Title: Understanding Factors Associated with Misclassification of Fatigue-related Accidents in Police Record

Authors: Yanyan Li[a,*], Toshiyuki Yamamoto[b], Guangnan Zhang[c,d]

[a] Department of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

[b] Institute of Materials and Systems for Sustainability, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan

[c] Institute of Guangdong, Hong Kong and Macao Development Studies, Sun Yat-Sen University, Xingang Xi Road, Guangzhou, China

[d] Center for Studies of Hong Kong, Macao and Pearl River Delta, Sun Yat-Sen University, Xingang Xi Road, Guangzhou, China

Abstract:

Introduction: Fatigue is one of the riskiest causes of traffic accidents threatening road safety. Due to lack of proper criteria, the identification of fatigue-related accident by police officers largely depends on inferential evidence and their own experience. As a result, many fatigue-related accidents are misclassified and the harmfulness of fatigue on road safety is misestimated. *Method:* In this paper, a joint model framework is introduced to analyze factors contributing to misclassification of a fatigue-related accident in police reports. Association rule data mining technique is employed to

[*] Corresponding author at: Department of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan.

E-mail address: lyy901207@gmail.com (Yanyan Li), yamamoto@civil.nagoya-u.ac.jp (Toshiyuki Yamamoto), sysuzgn@gmail.com (Guangnan Zhang).

identify the potential interactions of factors and logistic regression models are applied to analyze factors that hinder police officers' identification of fatigue-related accidents. Using the fatigue-related crash records from Guangdong Province during 2005-2014, factors contributing to the false positive and false negative detection of the fatigue-related accident have been identified and compared. *Results:* Some variables and interactions were identified to have significant impacts on fatigue-related accident detection. *Conclusions:* Based on the results, it can be inferred that the stereotype of certain groups of drivers, crash types, and roadway conditions affect police officers' judgment on fatigue-related accidents. *Practical applications:* This finding can provide useful information for training police officers and build better criteria for fatigue identification.

Keywords: fatigue; police report; misclassification; association rule; logistic regression model

1. Introduction

Fatigued driving is a serious problem threatening road safety all around the world. Police records from different countries indicate a range from 1% to 4% incidence of fatigue/sleep-related crashes of all registered crashes (Traffic Management Bureau, Ministry of Public Security, PRC, 2008; Radun & Radun, 2009). However, several questionnaire-based surveys suggest the role of fatigue in a traffic accident is misestimated. National Sleep Foundation (2008) reported that approximately 32% of respondents in Sleep in America Poll had driven while fatigued at least once a month. In China, a survey conducted in Guangdong province in 2007 also showed that 9.3% vehicle drivers had the fatigue driving experience in the past 30 days (Yan et al. 2010). The difference between police reports and surveys implies that police reports could have significantly misestimated the harmfulness of fatigue in road safety. One of the possible reasons is police officers are not so alert to the presence of fatigue and have difficulties in identifying fatigue-related accidents (Robertson et al. 2009).

Among all causes of traffic accidents, fatigue-related accidents are easily neglected or misclassified due to the difficulty in observing and identifying driver fatigue (Radun et al. 2013; Filtness et al. 2015). Unlike drunk driving crashes, no blood or breath test can be applied to quantify driver's fatigue level at crash scene (Pack et al. 1995; DaCoTA, 2012). As a result, there is currently no standard methodology for identifying fatigue as the cause of the crash (Crummy et al. 2008;

Filtness et al. 2015) and defining fatigue-related accident largely relies on inferential evidence or experience. For example, police officers may consider a crash to be fatigue-related, when the following conditions appear (Horne et al. 1995; NCSDR/NHTSA, 1998; Horne et al. 1999): occur during late night or mid-afternoon; single vehicle run off the roadway; occur on a high-speed road; absence of skid marks or braking. Some fatigue-related accidents were determined even by eliminating other causes of accidents (e.g. speeding, drunk driving, etc.).

To assist identification of fatigue in an accident, proxy measurements are developed aiming to improve reporting accuracy of fatigue-related accidents (Filtness et al. 2015). In Australia, the national Australian Transport Safety Bureau (ATSB, 2006) has developed the proxy definition for fatigue/sleep-related accident, and five jurisdictions in Australia have already incorporated proxy definition into their reporting process. In Queensland, for example, fatigue can be considered as a contributor to a crash when it fitted the proxy definition: single-vehicle crashes in more than 100 km/h speed zones which occur during midnight and in the afternoon, or where a vehicle runs out of roadway and the driver does not try to avoid the accident (Armstrong et al., 2013; Filtness et al. 2015). Although these proxy definitions are based on experience or scientific research, they are criticized for too specific (Crummy et al. 2008; Armstrong et al. 2013) and may provide misleading instructions for police officers. A questionnaire-based study conducted in Australia by Crummy et al. (2008) found that only a small proportion of participants that actually had a fatigue/sleep-related crash were correctly identified by ATSB proxy definitions (ATSB, 2006).

Reliable and accurate records are essential for assessing the scope of fatigue-related accident problems, monitoring and evaluating the effectiveness of intervention measures. A survey in Ontario showed that 56.6% of traffic police felt that they did not receive enough training to identify drivers who were fatigued or drowsy, or determined the role of fatigue in a crash (Robertson et al. 2009). Although several risk factors identified by prior research and public belief are believed to contribute to fatigue-related accidents, few works have been done to prove that whether these factors are useful for police officers to identify fatigue-related accidents. That is, some of the factors believed as associated with fatigue-related accidents are not helpful in judging whether an accident is fatigue-related, and may even lead to incorrect classification of the cause of accidents. Therefore, in this study, we proposed an analysis framework based on existing crash data to identify factors that easily make fatigue-related accidents misclassified by police officers, examine the interactive effects of those factors, and provide better inference for determining fatigue-related accident by removing some misleading terms, which help to improve enforcement strategies.

The paper is organized as follows: In Section 2, the detail of analysis strategies will be discussed; Section 3 describes the dataset and variables used in this study; the results will be presented in Section 4; Section 5 will further discuss the results; in the last section of the paper, a conclusion will be given.

2. Methodology

2.1 Objectives and research strategy

This study aims at investigating potential factors that hinder police officers' identification of fatigue-related accidents. However, some factors have not only individual effects but also combinatorial effects on the determining of fatigue-related accidents. Classic logistic regression model lacks appropriate criteria to incorporate interactions between independent variables when there are a large number of variables to be considered. Instead, ignoring interactions may cause biased estimation. Therefore, our strategies for this analysis are: (1) Association rule data mining technique is applied to identify important interactions between factors, which helps overcome the disadvantage of classic logistic regression model in selecting appropriate interactions; (2) Incorporating the interactions identified by association rules, binary logistic regression models are applied to find out factors that hinder police officers from correctly identifying fatigue-related accidents.

2.2 Association rule analysis

Regression models in road safety research focus on establishing and analyzing relationships between "dependent" and "independent". It is also important to take the correlation between "independent" variables into consideration since it may hamper the statistical analysis (Pande & Abdel-Aty, 2009). With the increasing number of independent variables, however, the number of interactions will grow at an accelerated rate. Thus, the methodology for identifying potential interaction among a large number of crash-related factors is needed. Association rule data mining technique can potentially identify relationships that are not well known from current research works and have been used in traffic safety research (Pande & Abdel-Aty, 2009; Montella et al. 2011; Das & Sun, 2014; Weng et al. 2016). Some studies have combined association rule data mining technique with logistic regression model for other purposes (Kamei et al. 2008; Shaharanee et al. 2009), but few of them use association rule analysis as a tool for selecting potential interactions among variables. Changpetch and Lin (2013) proposed a model selection method procedure for logistic (Changpetch & Lin, 2013a) and multinomial logit model (Changpetch & Lin, 2013b), which help to improve the classic model by considering potential interactions.

In this study, association rule analysis is performed using a priori algorithm according to the methodology introduced by Agrawal et al. (1993). A rule is defined as an implication of the form "A→B", where A is the antecedent (left-hand-side, LHS) and B is the consequent (right-hand-side, RHS). It is important to note that the rule should not be interpreted as a direct causation, but as associations between variables (Pande & Abdel-Aty, 2009; Montella et al. 2012). Three measures are commonly used in filtering rules: Support, Confidence and Lift. Support measures the frequency of LHS and RHS appearing in the dataset and is calculated as follows:

$$\text{Support}(A \rightarrow B) = P(AB) \quad (1)$$

where $P(AB)$ represents the probability of case containing A and B at the same time. Confidence determines how frequently RHS appears given that LHS occurs:

$$\text{Confidence}(A \rightarrow B) = P(B \mid A) \quad (2)$$

$$= P(AB)/P(A)$$

where $P(A)$ is the probability of case containing A. Lift is a measure of the statistical dependence of the rule. A lift value which is smaller than one indicates negative independence between LHS and RHS, a value equal to one indicates independence, and a value which is greater than 1 indicates positive interdependence (Montella et al. 2012). Higher lift value indicates stronger associations. Lift is defined as follows:

$$\text{Lift}(A \rightarrow B) = P(B \mid A)/P(B) \quad (3)$$

$$= P(AB)/P(A)\,P(B)$$

To make sure that the identified rules are reasonable and accurate, the minimum threshold values for these three indexes need to be specified. Since there are no clear criteria for choosing threshold values, different studies employed different threshold support and confidence values (Pande & Abdel-Aty, 2009; Montella, 2011; Montella et al., 2012; de Oña et al. 2013) based on the nature of the data (balanced or not) and sample size (small or large databases). For example, Pande and Abdel-Aty (2009) set 0.009 and 0.1 for them respectively. Thus, in this study the minimum threshold values for Support, Confidence and Lift are set as follows: Support $\geq$ 0.01, Confidence $\geq$ 0.1, and Lift $\geq$1.2. It also needs to be emphasized that only rules with two items in the

LHS are selected for ease of interpretation. We firstly generate rules with non-fatigue to fatigue crash or fatigue to non-fatigue crash in RHS from all the generated rules. Then, all the selected rules are descending ordered by confidence, and the top ten are changed into interactions and incorporated into logistic regression models as inputs.

2.3 Binary logistic regression model

Two assessment results will be recorded for normal procedure crash records in the database. The one recorded by the police officer at the crash scene is denoted as on-site assessment. Normally, *on-site assessment* was determined by a quick check at the crash scene, surrounding environment and simply ask those who were present for facts of the crash. The other one recorded in the final report is denoted as *final assessment*, which is the assessment result after an in-depth investigation. After placing a case on file for investigation, more detail information of driving condition is also collected for investigation. Information such as skid marks and scrub marks, the condition of vehicle mechanical (e.g. brakes, steering, tires, and lights). Besides road surveillance video and driving records, police officers will obtain and examine marks from the crash scene, statements from drivers and witnesses, collect off-scene information, and do vehicle mechanical inspection. Based on that information, an accident reconstruction can be performed to examine the real cause of the accident. Therefore, based on more detailed and reliable evidence, the final assessment is believed to be the accurate assessment result for the crash. It needs to be emphasized that on-site assessment and final assessment can be different, and this inconsistency of crash cause enables us to analyze the factors hindering making correct judgments. And the simple procedure crash records with on-site assessment only are not included in this study due to lack of information for judging whether they have been correctly classified or not. Thus, based on the combinations of on-site assessment and final assessment, three types of records are presented in Table 1:

- **Non-fatigue** to **Fatigue** (**N-F crash**): crashes with non-fatigue as *on-site assessment* and fatigue as *final assessment*;
- **Fatigue** to **Non-fatigue** (**F-N crash**): crashes with fatigue as *on-site assessment* and non-fatigue as *final assessment*;
- **Fatigue** to **Fatigue** (**F-F crash**): the *on-site assessment* and *final assessment* are both fatigue.

Table 1: Summary of Types of Records by Injury Severity Level

	Fatal and severe injury	**Minor injury**	**Total**
N - F crash	55	66	121
F - N crash	156	173	329
F - F crash	211	350	561
Total	422	589	

Similar to the terms widely used in medical screening (Stegeman et al. 2013), fatigue detection at the crash scene can be regarded as a test and the final result after full investigation as the real cause. Then, we can define the false negative fatigue-related accident as a fatigue-related accident that was misclassified into other causes at the crash scene. A false positive fatigue-related accident is defined as a crash believed to be fatigue-related but is actually not. Binary logistic regression models are employed to identify significant factors affecting false positive and false negative fatigue-related accident detection. The first model is established to identify factors related to false negative fatigue-related accident detection. The binary outcomes represented by a dummy variable (1 indicates false negative fatigue-related accident detection, 0 is correct fatigue-related accident detection), is used as dependent variables. The second model is built to identify factors affecting false positive fatigue-related accident detection. Similarly, the dependent variable is also dummy variable (1 indicates false positive fatigue-related accident detection, 0 is correct fatigue-related accident detection).

Furthermore, police officers may treat accidents differently under different injury severity level (e.g. if police suspect the accident may be related to fatigue driving, they tend to give an oral examination to determine whether the accident is fatigue-related level in less severe accidents (if the people involved are sober)). Thus, we separated our dataset into two groups based on their recorded injury severity level of the driver who is responsible for the accident: fatal and severe injury, and minor injury. We do not include property only accidents because part of them was applied to the simple procedure without further investigation and analysis based on selected sample will be biased.

3. Data

The data employed in this study were obtained from the Road Traffic Accident Database of China's Public Security Department (Zhang et al. 2016). All police-recorded fatigue crashes, relevant crash records occurred in Guangdong Province during 2005-2014, were filtered from the Traffic Accident Database. Only records in which the cause of an accident was convicted as fatigue-related and the involving driver who was fully or mainly responsible for the accident, were used in this study. According to definition by the database, an accident was defined as fatigue-related accident when fulfilled one of the following conditions: (a) driving cars more than eight hours a day, (b) engaging in other work excessive physical exertion, (c) lack of sleep which results in sleepy or lower reaction rate, so that the driver is having difficulty in assessing traffic conditions immediately and reacting accurately. In this study, 1101 general procedure crash records were extracted from the database. Among them, 561 are F-F crashes, 121 are N-F crashes, and 329 are F-N crashes. Although fatigue-related accidents account for a small percentage of all accidents in our dataset, the percentage of incorrect detection of fatigue-related accidents is really high.

To focus on the meaningful analysis, several variables will be considered in this study guided by prior research. These variables were selected into the final models: crash characteristics (crash type), driver characteristics (driver's gender, age, and occupation), vehicle characteristics (vehicle type and insurance condition), roadway characteristics (road type, lane type, and road segment) and environmental characteristics (lighting condition and weather). In order to consider non-linear effects in logistic regression models, driver's age was categorized: age (≤30, 31-40, 41-50, and ⩾51). The description of these variables is presented in Table 2.

Table 2: Descriptive Statistics of Variables

	Fatal and severe injury		Minor injury	
Variable	**Count**	**Percentage**	**Count**	**Percentage**
Crash characteristics				
Motor vehicle crash	233	55.21%	402	68.25%
Rollover	70	16.59%	75	12.73%
Hit fixed object	114	27.01%	88	14.94%
Driver characteristics				
Male	416	98.58%	553	93.89%
31-40 years old	156	36.97%	218	37.01%
41-50 years old	94	22.27%	108	18.34%
⩾51 years old	21	4.98%	47	7.98%
Clerk	24	5.69%	44	7.47%
Worker	93	22.04%	99	16.81%
Farmer	71	16.82%	131	22.24%
Self-employed	57	13.51%	64	10.87%
Migrant worker	52	12.32%	37	6.28%
Unemployed	3	0.71%	10	1.70%
Vehicle characteristics				
Heavy/medium truck	111	26.30%	117	19.86%
Light truck	32	7.58%	33	5.60%
Large/medium bus	15	3.55%	13	2.21%
Passenger car	89	21.09%	104	17.66%
Have insurance	346	81.99%	495	84.04%
Roadway characteristics				
Expressway	99	23.46%	110	18.68%
Urban expressway	15	3.55%	18	3.06%
Urban road	51	12.09%	128	21.73%
Motor vehicle lane	340	80.57%	414	70.29%
Non-motorized vehicle lane	14	3.32%	35	5.94%
Mix lane	50	11.85%	122	20.71%
Intersection	46	10.90%	78	13.24%
Special road segment	35	8.29%	31	5.26%
Environmental characteristics				
Dark with street light	97	22.99%	125	21.22%
Dark without street light	165	39.10%	169	28.69%
Cloudy	58	13.74%	98	16.64%
Total number of crash	422		589	

4. Results

R software 3.3.1 was applied to generate association rules and estimate the models.

4.1 Association rule analysis

In this study, 'arules' package in R software is employed for computation of association rules (Hahsler et al. 2007). To find out the potential associations or patterns among the items, association rule analysis is conducted using N-F crash and F-N crash sample sets separately for two injury severity levels. The results are presented in Table 3.

Then, these rules are converted into dummy variables. Suppose the rule being selected is "if $X_i = x_i$ and $X_j = x_j$ then $Y = y$", where x_i is the level of variable X_i, x_j is the level of variable X_j, and y is the level of response Y (Changpetch & Lin, 2013). The interaction between X_i and X_j is denoted as 1 if $X_i = x_i$ and $X_j = x_j$, and as 0 otherwise. For example, Rule 1 for N-F fatal and severe injured crash in Table 3 which is "Given a fatal and severe injured crash is N-F crash, the crash occurred in cloudy weather condition and the mainly responsible driver drove large/medium cargo vehicle" is converted into a dummy variables D1 (D1=1 if weather=cloudy and large/medium cargo vehicle; D1=0, otherwise).

Table 3: Rules for N-F Accident and F-N Accident by Injury Severity Level

	Fatal and Severe Injury	Support	Confidence	Lift	Minor Injury Crash	Support	Confidence	Lift
				N - F crash				
1	"cloudy" & "large/medium cargo vehicle"	0.01	0.40	3.07	"light cargo vehicle" & "motorized vehicle lane"	0.01	0.29	2.60
2	"age 31-40" & "cloudy"	0.02	0.32	2.44	"age 41-50" & "expressway"	0.01	0.29	2.55
3	"motor vehicle crash" & "urban road"	0.01	0.32	2.42	"male driver" & "light cargo vehicle"	0.01	0.24	2.16
4	"night without street light" & "cloudy"	0.02	0.30	2.27	"motor vehicle crash" & "expressway"	0.03	0.22	1.97
5	"self-employed" & "night without street light"	0.02	0.27	2.07	"vehicle with insurance" & "light cargo vehicle"	0.01	0.22	1.95
6	"vehicle with insurance" & "cloudy"	0.03	0.25	1.92	"age 41-50" & "might without street light"	0.01	0.22	1.95
7	"motor vehicle crash" & "night without street light"	0.02	0.24	1.85	"expressway" & "motorized vehicle lane"	0.04	0.21	1.89
8	"age 41-50" & "passenger car"	0.01	0.23	1.77	"night without street light" & "cloudy"	0.01	0.20	1.78
9	"age 41-50" & "night without street light"	0.01	0.23	1.74	"male driver" & "expressway"	0.04	0.20	1.78
10	"age 31-40" & "night without street light"	0.04	0.22	1.69	"larger/medium cargo vehicle" & "expressway"	0.02	0.19	1.69
				F - N crash				
1	"hit fixed object" & "non-motorized vehicle lane"	0.01	0.86	2.32	"urban road" & "night without street light"	0.01	0.57	1.95
2	"urban road" & "cloudy"	0.01	0.83	2.25	"night with street light" & "rollover"	0.01	0.55	1.86
3	"age 31-40" & "non-motorized vehicle lane"	0.01	0.75	2.03	"urban road" & "intersection"	0.02	0.50	1.70
4	"urban road" & "intersection"	0.01	0.71	1.93	"migrant worker" & "motorized vehicle lane"	0.02	0.48	1.64
5	"intersection" & "night without street light"	0.02	0.70	1.89	"migrant worker" & "motor vehicle crash"	0.02	0.48	1.63
6	"male driver" & "non-motorized vehicle lane"	0.02	0.69	1.87	"intersection" & "night with street light"	0.02	0.48	1.62
7	"passenger car" & "night without street light"	0.02	0.64	1.74	"mixed lane" & "hit fixed object"	0.02	0.47	1.61
8	"urban road" & "night without street light"	0.02	0.64	1.72	"migrant worker" & "urban road"	0.01	0.47	1.59
9	"motor vehicle crash" & "urban expressway"	0.01	0.63	1.69	"clerk" & "night with street light"	0.01	0.46	1.57
10	"passenger car" & "intersection"	0.01	0.63	1.69	"worker" & "large/medium cargo vehicle"	0.01	0.46	1.57

4.2 Binary logistic model

The independent variables are selected based on their significance and model fitness. To avoid neglecting variable problem, a conservative selection strategy was employed in the current study. First, all the variables were tested in the basic models. If no variable is significant in a particular group, log-likelihood ratio test would be conducted to compare models with or without those variables. Information criteria were also compared for the same purpose. If log-likelihood ratio test cannot reject the null hypothesis and information criteria also showed better fit in the model without insignificant variables, those variables would be removed from the basic model. At the final results, variables with 90% or higher levels of significance are kept in final results for examining more possible impact factors given relatively smaller sample size. The results of logistic regression models for factors associating with false negative and false positive fatigue-related accident detection on two injury levels are presented in Table 4 and Table 5. For fatal and severe accidents, seven variables (include one interaction) are found to be significant for false negative fatigue-related accident detection and eight variables (include one interaction) are significant impact factors for false positive fatigue-related accident detection at the 90% level. Among minor injury accident, nine significant variables were identified for false negative fatigue-related accident detection and eight variables (include two interactions) are found to be significant for false positive fatigue-related accident detection. More detailed discussion of results will be presented in next section.

Table 4: Factors Associating with False Negative Fatigue-related Accident Detection

	Fatal and severe injury			Minor injury		
	OR	95% CI		OR	95% CI	
Crash Characteristics						
Motor vehicle crash	0.334	0.075	1.485	0.422	0.104	1.709
Rollover	0.080***	0.014	0.472	0.248*	0.048	1.289
Hit fixed object	0.196**	0.041	0.930	0.469	0.096	2.287
Driver characteristics						
Male						
31 ~ 40 years old	3.536***	1.488	8.400	0.495**	0.251	0.974
41 ~ 50 years old	3.463**	1.319	9.096	0.860	0.390	1.897
≥ 51 years old	4.673**	1.179	18.520	0.380	0.098	1.478
Clerk				0.248**	0.066	0.939
Worker				0.301**	0.110	0.818
Farmer				0.876	0.397	1.932
Self-employed				0.469	0.175	1.258
Migrant worker				0.286	0.064	1.285
Vehicle-specific characteristics						
Large/Medium cargo vehicle				1.683	0.664	4.265
Light cargo vehicle				4.417***	1.439	13.550
Large/Medium passenger vehicle				8.054**	1.618	40.100
Light passenger car				1.420	0.591	3.412
Vehicle with insurance				0.475*	0.204	1.11
Roadway characteristics						
Expressway	1.054	0.472	2.354	1.761	0.772	4.017
Urban expressway	1.244	0.231	6.708	5.651**	1.356	23.550
Urban road	2.428*	0.894	6.593	1.150	0.497	2.661
Special road segment				0.134*	0.016	1.117
Intersection				0.509	0.201	1.286
Motorized vehicle lane						
Non-motorized vehicle lane						
Mixed lane						
Interaction						
Cloudy & Night without street light	6.667***	2.069	21.490			
Constant	0.314	0.069	1.423	1.223	0.229	6.527

*Note: *** Statistically significant at 1% level; ** Statistically significant at 5% level; * Statistically significant at 10% level*

Table 5: Factors Associating with False Positive Fatigue-related Accident Detection

	Fatal and severe injury			Minor injury		
	OR	95% CI		OR	95% CI	
Crash Characteristics						
Motor vehicle crash	0.588	0.157	2.205			
Rollover	0.248*	0.061	1.004			
Hit fixed object	0.227**	0.057	0.902			
Driver characteristics						
Male driver				0.269***	0.122	0.593
31 ~ 40 years old				0.632**	0.406	0.986
41 ~ 50 years old				0.910	0.532	1.556
≥ 51 years old				0.500*	0.231	1.085
Clerk	0.640	0.214	1.912	0.603	0.261	1.393
Worker	1.975**	1.010	3.859	0.476**	0.245	0.921
Farmer	1.518	0.747	3.085	1.588*	0.964	2.617
Self-employed	0.584	0.268	1.274	1.145	0.598	2.193
Migrant worker	1.671	0.774	3.607	1.921*	0.900	4.099
Vehicle-specific characteristics						
Large/Medium cargo vehicle						
Light cargo vehicle						
Large/Medium passenger vehicle						
Light passenger car						
Vehicle with insurance	0.412***	0.232	0.733			
Roadway characteristics						
Expressway	2.088**	1.116	3.908			
Urban expressway	1.652	0.466	5.855			
Urban road	3.127***	1.327	7.369			
Special road segment						
Intersection						
Motorized vehicle lane						
Non-motorized vehicle lane	5.334***	1.558	18.260			
Mixed lane	1.387	0.673	2.861			
Interaction						
Passenger car & Night without street light	4.372*	0.921	20.750			
Urban road & Intersection				3.309**	1.234	8.868
Worker & Large/medium cargo vehicle				4.863**	1.340	17.650
Constant	1.885	0.468	7.599	1.924	0.819	4.521

*Note: *** Statistically significant at 1% level; ** Statistically significant at 5% level; * Statistically significant at 10% level*

4.3 Model evaluation

We conducted likelihood ratio tests to compare the overall fitness between models with interactions and without interactions. The likelihood ratio is calculated as follow:

$$\mathrm{LR} = -2[\mathrm{LL_r} - \mathrm{LL_u}] \tag{4}$$

where $\mathrm{LL_r}$ represents the log-likelihood at convergence of restricted model (model without interactions) and $\mathrm{LL_u}$ is the log-likelihood at convergence of unrestricted model (model with interactions). Under the null hypothesis that the coefficient of interaction is equal to zero, LR statistic is chi-square distributed with degree of freedom equal to the number of interactions. The results of LR statistic and several goodness-of-fit statistics are shown in Table 6.

Table 6: Goodness-of-fit Measures

	N-F Crash		**F-N Crash**	
	Fatal and severe	Minor	Fatal and severe	Minor
	With Interactions			
Number of coefficients	11	22	16	12
McFadden pseudo-R2	0.132	0.128	0.105	0.057
Log-likelihood at convergence	-117.781	-158.768	-223.993	-313.076
Log-likelihood at null	-135.565	-181.971	-250.248	-331.965
Akaike Information Criteria (AIC)	267.150	361.535	479.987	650.152
Bayesian Information Criteria (BIC)	296.853	450.210	542.472	701.267
Hosmer-Lemeshow test (p-value)	0.376, g=13	0.833, g=24	0.401, g=18	0.977, g=14
	Without Interactions			
Number of coefficients	10	22	15	10
McFadden pseudo-R2	0.095	0.128	0.097	0.040
Log-likelihood at convergence	-122.698	-158.768	-225.878	-318.545
Log-likelihood at null	-135.565	-181.971	-250.248	-331.965
Akaike Information Criteria (AIC)	265.397	361.535	481.757	657.091
Bayesian Information Criteria (BIC)	301.232	450.210	540.338	699.687
Hosmer-Lemeshow test (p-value)	0.467, g=12	0.833, g=24	0.612, g=17	0.935, g=12

The results indicate that models with interactions outperformed the models without interactions in both fatal and severe injury accident sample and minor injury accident sample. As shown in Table 6, the p-value of Hosmer-Lemeshow Test (Hosmer & Lemeshow, 1980) for all four models is greater than 0.05 that show no evidence of poor fit.

5. Discussion

5.1 False negative fatigue-related accident detection

With regards to crash type, the odds for rollover or hitting fixed object fatigue-related accident with fatal and severe injury being misclassified are lower than being correctly classified (OR=0.080 and OR=0.196, respectively). Some researchers had already pointed out that single car accidents were closely associated with driver fatigue (Radun et al. 2009; Armstrong et al. 2008), and hitting fixed object and rollover were two major types of single vehicle accidents. A statistics from Australia has shown that hitting fixed object accidents and rollover accidents accounted for 54% and 28% of all fatigue-related accidents during 2005-2009 in South Australia (Government of South Australia, 2010). In line with those findings, these two types of accidents have lower odds of being misclassified into non-fatigue accidents than other types of accidents. In serious injury accidents, lacking witnesses makes it almost impossible to observe the syndrome of fatigue from a dead body and determining accident causes based on some specific types of crash seems to be a useful and effective way. However, rollover was found to be significant only at 90% level (OR=0.248) and hitting fixed object is not significant for minor injury accidents. In a minor injured accident, besides of crash types, police officers might also ask the witness for relevant information about driver fatigue condition to assist their judgment.

Driver's age and occupation are found to have a significant influence on false negative fatigue-related detection for serious injured accidents. The odds of false negative fatigue-related accident detection for drivers in age groups of 31-40 (OR=3.536), 41-50 (OR=3.463), ≥51 (OR=4.673) are higher than younger drivers (≤30). Young drivers were believed to be frequently involved in fatigue-related accidents because of their lifestyle (Horne & Reyner, 1995; Maycock, 1996; McKernon, 2009). Moreover, some practical guidance for identifying for police officers also suggested young driver to be one of the high-risk groups. On the other hand, compared to the young drivers, drivers in other age groups may not attract sufficient attention. Table 2 shows that drivers who are 31-40 years old and 41-50 years old occupied 37.0% and 22.3% of all fatigue-related fatal and severe injured accidents, and 37.0% and 18.3% for minor injured accidents. Without enough attention, fatigue-related accidents in which the responsible drivers who are older than 31 years old, have high odds for being misclassified into non-fatigue accidents than younger drivers. On the contrary, among minor injury accidents, the odds of a fatigue-related accident involving drivers aged between 31-40 being misclassified into non-fatigued accidents is lower than younger drivers (OR=0.495). For clerk and normal worker, the odds of being misclassified into non-fatigue-related accidents are 0.248 and 0.301 the odds for other occupation in minor injured accidents. Studies showed that shift workers and commercial vehicle drivers were more likely to drive under fatigue (Morrow & Crum, 2004; Philip, 2005). According to the coding rules of this database, both professional drivers and shift workers were coded as "worker" or "clerk". If the driver belongs in these two types of occupations and there is no any other extra information for determining accident cause, police officers may easily connect them to fatigue driving. Thus, they are less likely to be misclassified into other causes.

The odds for both large/medium passenger vehicle and light cargo vehicle involving fatigue-related accidents being assigned to other cause are approximately 8.05 times and 4.42 times the odds for other types of vehicles in minor injured accidents (OR=8.054 and OR=4.417, respectively). Even though large/medium cargo or passenger vehicles are recommended to install driving recorders by road management authorities, not all of them would actually install them since it is not mandatory, especially for privately-owned cargo vehicles. Driving records can help to identify the cause of accidents. However, police officers may not have enough time to check them at accident scene due to complaints from drivers and passengers. Therefore, a more common way for officers is to ask the drivers how long they had driven or whether they felt fatigued or sleepy at the crash scene to determine whether fatigue involved in the crash (Robertson et al. 2009). For fatigue-related accidents involving light cargo vehicles, they are also easy to be assigned to other cause of the accident. This may be partly due to the difficulty in proving fatigue driving behavior without driving recorders as well as a similar working pattern as heavy cargo vehicles. Previous research noticed that the long and monotonous journey made heavy cargo vehicles more likely to involve in fatigue-related accidents (Summala & Mikkola, 1994; Chang & Mannering, 1999). The combination effects make fatigue-related accident involving those vehicles easily be misclassified.

The odds of minor injured fatigue-related accidents took place on urban expressways being misclassified into other cause is 5.65 times higher than correctly classified (OR=5.651). These accidents often occurred on expressways because of the average trip length and high-speed limit (Pack et al. 1995; Diamantopoulou et al. 2003). Expressway has been widely considered to be of high risk of fatigue-related accident (NCSDR/NHTSA, 1998). However, less attention has been paid to urban expressways which have similar road condition with expressway, that driving on them also easily leads to driver fatigue (Li et al. 2010). With the rapid urbanization in Guangdong, more and more urban expressways were built to serve the city traffic. Therefore, police officers have lower sensitivity and fail to correctly identify driver fatigue when crashes occurred on urban expressways for minor injured accidents, which make them easily being misclassified.

5.2 False positive fatigue-related accident detection

The odds of fatal and severe injured non-fatigue hitting fixed object crash being misclassified into the fatigue-related accident is lower than other types of crashes (OR=0.227). Rollover accidents are also found significant at 90% level. This finding is similar to false negative fatigue-related accident detection indicating that crash types is a good indicator for identifying fatigue-related accidents.

Some driver characteristics significantly contribute to false positive fatigue-related accident detection. Accidents involving drivers whose occupation were categorized as "worker" have higher odds for false positive fatigue detection in fatal and severe injury accidents (OR=1.975) than other occupations. A survey conducted by police officers in Ontario confirmed that approximately 61% of them believed that night or shift workers tended to involve in fatigue-related accidents (Robertson et al. 2009). This image of shift workers is in line with some previous research indicating that workers with non-fixed working schedule were more likely to have sleep problems (Marcus & Loughlin, 1996; McCartt et al. 1996; Dalziel & Job, 1997), which contributed to fatigue-related accidents (Connor et al. 2001). Thus, these accidents are easily being considered as fatigue-related. For minor injury accidents, the odds of workers involving accidents being false positive fatigue detected is lower than other occupations (OR=0.476) even if police officers can ask those drivers about the fatigue condition at the crash scene. Moreover, farmers (OR=1.588, significant at 90% level) and migrant workers (OR=1.921) have higher odds of being misclassified into other cause in minor injured accidents. Some common features are shared by farmers and migrant workers: low salary, non-fixed working schedule, and low social status. Drivers with these features are widely thought to be related to fatigue driving, and this stereotype can influence the judgment of police officers on fatigue-related accident detection. Therefore, police officers tend to believe that they are more likely to involve in traffic violation related to fatigue. The odds of male drivers for false positive fatigue detection is significantly lower than the odds of female drivers (OR=0.269). While the male driver was believed were more at risk of driving while fatigued (Robertson et al. 2009; Horne & Reyner, 1995), they also found to be at high risk of other violations. As a result, they were not easily to misclassify into fatigue-related accidents. Drivers' age (31-40, OR=0.632) also shows similar results that the odds of false positive fatigue-related accident detection is lower for them compared to young drivers, which is similar to the previous discussion.

The odds of a vehicle with insurance for false positive fatigue-related accident detection is lower than vehicles without insurance for fatal and severe injury accident (OR=0.421). According to the insurance claim process, investigators from insurance companies need to do site investigation which may help police officer to determine the crash cause better.

As for roadway characteristics, expressways and urban roads have higher odds of false positive fatigue-related accident detection in fatal and severe injury accidents (OR=2.088 and OR=3.127) since many researches discussed the relationship between urban expressways and fatigue-related accident detection and their potential danger may be overemphasized. In addition, there are dozens of monitoring facilities on expressways and urban roads, it is still difficult to identify whether the cause of accidents is fatigue-related at the crash scene immediately. Accidents occurred in non-motorized vehicle lanes have higher odds of false positive fatigue-related accident detection (OR=5.334) since most of the fatigue-related accidents since most of the fatigue accident are vehicle-related.

5.3 Interactions

Even though the effect of some individual factors may not have significant impacts, their combination with other factors did influence police officers' judging on determining fatigue-related accidents. For example, lightening condition and weather do not show significant influence for failing to recognize fatigue-related accidents. But driving at night without street light in a cloudy day was identified to contribute to false negative fatigue-related accident detection in serious accidents (OR=6.667). Therefore, when a serious accident take place at night without street light in a cloudy day, the police officer should carefully consider fatigue might be one of the causes of accidents.

Some other interactive factors are identified to hamper the judgment of police officers. An accident involving passenger car during night time without street lighting are found more likely to be false positive fatigue-related accident detection for fatal and severe injured accidents (OR=4.372, significant at 90% level). For minor injured accident, accidents occurred on interactions of urban roads (OR=3.309) is easy to be considered as fatigue-related accidents when fatigue is actually not the primary cause. In addition, non-fatigue accidents in which driver is labeled as "worker" that drives trucks or other large size cargo vehicles, are more likely to be considered as fatigue-related (OR=4.836). Commercial truck drivers are commonly believed to be associated with fatigue driving, thus, they are also more easily to be mistakenly believed to fatigue driving. One possibility is commercial large/medium cargo vehicle drivers are more skillful that they have enough experience and ability to avoid fatigue-related accidents. Therefore, fatigue may not be the major cause of accidents that have these three interactive features, and other possible causes of accidents should be considered.

6. Conclusion and practical applications

Due to lack of proper criteria, the identification of fatigue-related accident by police officers largely depend on inferential evidence and their own experience, and may even lead to incorrect classification of the cause of accidents. Even though some risk factors identified were believed to contribute to fatigue-related accidents, less research has been done to prove that whether these factors are helpful for fatigue-related accident identification. The purpose of this study is to find out factors affecting police officers' judgment when dealing with fatigue-related accidents.

The results show that single vehicle rolling over or hitting fixed object accidents are good indicators for determining fatigue involvement. Accidents that include two or more vehicles have not been found to have significant influence both misclassification types since fatigue could be one of the causes of accidents that have not been noticed. Driving light cargo vehicle, driving large/medium passenger vehicle and urban expressway should attract more attention on determining whether an accident is fatigue-related. Moreover, some stereotypes should be abandoned. Some occupations (e.g. workers, farmer and migrant workers) should not be labeled as "high risk of fatigue driving" when investigating cause of accidents. Expressways and urban roads are also easy to be viewed as high-risk places for fatigue-related accidents. These images will hinder the judgment of fatigue detection whey there are used as an evidence for convicting an accident is fatigue-related. Fatigue should be considered as a possibility rather than a conclusion for cause of the accidents. We also recognized some interactive effects between variables that may also affect fatigue-related accident detection. However, it should be re-emphasized that significant combinatorial effects of factors in this study may only reveal the characteristics of a small subset of fatigue-related accidents. More rules can be generated based on this dataset and it would be better to be used as additional information for training police officers to identify fatigue-related accidents correctly.

Based on the findings from this study, some countermeasures should be considered to improve the fatigue-related accident detection. A clear and easy to implement fatigue definition is an essential solution to this problem. However, up until now, we still do not have a completely general method to quantify driver fatigue and to determine what kind of fatigue level should be considered as fatigue driving. Finding an appropriate definition and quantification method for driver fatigue is one of the challenges in fatigue research during the coming years. In the current stage, more partial countermeasures are needed to improve fatigue-related accident detection. First of all, raising fatigue driving violation penalty for passenger vehicles can prevent drivers' from fatigue driving as well as stimulate police officer to put more attention on identifying the involvement of fatigue in an accident. Secondly, providing training for identifying fatigue in traffic accidents can be beneficial for more police officers and give them a better understanding of fatigue considering the experience of Finland (Radun & Radun, 2013). In this case, misleading factors in identifying fatigue-related accident should be addressed in training process. Moreover, useable and reliable vehicle-based fatigue measurement devices should be encouraged. These devices not only can be used in monitoring drivers' behaviors and the level of driver fatigue by placing sensors on the steering wheels and acceleration pedals, but also provide useful information for traffic police officers to determine the role of fatigue in an accident (Liu et al., 2009; Sahayadhas et al., 2012). More information should be collected to identify fatigue such as the length of time spent driving, detail previous work, sleeping condition and rest schedules of the drivers involved.

There are several limitations which need to be acknowledged. Firstly, for the purpose of identifying factors contributing to the detection of fatigue-related accidents, N-F crashes and F-N crashes are compared with F-F crashes in our study. However, we have not examined whether these factors also influence police officers' judgment on other types of crash. To understand whether those factors are unique for fatigue-related accidents, comparison studies of factors contributing to misclassification of other types of crashes should be conducted in the future. Furthermore, the misclassification scale of property only fatigue-related accident may be underestimated. Filtness et al. (2015) also mentioned that identifying fatigue among less serious accidents may be inaccurate. Moreover, due to the difficulties in proving fatigue as a contributor in an accident, it is possible that not all fatigue-related accidents have been detected even after full investigation process. Some fatigue-related accidents cannot be detected even after in-depth investigation since there is no extra information to evaluate the scale of miscoding problem. Thus, additional data should be collected (e.g. self-report fatigue questionnaire) for better assessing fatigue identification. Several important variables should be considered such as time of day and pre-crash activity. These variables can provide valuable information for identify fatigue-related accident. Unfortunately, they are not included in our dataset.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China, grant 71573286. The first author would like to thank the China Scholarship Council (CSC) for financial support.

Reference

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, pp. 207-216.

Armstrong, K., Filtness, A.J., Watling, C.N., Barraclough, P., & Haworth, N. (2013). Efficacy of proxy definitions for identification of fatigue/sleep-related crashes: An Australian evaluation. Transportation Research Part F: Traffic Psychology and Behaviour, 21, 242-252.

Armstrong, K.A., Smith, S.S., Steinhardt, D.A., & Haworth, N.L. (2008). Fatigue crashes happen in urban areas too: characteristics of crashes in low speed urban areas. 2008 Road Safety Conference.

Australian Transport Safety Bureau (2006). Characteristics of fatal road crashes during national holiday periods. Canberra: Commonwealth of Australia.

Chang, L.Y., & Mannering, F. (1999). Analysis of injury severity and vehicle occupancy in truck-and non-truck-involved accidents. Accident Analysis & Prevention, 31(5), 579-592.

Changchien, S.W., Lu, & T.C. (2001). Mining association rules procedure to support on-line recommendation by customers and products fragmentation. Expert Systems with Applications, 20 (4), 325-335.

Changpetch, P. & Lin, D.K., (2013a). Model selection for logistic regression via association rules analysis. Journal of Statistical Computation and Simulation, 83 (8), 1415-1428.

Changpetch, P., & Lin, D.K. (2013b). Selection of multinomial logit models via association rules analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 5 (1), 68-77.

Connor, J., Norton, R., Ameratunga, S., Robinson, E., Wigmore, B., & Jackson, R. (2001). Prevalence of driver sleepiness in a random population-based sample of car driving. Sleep, 24 (6), 688-694.

Crummy, F., Cameron, P.A., Swann, P., Kossmann, T., & Naughton, M. (2008). Prevalence of sleepiness in surviving drivers of motor vehicle collisions. Internal Medicine Journal, 38 (10), 769-775.

DaCoTA (2012). Fatigue. Deliverable 4.8h of the EC FP7 Project DaCoTA. http://www.dacota-project.eu/Deliverables/Webtexts/Fatigue.pdf

Dalziel, J.R., & Job, R.S. (1997). Motor vehicle accidents, fatigue and optimism bias in taxi drivers. Accident Analysis & Prevention, 29 (4), 489-494.

Das, S., & Sun, X. (2014). Investigating the pattern of traffic crashes under rainy weather by association rules in data mining. In: Proceedings of the Transportation Research Board 93rd Annual Meeting.

De Oña, J., López, & G., Abellán, J. (2013). Extracting decision rules from police accident reports through decision trees. Accident Analysis & Prevention, 50, 1151-1160.

Diamantopoulou, K., Hoareau, E., Oxley, P. E., & Cameron, M. H. (2003). The Feasibility of Identifying Speeding-Related & Fatigue-Related Crashes in Police-Reported Mass Crash Data. (197 ed.) Clayton Vic Australia: Monash University.

Filtness, A.J., Armstrong, K., Watson, A., & Smith, S. (2015). Sleep-related crash characteristics: Implications for applying a fatigue definition to crash reports. Accident Analysis & Prevention, 99, 440–444.

Government of South Australia (2010). Fatigue-related Crashes in South Australia 2010, from http://www.dpti.sa.gov.au/__data/assets/pdf_file/0010/51022/Fatigue_Fact_Sheet_2010.pdf

Hahsler, M., Grün, B., & Hornik, K. (2007). Introduction to arules–mining association rules and frequent item sets. SIGKDD Explorations, 2 (4).

Horne, J., & Reyner, L. (1999). Vehicle accidents related to sleep: A review. Occupational and Environmental Medicine, 56 (5), 289-294.

Horne, J.A., & Reyner, L.A. (1995). Sleep related vehicle accidents. BMJ, 310 (6979), 565-567.

Hosmer, D.W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. Communications in Statistics-Theory and Methods, 9 (10), 1043-1069.

Kamei, Y., Monden, A., Morisaki, S., & Matsumoto, K.-I. (2008). A hybrid faulty module prediction using association rule mining and logistic regression analysis. In: Proceedings of the Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, pp. 279-281.

Li, D. H., Liu, Q., Yuan, W., & Liu, H. X. (2010). Relationship between fatigue driving and traffic accident. Journal of Traffic and Transportation Engineering, 2, 021.

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. Journal of Safety Research, 40(4), 239-245.

Marcus, C.L., & Loughlin, G.M. (1996). Effect of sleep deprivation on driving safety in housestaff. Sleep, 19 (10), 763-766.

McCartt, A.T., Ribner, S.A., Pack, A.I., & Hammer, M.C. (1996). The scope and nature of the drowsy driving problem in new york state. Accident Analysis & Prevention 28 (4), 511-517.

McKernon, S. (2009). A literature review on driver fatigue among drivers in the general public. Land Transport New Zealand.

Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. Transportation Research Record: Journal of the Transportation Research Board, (2237), 107-116.

Montella, A., Aria, M., D’ambrosio, A., & Mauriello, F. (2012). Analysis of powered two-wheeler crashes in italy by classification trees and rules discovery. Accident Analysis & Prevention, 49, 58-72.

Morrow, P. C., & Crum, M. R. (2004). Antecedents of fatigue, close calls, and crashes among commercial motor-vehicle drivers. Journal of Safety Research, 35(1), 59-69.

National Sleep Foundation (2008). Sleep in America Poll - Sleep, Performance and Workplace. Sleep Health: Journal of the National Sleep Foundation, 1 (2), e7.

NCSDR/NHTSA (1998). Drowsy driving and automobile crashes. NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness. DOT Report HS, 808, 707.

Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., & Schwab, C.W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. Accident Analysis & Prevention, 27 (6), 769-775.

Pande, A., & Abdel-Aty, M. (2009). Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. Safety Science 47 (1), 145-154.

Philip, P. (2005). Sleepiness of occupational drivers. Industrial Health, 43(1), 30-33.

Radun, I., Ohisalo, J., Radun, J., Wahde, M., & Kecklund, G. (2013). Driver fatigue and the law from the perspective of police officers and prosecutors. Transportation Research Part F: Traffic Psychology and Behaviour, 18, 159-167.

Radun, I., & Radun, J.E. (2009). Convicted of fatigued driving: Who, why and how? Accident Analysis & Prevention, 41 (4), 869-875.

Radun, I., Radun, J., Ohisalo, J. (2009). Fatigued driving and the law. Drivers punished of fatigued driving. Liikenneturvan Tutkimusmonisteita, (107/2009).

Robertson, R., Holmes, E., & Vanlaar, W. (2009). The facts about fatigued driving in Ontario: A guidebook for police.

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. Sensors, 12(12), 16937-16953.

Summala, H., & Mikkola, T. (1994). Fatal accidents among car and truck drivers: effects of fatigue, age, and alcohol consumption. Human Factors, 36(2), 315-326.

Shaharanee, I.N.M., Hadzic, F., & Dillon, T.S. (2009). Interestingness of association rules using symmetrical tau and logistic regression. In: Proceedings of the Australasian Joint Conference on Artificial Intelligence, pp. 422-431.

Stegeman, I., Wijkerslooth, T.R., Stoop, E.M., Leerdam, M., Ballegooijen, M., Kraaijenhagen, R.A., Fockens, P., Kuipers, E.J., Dekker, E., & Bossuyt, P.M. (2013). Risk factors for false positive and for false negative test results in screening with fecal occult blood testing. International Journal of Cancer, 133 (10), 2408-2414.

Traffic Management Bureau, Ministry of Public Security, PRC (2008). China road traffic accidents annual statistical report 2007.

Weng, J., Zhu, J. Z., Yan, X., & Liu, Z. (2016). Investigation of work zone crash casualty patterns using association rules. Accident Analysis & Prevention, 92, 43-52.

Yan, Q.-H., Ma, W.J., & Xu, Y.J. (2010). Incidence of unsafe driving behaviors and its related risk factors. Chinese Journal of Public Health, 8, 033.

Zhang, G., Yau, K.K., Zhang, X., & Li, Y. (2016). Traffic accidents involving fatigue driving and their extent of casualties. Accident Analysis & Prevention, 87, 34-42.