of the causes of traffic accidents. DOT Hs-034-3-535.

U.S. Department of Transportation (USDOT). (2015). Critical reasons for crashes investigated in the national motor vehicle crash causation survey: A brief statistical summary. Report No. DOT HS 812 115.

U.S. Department of Transportation (USDOT). (1975). National traffic safety newsletter, NHTSA Washington D.C.

Vaughan, R. (1992). *Vehicle ageing and safety*. Paper presented at Wheels '92 Conference and Workshop, Sydney, Australia.

Weng, J., & Li, G. (2019). Exploring shipping accident contributory factors using association rules. *Journal of Transportation Safety & Security*, *11*(1), 36–57. doi:10.1080/19439962.2017.1341440

Wolf, R. (1968). Truck accidents and traffic safety – An overview. SAE meeting, Detroit.