SPATIAL-TEMPORAL ANALYSIS OF DRINK-DRIVING PATTERNS IN HONG KONG

Y.C. LI
Department of Civil Engineering, The University of Hong Kong,
Pokfulam Road, Hong Kong
E-mail: joeyliyc@hku.hk; Tel.: +852 2859-2662; Fax: +852 2517-0124

N.N. SZE (Corresponding author)
Department of Civil and Natural Resources Engineering, University of Canterbury,
Christchurch, New Zealand
E-mail: tony.sze@canterbury.ac.nz; Tel.: +64-3-364 2238; Fax: +64-3-3642758

S.C. WONG
Department of Civil Engineering, The University of Hong Kong,
Pokfulam Road, Hong Kong
E-mail: hhecwsc@hku.hk; Tel.: +852 2859-1964; Fax: +852 2559-5337

ABSTRACT

Normally, bars and restaurants are the preferred locations for drinking. Therefore, there is concern that the roads in bar and restaurant areas could have a higher probability of drink-drivers and alcohol-related road crashes. Many studies have been conducted to model the association between drinking locations and the prevalence of drink-driving, so that cost-effective enforcement strategies can be developed to combat drink-driving. In this study, a cluster analysis approach was applied to model the spatial-temporal variation of drink-driving distribution in Hong Kong. Six spatial-temporal clusters of drink-driving distribution emerged from the data: (i) bar and restaurant area, weekend-overnight; (ii) bar and restaurant area, other timespan; (iii) urban area, weekend-overnight; (iv) urban area, other timespans; (v) rural area, weekend-overnight; and (vi) rural area, other timespans. Next, separate zero-inflated regression models were established to identify the factors contributing to the prevalence of drink-driving for each of the six recognized clusters. The results indicated that drivers in rural areas tend to consume more alcohol than those in urban areas, regardless of the time period. In addition, both seasonal variation and vehicle class were found to determine the breath alcohol concentration (BrAC) levels among drivers.

Keywords: Drink-driving; Cluster analysis; Zero-inflated regression model; Random breath test

1 Introduction

Bar and restaurant areas are the preferred drinking locations, and there is high concern that these areas may have a large number of drink-driving offences. In view of this concern, many studies have been conducted to measure the association between drinking locations and drink-driving behavior (Gruenewald and Ponicki, 1995; Gruenewald and Treno, 2000; Lee et al., 1997; Chen et al., 2010). A study by Baum (1999) confirmed that the frequency of drink-driving incidents in Australia’s central business districts (CBDs) and tourist resort areas was higher than in other locations. Furthermore, studies of historical crash records have affirmed that the local availability of alcoholic drinks can increase an area’s risk of road crashes (Scribner et al., 1994; Jewell and Brown, 1995; Gruenewald and Johnson, 2006; Treno et al., 2007; Taylor et al., 2010). A local Hong Kong study showed that about 10.3% of all road crashes involving death were attributed to alcohol (Cameron, 2004). From 2007 to 2011, 162 road crashes were found to be related to drink-driving, of which 32 (20%) were KSI (killed or seriously injured) crashes (Transport Department, 2012). Recent studies have revealed the association between drink-driving and crash risk (Li et al., 2012) and injury severity (Tsui et al., 2010). Although an understanding of drink-driving patterns is indispensable to the analysis of alcohol-related crash risks, little research has been conducted on drink-driving patterns with respect to geographical area, road class, or temporal distribution.

It is possible to develop cost-effective anti-drink-driving strategies or policies to reduce alcohol-related road crashes through identification of areas with high accessibility to selling points of alcohol, and examination of associated drink-driving patterns. In Hong Kong, pubs, bars and restaurants with valid liquor license to sell or supply liquor for consumption are concentrated in Wanchai (14.3%), and Tsim Sha Tsui (18.3%) (Shaded in black colour in Figure 1) (Liquor Licensing Board, 2013). However, quantitative studies are lacking on the relationship between drink-driving patterns, geographical areas, road types, and selling points of alcohol.

[Insert Figure 1 here]

Some cross-sectional studies have explored possible contributing factors such as age and gender that may be correlated to the likelihood of drink-driving (Begg et al., 2003; Peck et al., 2008; Kim et al., 2010; Moan and Rise, 2011). In Belgium, a multilevel discrete choice model was established to examine the association between drink-driving, driver demographics, and driver perceptions, based on data from a roadside survey (Vanlaar, 2005). Further understanding of the relationships between the spatial-temporal distribution of drink-driving incidents and contributory factors should be worth exploring. The Hong Kong Police Force collects comprehensive drink-driving data, based on round-the-clock random breath tests (RBTs) at roadblocks throughout the territories. By taking advantage of this data, it is possible to obtain detailed information on drink-driving incidents, including the drivers' breath alcohol concentrations (BrAC), times, and locations. This data resource offers an opportunity for us to model the spatial-temporal drink-driving patterns in Hong Kong.

This study established an integrated approach for capturing the spatial-temporal variation of drink-driving distributions and the factors contributing to the prevalence of drink-driving. First, the drink-driving data were mapped to the respective spatial-temporal units by use of a Geographical Information System (GIS) program. Then, cluster analysis was conducted to identify clusters of drink-driving distributions. The method and the results of the cluster analysis for the identified spatial-temporal patterns of drink-driving distributions are explained in Section 2. In Section 3, zero-inflated regression models help to identify factors contributing to the prevalence of drink-driving and of high breath alcohol concentration levels in each of the recognized clusters. Finally, we conclude the study by providing recommendations for further research in Section 4.

2 Cluster Analysis for Distinguishing Drink-Driving Patterns

2.1 Random Breath Tests (RBTs) and Drink-Driving Distributions

In Hong Kong, the first legal drink-driving limit was introduced in 1995, requiring that a driver's breath alcohol concentration (BrAC) be no higher than 35 μg/100ml. This legal limit was subsequently lowered to 22 μg/100ml in 1999. In February 2009, new legislation came into effect that empowered the police to conduct random breath tests (RBTs). All RBTs are conducted at roadblocks or checkpoints, at which the police may stop any vehicle at any time, even in the absence of evidence or reasonable cause to suspect that the driver has consumed alcohol, and all stopped drivers are tested. In this study, information on 33,472 RBT measurements at the roadblocks, which were evenly distributed throughout the territories, around the clock, during the one-year period from February 9, 2009 to February 8, 2010 was obtained. Each of the 33,472 records refers to a single BrAC measurement of individual driver stopped at the roadblocks, and information on the driver's breath alcohol concentration level, time and location of measurement, driver gender and vehicle type was recorded. By use of a Geographical Information System (GIS), these 33,472 records were mapped onto a spatial-temporal grid with respect to test location (144 levels), day of the week (7 levels) and time of a day (6 levels), using the Geographical Information System (GIS) technique. For each spatial-temporal unit, as shown in Table 1, a profile of breath alcohol concentration (BrAC) frequency distribution could be set out. In particular, a single distribution profile of spatial-temporal unit *k* could be denoted by,

$$\overrightarrow{\mathrm{N}}_{\mathrm{k}} = \{\mathrm{n}_{\mathrm{k},1}, \mathrm{n}_{\mathrm{k},2}, \dots, \mathrm{n}_{\mathrm{k},21}\} \forall \mathrm{k} \in 1,2, \dots \mathrm{K} \tag{1}$$

where *K* equal to 6,048, and $\sum_k \sum_{i=1}^{21} n_{k,i} = \sum_k n_k = 33{,}472$

[Insert Table 1 here]

Not all the 6,048 spatial-temporal unit have measurement because there was no RBT operation in certain time periods at particular location, we therefore have 745 drink-driving distribution profiles set out. Table 2 summarizes the characteristics of the 745 drink-driving distribution profiles.

[Insert Table 2 here]

2.2 Cluster Analysis

Cluster analysis is an exploratory data evaluation technique for grouping individuals into meaningful clusters, such that individuals within a cluster are similar in some respects and dissimilar to individuals in other clusters. In terms of road safety research, many studies (Sigve and Torbjorn, 2007; Depaire et al., 2008; Wong and Chung, 2008; Anderson, 2009) have been conducted using cluster analysis to develop various crash prediction models. For example, a local Hong Kong study examined the geographical distribution of road crashes using the cluster technique (Ng et al., 2002).

Basically, cluster analyses can be classified into three different types: visual techniques, hierarchical methods, and non-hierarchical methods. In this study, an hierarchical agglomerative technique using the Wald method was used to generate clusters of drink-driving distribution profiles, that we group individuals having similar characteristics into a cluster. The agglomerative approach has been one of the most common cluster analysis methods. Specifically, the degree of similarity between two individuals can be defined by the squared Euclidean distance between their respective centroids, or means (Jarrell, 1994).

$$\text{Euclidean distance, } \mathrm{d}_{\mathrm{ij}} = \left(\sum_{\mathrm{r}=1}^{\mathrm{R}} \left(\mathrm{x}_{\mathrm{ir}} - \mathrm{x}_{\mathrm{jr}}\right)\right)^{\frac{1}{2}} \tag{2}$$

where x_{ir} and x_{jr} refers to the characteristic score of variable *r* of individual *i* and *j* respectively.

To this end, the possible variables in the cluster analysis are the BrAC frequency distribution, district, day of the week, and time period. Following the hierarchical agglomerative approach, the 745 drink-driving distribution profiles were gradually merged into clusters at different steps or stages, until all individuals were assigned to respective clusters. In the initial stage, pairs of individuals that shared the greatest similarity were merged first. Similar pairs or individuals were then being merged in subsequent steps, based on the same mechanism. Eventually, a dendrogram (tree diagram) could be established to illustrate the merging process and the final results of cluster analysis.

2.3 Cluster Analysis Results

A cluster analysis on the 745 drink-driving distribution profiles was performed using statistical software SPSS 20.0.

The variables entailed in the cluster analysis are BrAC frequency distribution, geographical district (18 levels as shown in Figure 1), day of the week (7 levels), and time period (6 levels, i.e., 7:00 a.m.–11:00 a.m. [morning], 11:00 a.m.–3:00 p.m. [afternoon], 3:00 p.m.–7:00 p.m. [evening], 7:00 p.m.–11:00 p.m. [night], 11:00 p.m.–3:00 a.m. [midnight] and 3:00 a.m.–7:00 a.m. [dawn]).

In cluster analysis, parsing the dendrogram to determine the number of clusters is a subjective process, which is likely to be the most common approach (Baxter, 1994). Based on the dendrogram obtained in this study, we therefore proposed to group individuals into six natural clusters: (i) bar and restaurant areas[1], weekend-overnight[2] (Number of profiles = 172, Percentage of total = 23.1%); (ii) bar and restaurant areas, other timespans (98, 13.2%); (iii) urban areas, weekend-overnight (122, 16.4%); (iv) urban areas, other timespans (77, 10.3%); (v) rural areas, weekend-overnight (111, 14.9%); and (vi) rural areas, other timespans (165, 22.1%). Chi-square test was used to assess the independence of drink-driving distribution profiles between different clusters. As Table 3 shows, the chi-square test statistics for pairs of clusters were all greater than the critical value of 32.671 (Degree of freedom = 21, *p-level* = 0.05). We therefore concluded that the six clusters of drink-driving distribution profiles were remarkably distinguished from each other at the 5% level. Figure 2 illustrates the cluster tree adopted in this study.

[Insert Table 3 here]

[Insert Figure 2 here]

Results of cluster analysis formed the basis of the development of separated drink-driving prediction models in the subsequent section.

3 Zero-inflated Models for Factors Contributing to Drink-driving and High BrAC

3.1 Model framework

Based on the characteristics of the above six clusters set out, separated prediction models were established to identify factors contributing to the prevalence of drink-driving and high BrAC levels, including the drivers' characteristics, road types, and vehicle classes involved in each of the different clusters. As driver characteristics do affect the BrAC level, the unit of prediction model is individual drivers.

In the proposed prediction model, the 33,472 RBT records, in each of which comprehensive information on BrAC levels and personal characteristics of the concerned drivers were recorded, were used. Table 4 summarized the characteristics of the 33,472 measurements. The 33,472 measurements were segregated into six: Cluster 1 – bar and restaurant areas, weekend-overnight (Number of measurements = 1,034); Cluster 2 – bar and restaurant areas, other timespans (8,201); Cluster 3 – urban area, weekend-overnight

---

[1] Bar and restaurant areas refers to Yau Tsim Mong (Tsim Sha Tsui and Mong Kok District) and Wan Chai areas, which represent 25.4% and 14.3% bars and restaurants with valid liquor licence in Hong Kong, respectively (Liquor Licensing Board, 2013)

[2] Weekend refers to Friday and Saturday, and overnight refers to the time periods from 11:00 p.m. to 07:00 a.m. of the next day

(965); Cluster 4 – urban area, other timespans (9,469); Cluster 5 – rural area, weekend-overnight (2,359); and Cluster 6 – rural area, other timespans (11,444).

[Insert Table 4 here]

3.2 Zero-inflated (ZI) regression models

3.2.1 Model formulation

In the proposed drink-driving prediction model, the dependent variable was the breath alcohol concentration level, and the independent variables were the season of year (spring, summer, fall or winter), the road type (major and minor), the driver's gender, and the vehicle class (taxis, other commercial vehicles, or non-commercial vehicles). As the dependent variable, breath alcohol concentration (BrAC), is of a positive and continuous nature, we considered three different model forms: (i) lognormal; (ii) gamma; and (iii) exponential. To address the problem that zeroes were prevalent in the BrAC data, zero-inflated modifications were incorporated into the regression models.

Although the application of zero-inflated models for crash predictions was questioned due to the difference in the zero generation process between crash occurrence and manufacture failure (Lambert, 1992), it could have potential to model BrAC levels (Ghosh et al., 2006; Ospina et al., 2012). Indeed, zero BrAC could be generated by two different processes: one is governed by the binary distribution of compliance drivers, while another is governed by the continuous distribution of non-compliance drivers who have zero BrAC. For instance, the formulation of a proposed zero-inflated continuous model can be specified by establishing the following probability function (Ntzoufras, 2009):

$$f_{ZID}(y) = \pi_0 I(y = 0) + (1 - \pi_0) f_D(y|\theta) \tag{3}$$

where $f_D(y|\theta)$ is the probability function of θ and $f_{ZID}(y)$ is the modified probability function, with an additional parameter π_0 indicating the proportion of additional zeros.

Furthermore, the probability of zeros can be expressed in the following form:

$$P(y = 0) = \pi_0 + (1 - \pi_0) f_D(0|\theta), \tag{4}$$

and the probability of $y > 0$ is given by

$$P(y > 0) = (1 - \pi_0) f_D(y|\theta). \tag{5}$$

3.2.2 Bayesian approach

In this study, the Bayesian method was applied, using the Markov Chain Monte Carlo (MCMC) simulation approach to estimate the parameters of the zero-inflated continuous regression models (Scollnik, 2002; Gelman, 2004; Ghosh et al., 2006; Ntzoufras, 2009). These calculations were performed on the *WinBUGS* platform. The MCMC parameter estimate technique generates sequences of random points, whose distributions should converge to the target posterior distributions.

Based on Bayes' theorem, the posterior distribution of parameters can be derived by integrating the prior distribution and likelihood functions, as follows:

$$f(\theta|y) = \frac{f(y,\theta)}{f(y)} = \frac{f(y|\theta)f(\theta)}{f(y)}, \quad (6)$$

where y is the observed outcome and θ is the parameter estimate. The marginal distribution of y can be specified as follows:

$$f(y) = \int f(y|\theta)f(\theta)d\theta, \quad (7)$$

which is a constant with fixed value of y. The posterior distribution can be further expressed as an unnormalized posterior distribution by omitting the factor $f(y)$ as follows:

$$f(\theta|y) \propto f(y|\theta)f(\theta) \quad (8)$$

3.2.3 Goodness-of-fit

Spiegelhalter et al. (2002) proposed the deviance information criteria (DIC) as a measure of model complexity and fit. As with the Akaike or Bayesian information criterion indicators, a lower DIC value indicates a better model fit.

$$DIC = D(\overline{\theta}) + 2p_D = \overline{D(\theta)} + p_D, \quad (8)$$

where $D(\overline{\theta})$ is the usual deviance evaluated at the posterior means of parameter θ, and $\overline{D(\theta)}$ is the posterior mean of deviance, and p_D is the effective number of parameters. Usual deviance can be defined as follows:

$$D(\theta) = -2logf(y|\theta). \quad (9)$$

3.3 Analysis Results

Possible factors contributing to the prevalence of drink-driving and higher BrAC levels were examined, including the season of year (e.g. spring, summer, fall or winter), road type (major road versus minor road), driver gender, and vehicle class (taxi, other commercial vehicle, or non-commercial vehicle). The characteristics of individual drivers (e.g., driver gender and vehicle class), were also incorporated into the probability function for the zero-state. Table 5 illustrates the results of goodness-of-fit analysis for three candidate zero-inflated continuous regression models: (i) Zero-inflated lognormal (ZILogNorm); (ii) Zero-inflated gamma (ZIGamma); and (iii) Zero-inflated exponential (ZIExpon), as applied to the six concerned models. The number of iteration for posterior estimates was 100,000, and the corresponding DIC values are presented in Table 5.

[Insert Table 5 here]

The results of parameter estimates were generated for each cluster, based on the zero-inflated regression model having the lowest DIC value. These results are presented in Table 6..

[Insert Table 6 here]

3.3.1 Overall pattern

As shown in Table 5, the zero-inflated exponential (ZIExpon) regression model exhibited the best fit with the observed breath alcohol concentration data (*DIC* = 15,026.2). Predicted distributions of breath alcohol concentrations and respective root mean square errors (RMSE) are also presented in the Table 6. For instance, less than 3% (the proportion of BrAC positive drivers are 2.67%) of the tested drivers were found positive, and their average breath alcohol concentration was 0.097 μg/100ml.

Table 6 illustrates the results of the zero-inflated exponential regression model of overall breath alcohol concentrations. For the regression state, both seasonal variation and vehicle class were found significantly correlated to BrAC at the 5% significance level. BrAC levels in the spring (coefficient = 29.530), summer (13.670) and autumn (9.018) were all higher than levels found in the winter at the 1% significance level. Also, the BrAC levels of taxi drivers (-2.796) were lower than those of other vehicle drivers at the 1% significance level. However, no evidence could be established for an association between BrAC levels and other factors, including road type and driver gender. For the zero-state, the probabilities of zero BrAC among taxi drivers (1.558) and other commercial vehicle drivers (1.091) were both found to be higher than among non-commercial vehicle drivers at the 1% significance level. This implies that occupational drivers (both taxi and other commercial vehicle drivers), generally have a lower tendency to drive after drinking.

To evaluate the possible effects of spatial and temporal factors on overall BrAC levels, we have incorporated the six cluster types in the regression equation. As also shown in Table 6, BrAC levels in Cluster 1 (bar and restaurant area, weekend-overnight, -1.045) and Cluster 3 (urban area, weekend-overnight, -0.570) were lower than that of other clusters, both at the 1% level of significance. Besides, BrAC levels in Cluster 2 (bar and restaurant area, other timespans, 0.399) and Cluster 4 (urban area, other timespans, 0.252) were higher than that of other clusters, at the 5% level of significance.

3.3.2 Cluster 1 – Bar and restaurant area, weekend-overnight

As shown in Table 5, the zero-inflated gamma (ZIGamma) regression model (DIC = 767.8) was superior to other models for the association measure of BrAC levels in Cluster 1. In this model, the estimated mean BrAC of drink-drivers was 18.460 μg/100ml (refer to Table 6). This measure was remarkably higher than the overall Hong Kong average.

The seasonal variation and vehicle class were both found significantly correlated to BrAC in the regression state at the 5% significance level. The BrAC levels in spring (1.910) and summer (0.904) were noticeably higher than those measured in winter. The BrAC level for taxi drivers (-1.861) was in particular lower than that for other vehicle drivers. For the zero-state, the probability of zero BrAC among taxi drivers in bar and restaurant areas (2.477) was much higher than that of other vehicle drivers, at the 5% significance level.

3.3.3 Cluster 2 – Bar and restaurant area, other timespans

The BrAC distribution of Cluster 2 was similar to that of the overall situation in Hong Kong. As shown in Table 5, the zero-inflated exponential (ZIExpon) regression model was superior (DIC = 2,817.5) for this cluster. Also, only a small proportion (2.11%) of tested drivers were found to be driving after drinking, and their mean estimated BrAC was 0.132 μg/100ml (refer to Table 6).

Again, seasonal variation and vehicle type were found significantly correlated to BrAC in the regression state, both at the 5% significance level. Compared with the winter period, the BrAC levels in other seasons, spring (24.900), summer (20.690), and autumn (6.505), were all substantially higher, at the 5% significance level. For taxi drivers who drink and drive, the predicted BrAC was lower (-3.025) than that of other vehicle drivers, at the 5% significant level. However, the likelihood of zero BrAC among both commercial vehicle (0.840) and taxi (1.028) drivers was noticeably higher than that of non-commercial vehicle drivers, at the 5% significance level. This result is comparable to that of overall probabilities in Hong Kong.

3.3.4 Cluster 3 – Urban area, weekend-overnight

In this cluster, the zero-inflated gamma (ZIGamma) regression model (DIC = 548.2) outperformed all other models. As illustrated in Table 6, the BrAC distribution of Cluster 3 was similar to that of Cluster 1 (bar and restaurant areas, weekend-overnight). The proportion of positive BrAC measurements in Cluster 3 (3.38%) was slightly lower than that of Cluster 1 (5.00%). The mean estimated BrAC was 19.810 μg/100ml, which was comparable to that of Cluster 1.

Furthermore, seasonal variation was found significantly correlated to the BrAC of drink-drivers, at the 5% significance level. The mean BrAC level of drivers in spring (1.212) and summer (1.451) was noticeably higher than that measured in winter at the 5% significance level. No evidence could be established for an association between the mean BrAC level and other factors, including road type, gender or vehicle class. No other factor was found significantly correlated to the prevalence of drink-driving in this cluster.

3.3.5 Cluster 4– Urban area, other timespans

For Cluster 4, the zero-inflated lognormal (ZILogNorm) regression model was found to be the best fit, with the observed BrAC distribution (of DIC = 2,742.4). As shown in Table 6, the proportion of positive BrAC levels for this cluster was the lowest (1.65%) among the six clusters, and the mean estimated BrAC was 1.849 μg/100ml.

Through their model forms can be differentiated, the contributory factors to mean BrAC levels and the likelihood of positive BrAC were similar in Clusters 2 and 4 (other timespans for both bar and restaurant areas and urban areas). For instance, in Cluster 4 both seasonal variation and vehicle class were found as contributing to the mean BrAC level at the 5% significance level. Likewise, the BrAC levels in summer (1.031) and autumn (0.968) were higher than in the winter at the 1% significance level. Furthermore, the mean BrAC of taxi drivers was lower (-0.877) than that of other vehicle drivers at the 5% significance level. For the zero-state, vehicle class was a deterministic factor, and the likelihood of positive BrAC among both commercial vehicle (0.989) and taxi (1.041) drivers was higher than that for non-commercial vehicle drivers, at the 5% and 1% significance levels respectively.

3.3.6 Cluster 5 – Rural area, weekend-overnight

Not surprisingly, the BrAC distributions in Clusters 3 and 5 (the weekend-overnight period in urban and in rural areas) were similar. For instance, the zero-inflated gamma (ZIGamma) regression model best fit with the observed BrAC distribution in Cluster 5 (DIC = 2,186.5). However, both the proportion of drink-drivers (6.06%) and the mean BrAC (20.850 μg/100ml) in Cluster 5 were the highest among all of the six clusters. Obviously, there is cause for great concern over the prevalence of drink-driving during the weekend-overnight period in rural areas.

Again, seasonal variation was found deterministic to the mean BrAC level, at the 1% significance level. The mean BrAC in winter was lower than that in the spring (1.574), summer (0.912) and autumn (0.819). For the zero-state, no evidence could be established for an association between the likelihood of drink-driving and the other potentially contributing factors considered.

3.3.7 Cluster 6 – Rural area, other timespans

The zero-inflated exponential (ZIExpon) regression model fit well with the BrAC distribution in Cluster 6 (DIC = 6,019.7). As illustrated in Table 6, some 3.29% of the tested drivers were found to have positive BrAC, and their mean BrAC level was 0.088 μg/100ml.

Again, seasonal variation was related to BrAC, at the 1% significance level. For instance, the mean BrAC levels in spring (29.990), summer (9.369), and autumn (10.740) were all higher than levels found in the winter at the 1% significance level. Nevertheless, no evidence could be established for the association between BrAC and other factors, including road type, gender or vehicle class. For the zero-state, occupational drivers, including commercial vehicle (1.202) and taxi (2.082) drivers, had a lower likelihood of positive BrAC than non-commercial vehicle drivers in Cluster 6 at the 1% significance level.

4 Discussion

4.1 Contributory Factors to the Prevalence of Drink-driving

Generally, the zero-inflated exponential (ZIExpon) regression model fit well with the observed BrAC data for overall RBT records, in which 2.82% of all tested drivers were found to have positive BrAC. Seasonal variation and vehicle class were found significantly correlated with the BrAC of drink-drivers. According to traditional wisdom, alcohol consumption tends to be higher in Hong Kong’s festive seasons, e.g., Christmas to New Year’s day (December-January) and Chinese New Year (January or February). However, the results of this study reveal that, surprisingly, the BrAC of drink-drivers in the winter festive period (December-February) was lower than in all other seasons, regardless of district or time period. This pattern may be attributed to the more frequent publicity campaign activities and targeted enforcement actions against drink-driving immediately before and during the festive season. Evidence has been established that such targeted enforcement and publicity campaign activities, including use of electronic media, are effective in discouraging drink-driving, and thus alcohol-related road crashes can possibly be reduced (Elder et al., 2004; Tay, 2005). The study’s results also demonstrated that occupational drivers, including taxi and commercial vehicle drivers, have a lower likelihood of drink-driving. Even if convicted of drinking while driving, the BrAC levels of occupational drivers were found noticeably lower than the levels of convicted non-commercial vehicle drivers. In Israel, occupational drivers were found to be more law-abiding than non-commercial vehicle drivers (Rosenbloom et al., 2009), but in Thailand and Nigeria, occupational drivers were found to be remarkably less law-abiding, especially in driving while impaired by alcohol (Ingsathit et al., 2009; Balogun et al., 2012). However, diagnostic analyses on the propensity of convicted driving behavior of occupational drivers are limited. It should be worthwhile for future research to explore the relationship between the likelihood of drink-driving and other possibly contributing factors among occupational drivers.

Investigators of driver behavior have long considered gender a deterministic factor, and have generally found that male drivers are more aggressive, less cautious (Glendon and Cernecca, 2003; Vanlaar, 2005; Fernandes et al., 2010), and more likely to be convicted of drink-driving (Kim et al., 2010) than female drivers. However, no evidence could be found for an association between driver gender and the likelihood of drink-driving in the results of this study, or in some of the other diagnostic analyses of driver behavior

(Kuntsche et al., 2010). It is possible that the sample size of female drivers (5.5% in this study) might have been too small for efficient quantitative analysis.

4.2 Variations in Drink-driving Patterns by Geographical Area

One of the primary objectives of this study was to understand the variations in drink-driving patterns by geographical area, e.g., rural, urban, or bar and restaurant areas. Understanding such patterns is essential for the development of effective and appropriate anti-drink-driving measures, which if implemented can enhance road safety performance in the problematic areas recognized. Therefore, separate prediction models for BrAC levels and the prevalence of drink-driving were developed for different clusters by geographical area and time period. BrAC levels of convicted drivers tended to be higher during weekend-overnight periods, regardless of geographical area. For these periods, zero-inflated gamma (ZIGamma) models had the best predictive performance. In contrast, the zero-inflated lognormal (ZILogNorm) or zero-inflated exponential (ZIExpon) alternatives showed better predictive performance for BrAC levels of convicted drivers in other timespans, regardless of geographical area. This implies that the BrAC distributions in the weekend-overnight periods are more likely to be dispersed. The predicted mean BrAC levels in weekend-overnight periods (ranging from 18 μg/100ml to 21 μg/100ml) were close to the prescribed legal limit (22μg/100ml), but those in all other timespans were far lower (ranging from 0 μg/100ml to 2 μg/100ml). The above findings are consistent with findings from Switzerland and Northern Ireland (Vanlaar, 2005; Evans et al., 2006; Kuntsche et al., 2010).

However, this study's results also indicate that drivers in rural areas, especially during the weekend-overnight period, tend to have higher BrAC levels than those in bar and restaurant areas during the same time periods, even through the intensity of alcohol selling points in the bar and restaurant areas could be higher. This finding seems to contradict the findings of previous studies that suggested the incidence of drink-driving is positively correlated with the availability of alcohol (Gruenewald and Ponicki, 1995; Gruenewald and Treno, 2000; Treno et al., 2007; Schonlau et al., 2008). Jackson and Owens (2011) had revealed that reduction in walking distance to public transport stations can reduce the prevalence of drink-driving in the United States. Different from the U.S. and other countries, all areas (even rural areas) in Hong Kong are well connected by public transport, with very high ridership (over 90% of person trips are by public transport) (Transport Department, 2013). Despite of this, the findings of current study revealed that the prevalence of drink driving in rural areas were higher than that in urban areas. This implies that focusing on the accessibility to public transport is not always the most cost effective approach to combat drink-driving. Besides, at least one transfer might be required for bus routes connecting rural areas, this may discourage drivers from switching to public transport even if the public transport stations are highly accessible. . If this is the case, then better approaches like "dial-a-driver" would be essential to metropolitans like Hong Kong. Hence, drink-drivers would be encouraged to switch to alterative transport modes, instead of driving by themselves after drink.

Conclusively, the results of this study have implications for the transport authorities and for police. These findings should lead to the development of targeted enforcement and campaign activities in rural areas, in which the average BrAC level of convicted drivers was 12.9%, or 2.39 μg/100ml higher than that in other areas. The authorities should also be especially vigilant during the weekend-overnight period.

5 Conclusion

Alcohol-related road crashes and injuries can be viewed as a result of deficiencies in traffic control strategies and in enforcement measures against drink-driving (Holder et al., 2000). To combat drink-driving behavior and alcohol-related road crashes, effective and appropriate remedial measures based on the results of this study can be developed from two perspectives. First, the data showed that the occurrence of illegal BrAC levels and the effects of factors contributing to the prevalence of drink-driving varied with respect to geographical area and time period. Targeted enforcement measures should therefore be imposed in the weekend-overnight periods for a stronger deterrent effect against drink-driving. Also, both the transport authorities and the police should pay greater attention, not only to the bar and restaurant areas, but also to the rural areas in which the BrAC levels of drink-drivers are higher. Second, the results demonstrated that the BrAC levels of drink-drivers in the festive season were lower than those in other periods of the year. This could be favorably attributed to the success of driver safety campaigns and publicity, e.g., through electronic media, in the festive season. Undoubtedly, better driver education and publicity can be effective in enhancing public awareness concerning road accident risks, and thus reducing the likelihood of drink-driving in other seasons as well.

This study has a number of limitations. Generally, factors such as socioeconomics and driving experiences, e.g., income, trip purpose and driving offense record, could be deterministic to the prevalence of drink-driving and the BrAC levels of convicted drivers. Data on these factors, however, were not available for this study. The data on some other potentially contributing factors, including vehicle class and gender, were examined. Occupational drivers were found to have a lower likelihood of drink-driving than non-commercial vehicle drivers. If the relevant information on trip characteristics and driving experience becomes available for future attitudinal surveys, it would be worthwhile exploring the relationship between aggressive driving and the BrAC levels of convicted drivers, and the prevalence of drink-driving among novice drivers. Furthermore, results of association measure could be subject to the influences of differences in the levels of testing between geographical areas and time periods. The deterrent effects of RBT could be different due to the variation in the levels of testing. Last but not least, drink drivers could have avoided the RBT if some RBT sites are easy to avoid. Therefore, the results could be conservative. No information is however available to explicitly reveal their influences on the association.

Acknowledgements

We gratefully acknowledge the Hong Kong Police Force for providing the RBT data. The research described here was supported by a Research Postgraduate Studentship, the Engineering Postdoctoral Fellowship Programme, and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. HKU 7188/11E).

References

Anderson T.K. (2009) “Kernel density estimation and K-means clustering to profile road accident hotspots.” *Accident Analysis and Prevention*, **41**(3), 359-364.

Balogun S.K., Shenge N.A., Oladipo S.E. (2012) “Psychosocial factors influencing aggressive driving among commercial and private automobile drivers in Lagos metropolis.” *The Social Science Journal*, **49**(1), 83-89.

Baum A. (1999) “An aggregate level analysis of the socioeconomic correlates of drink driving offenders.” *Accident Analysis and Prevention*, **31**(3), 213-220.

Baxter M.J. (1994) *Exploratory Multivariate Analysis in Archaeology*. Edinburgh University, Edinburgh.

Begg D.J., Langley J.D., Stephenson S. (2003) “Identifying factors that predict persistent driving after drinking, unsafe driving after drinking, and driving after using cannabis among young adults.” *Accident Analysis and Prevention*, **35**, 669-675.

Cameron P.A., Rainer T.H., Mak P. (2004) “Motor vehicle deaths in Hong Kong: opportunities for improvement.” *Journal of Trauma*, **56**(4), 890-893.

Chen M.J., Grube J.W., Gruenewald P.J. (2010) “Community alcohol outlet density and underage drinking.” *Addiction*, **105**(2), 270-278.

Depaire B., Wets G., Vanhoof K. (2008) "Traffic accident segmentation by means of latent class clustering." *Accident Analysis and Prevention*, **40**(4), 1257-1266.

Elder R.W., Shults R.A., Sleet D.A., Nichols J.L., Thompson R.S., Rajab W., Task Force on Community Preventive Services (2004) "Effectiveness of mass media campaigns for reducing drinking and driving and alcohol-involved crashes: a systematic review." *American Journal of Preventive Medicine*, **27**(1), 57-65.

Evans A., Marques-Vidal P., Ducimetière P., et al. (2006) "Patterns of alcohol consumption and cardiovascular risk in Northern Ireland and France." *Annals of Epidemiology*, **17**(5)S, S75-S80.

Fernandes R., Hatfield J., Soames Job, R.F. (2010) "A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers." *Transportation Research Part F: Traffic Psychology and Behaviour,* **13**(3), 179-196.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (2004) *Bayesian Data Analysis* (2nd Edition). Chapman & Hall/CRC, Boca Raton, FL.

Ghosh S.K., Mukhopadhyay P., Lu J.C. (2006) "Bayesian analysis of zero-inflated regression models." *Journal of Statistical Planning and Inference*, **136**, 1360-1375.

Glendon A.L., Cernecca L. (2003) "Young drivers' responses to anti-speeding and anti-drink-driving messages." *Transportation Research Part F: Traffic Psychology and Behaviour,* **6**(3), 197-216.

Gruenewald P.J., Johnson F.W. (2006) "The stability and reliability of self reported drinking measures." *Journal of Studies on Alcohol*, **67**(5),738-745.

Gruenewald P.J., Ponicki W.R. (1995) "The relationship of the retail availability of alcohol and alcohol sales to alcohol-related traffic crashes." *Accident Analysis and Prevention,* **27**(2), 249–259.

Gruenewald P.J., Treno A.J. (2000) "Local and global alcohol supply: economic and geographic models of community systems." *Addition*, **12**(4), 537-549.

Holder H.D., Gruenewald P.J., Ponicki W.R., Treno A.J., Grube J.W., Saltz R.F., Voas R.B., Reynolds R., Davis J., Sanchez L., Gaumont G., Roeper P. (2000) "Effect of community-based interventions on high-risk drinking and alcohol-related injuries." *Journal of the American Medical Association,* **284**, 2341-2347.

Ingsathit A., Woratanarat P., Anukarahanonta T., Rattanasiri S., Chatchaipun P., Wattayakorn K., Lim S., Suriyawongpaisal P. (2009) "Prevalence of psychoactive drug use among drivers in Thailand: a roadside survey." *Accident Analysis and Prevention,* **41**(3), 474-478.

Jackson C.K., Owens E.G. (2011) "One for the road: public transportation, alcohol consumption, and intoxicated driving." *Journal of Public Economics,* **95**(1-2), 106-121.

Jarrell S.B. (1994) *Basic Statistics*. Wm. C. Brown Communication, Inc., USA.

Jewell, R.T., Brown, R.W. (1995) "Alcohol availability and alcohol-related motor vehicle accidents." *Applied Economics*, **27**, 759-765.

Kim J.H., Lee S, Chan K.W.C., Lau J., Tsang A., Griffiths S.M. (2010) "A population-based study on the prevalence and correlates of drinking and driving in Hong Kong." *Accident Analysis and Prevention,* **42**(4), 994-1002.

Kuntsche E., Cooper M.L. (2010) "Drinking to have fun and to get drunk: motives as predictors of weekend drinking over and above usual drinking habits." *Drug and Alcohol Dependence,* **110**(3), 259-262.

Lambert D. (1992) "Zero-inflated Poisson regression, with an application to defects in manufacturing." *Technometrics*, **34**, 1-14.

Lee J.A., JonesWebb R.J., Short B.J., Wagenaar, A.C. (1997) "Drinking location and risk of alcohol-impaired driving among high school seniors." *Addictive Behaviors*, **22**(3), 387-39.

Li Y.C., Sze N.N., Wong S.C., Tsui K.L., So F.L. (2013) "Effects of drink driving on crash risk based on random breath test data." *Hong Kong Journal of Emergency Medicine* **20**, 146-154.

Liquor Licensing Board. (2013). *List of Premises with Liquor Licence. Available at* <http://www.fehd.gov.hk/english/LLB_web/epremis.html> (Accessed on 06 April 2013).

Moan I.S., Rise J. (2011) "Predicting intentions not to 'drink and drive' using an extended version of the theory of planned behaviour." *Accident Analysis and Prevention,* **43**(4), 1378-1384.

Ng K.S., Hung W.T., Wong W.G. (2002) "An algorithm for assessing the risk of traffic accident." *Journal of Safety Research.*, **33**(3), 387-410.

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics, Hoboken, USA.

Ospina R., Ferrari S.L.P. (2012) "A general class of zero-or-one inflated beta regression models." *Computational Statistics and Data Analysis*, **56**(6)¸ 1609-1623.

Peck R.C., Gebers M.A., Voas R.B., Romano E. (2008) "The relationship between blood alcohol concentration (BAC), age and crash risk." *Journal of Safety Research,* **39**, 311-319.

Rosenbloom T., Eldror E., Shahar A. (2009) "Approaches of truck drivers and non-truck drivers toward reckless on-road behaviour." *Accident Analysis and Prevention,* **41**(4), 723-728.

Schonlau M., Scribner R., Farley T.A., Theall K., Bluthenthal R.N., Scott M., Cohen D.A. (2008) "Alcohol outlet density and alcohol consumption in Los Angeles county and southern Louisiana." *Geospatial Health,* **3**(1), 91-101.

Scollnilk D.P.M. (2002). "Modeling size-of-loss distributions for exact data in WinBUGS." *Journal of Actuarial Practice,* **10**, 193-218.

Scribner R.A., MacKinnon D.P., Dwyer J.H. (1994) "Alcohol outlet density and motor vehicle crashes in Los Angeles country cities." *Journal of Studies on Alcohol,* **55**(4), 447-453.

Spiegelhalter D.J., Best N.G., Carlin B.P., Van Der Linde A. (2002) "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* **64**(4), 583-639.

Sigve O., Torbjørn R. (2007) "Using cluster analysis to test the cultural theory of risk perception." *Transportation Research Part F: Traffic Psychology and Behaviour*, **10**(3), 254-262.

Tay R. (2005) "Drink driving enforcement and publicity campaigns: are the policy recommendations sensitive to model specification?" *Accident Analysis and Prevention*, **37**(2), 259-266.

Taylor B., Irving H.M., Kanteres F., Room R., Borges G., Cherpitel C., Greenfield T., Rehm J. (2010). "The more you drink, the harder you fall: a systematic review and meta-analysis of how acute alcohol consumption and injury or collision risk increase together." *Drug and Alcohol Dependence,* **110**(1-2), 108-116.

Transport Department (2012) *Road Traffic Accident Statistics, Year 2007 to Year 2011*, Transport Department of Hong Kong. Available at <http://www.td.gov.hk/en/road_safety/road_traffic_accident_statistics/index.html> (Accessed on 29 July 2012).

Transport Department (2013) *Transport in Hong Kong – Public Transport.* Transport Department of Hong Kong. Available at http://www.td.gov.hk/en/transport_in_hong_kong/public_transport/introduction/index.html (Accessed on 21 May 2013).

Treno A.J., Johnson F.W., Remer L.G., Gruenewald P.J. (2007) "The impact of outlet densities on alcohol-related crashes: a spatial panel approach." *Accident Analysis and Prevention,* **39**(5), 894-901.

Tsui K.L., Sze N.N., So F.L., Wong S.C. (2010) "Association between drink driving and severity of crash injuries to road users." *Hong Kong Journal of Emergency Medicine*, **17**, 34-39.

Vanlaar W. (2005) "Drink driving in Belgium: results from the third and improved roadside survey." *Accident Analysis and Prevention,* **37**(3), 391-397.

Wong J.T., Chung Y.S. (2008). "Analyzing heterogeneous accident data from the perspective of accident occurrence." *Accident Analysis and Prevention,* **40**(1), 357-367.

Table 1 Notation of the BrAC Frequency Distribution Profiles

Spatial-temporal unit	BrAC Level (µg/100ml)					Total
	0	0<x<=5	5<x<=10	…	x>95	
1	n_{11}	n_{12}	n_{13}	…	$n_{1\ 21}$	n_1
2	n_{21}	n_{22}	n_{23}	…	$n_{2\ 21}$	n_2
…						
k	n_{k1}	n_{k2}	n_{k3}	…	$n_{k\ 21}$	n_k
k+ 1	$n_{k+1\ 1}$	$n_{k+1\ 2}$	$n_{k+1\ 3}$	…	$n_{k+1\ 21}$	n_{k+1}
…						
K	n_{K1}	n_{K2}	n_{K3}	…	$n_{K\ 21}$	n_K

Table 2 Summary the 745 Drink-driving Distribution

Factor	Attribute	Frequency	Percent (%)
District	**Bar & Restaurant areas**	**189**	**25.4**
	- Yau Tsim Mong	125	16.8
	- Wan Chai	64	8.6
	Urban areas	**220**	**29.5**
	- Kowloon City	38	5.1
	- Kwun Tong	87	11.7
	- Sham Shui Po	13	1.7
	- Wong Tai Sin	6	0.8
	- Central and Western	21	2.8
	- Eastern	29	3.9
	- Southern	26	3.5
	Rural areas	**336**	**45.1**
	- Islands	7	1.0
	- Kwai Tsing	65	8.7
	- North	13	1.7
	- Sai Kung	60	8.1
	- Shatin	22	3.0
	- Tai Po	46	6.2
	- Tsuen Wan	48	6.4
	- Tuen Mun	10	1.3
	- Yuen Lonh	65	8.7
Day of week	**Weekdays & Sunday**	**493**	**66.2**
	- Monday	106	14.2
	- Tuesday	87	11.7
	- Wednesday	70	9.4
	- Thursday	66	8.9
	- Sunday	164	22.0
	Weekends	**252**	**33.8**
	- Friday	107	14.4
	- Saturday	145	19.4
Time period	**Daytime & Nighttime**	**523**	**70.2**
	- 7:00 a.m.–11:00 a.m.	111	14.9
	- 11:00 a.m.–3:00 p.m.	118	15.8
	- 3:00 p.m.–7:00 p.m.	97	13.1
	- 7:00 p.m.–11:00 p.m.	197	26.4
	Overnight	**222**	**29.8**
	- 11:00 p.m.–3:00 a.m.	152	20.4
	- 3:00 a.m.–7:00 a.m.	70	9.4

Number of observations = 745.

Table 3 Chi-square test of independence between different clusters

	Chi-square test statistic, χ^2					
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	-	-	-	-	-	-
Cluster 2	89.742**	-	-	-	-	-
Cluster 3	85.724**	42.266**	-	-	-	-
Cluster 4	90.226**	33.217*	39.468**	-	-	-
Cluster 5	86.708**	311.593**	541.004**	203.902**	-	-
Cluster 6	83.452**	219.414**	403.168**	184.399**	88.344**	-

* Statistically significant at 5% level
** Statistically significant at 1% level

Table 4 Summary of the 33,472 RBT Measurements

Factor	Attribute	Frequency	Percent (%)
District	**Bar & restaurant area**	**9,235**	**27.6**
	- Yau Tsim Mong	6,521	19.5
	- Wan Chai	2,714	8.1
	Urban area	**10,434**	**31.2**
	- Kowloon City	1,254	3.7
	- Kwun Tong	5,114	15.3
	- Sham Shui Po	393	1.2
	- Wong Tai Sin	129	0.4
	- Central and Western	769	2.3
	- Eastern	1,807	5.4
	- Southern	968	2.9
	Rural area	**13,803**	**41.2**
	- Islands	146	0.4
	- Kwai Tsing	1,827	5.4
	- North	559	1.7
	- Sai Kung	2,359	7.0
	- Shatin	530	1.6
	- Tai Po	2,467	7.4
	- Tsuen Wan	2,805	8.4
	- Tuen Mun	434	1.3
	- Yuen Lonh	2,676	8.0
Day of week	**Weekdays & Sunday**	**21,848**	**65.3**
	- Monday	4,964	14.8
	- Tuesday	3,293	9.8
	- Wednesday	2,495	7.5
	- Thursday	2,437	7.3
	- Sunday	8,659	25.9
	Weekends	**11,627**	**34.7**
	- Friday	4,563	13.6
	- Saturday	7,064	21.1
Time period	**Daytime & Nighttime**	**22,206**	**66.3**
	- 7:00 a.m.–11:00 a.m.	5,901	17.6
	- 11:00 a.m.–3:00 p.m.	4,118	12.3

	- 3:00 p.m.–7:00 p.m.	4,033	12.0
	- 7:00 p.m.–11:00 p.m.	8,154	24.4
	Overnight	**11,266**	**33.7**
	- 11:00 p.m.–3:00 a.m.	8,534	25.5
	- 3:00 a.m.–7:00 a.m.	2,732	8.2
Month	**Spring**	**7,723**	**23.0**
	- March	2,320	6.9
	- April	2,410	7.2
	- May	2,993	8.9
	Summer	**6,892**	**20.6**
	- June	2,302	6.9
	- July	2,064	6.2
	- August	2,526	7.5
	Fall	**8,683**	**26.0**
	- September	2,510	7.5
	- October	3,272	9.8
	- November	2,901	8.7
	Winter	**10,174**	**30.4**
	- December	3,532	10.6
	- January	4,253	12.7
	- February	2,389	7.1
Road type	**Minor**	**12,278**	**36.7**
	Major	**21,194**	**63.3**
Gender	**Male**	**31,641**	**94.5**
	Female	**1,831**	**5.5**
Vehicle class	**Non-commercial vehicles**	**25,370**	**75.8**
	- Motor cycle	1,185	3.5
	- Private car	18,536	55.4
	- LGV	5,589	16.7
	- Government	23	0.1
	- Others	37	0.1
	Other commercial vehicles	**2,383**	**7.1**
	- MGV	1,043	3.1
	- HGV	63	0.2
	- Public light bus	777	2.3
	- Public bus	489	1.5
	- Tram	11	0.0
	Taxi	**5,719**	**17.1**

Number of observations = 33,472.

Table 5 DIC values of different zero-inflated regression models for different clusters

	Overall	Cluster 1 Bar & restaurant area, weekend-overnight	Cluster 2 Bar & restaurant area, other timespans	Cluster 3 Urban area, weekend-overnight	Cluster 4 Urban area, other timespans	Cluster 5 Rural area, weekend-overnight	Cluster 6 Rural area, other timespans
ZILogNorm	15,096.8	770.6	2,827.2	548.4	**2,742.4**	2,186.5	6040.9
ZIGamma	15,228.5	**767.8**	2,867.6	**548.2**	2,775.3	**2,186.5**	6099.7
ZIExpon	**15,026.2**	772.1	**2,817.5**	N.A.*	2,745.0	2,190.8	**6019.7**
Sample size	33,472	1,034	8,201	965	9,469	2,359	11,444

Notes: ZILogNorm – Zero-inflated Lognormal Regression Model
ZIGamma – Zero-inflated Gamma Regression Model
ZIExpon – Zero-inflated Exponential Regression Model

* The respective Bayesian model could not converge.

Table 6 Parameter estimates of zero-inflated regression models for different clusters

	Overall	Cluster 1 Bar & restaurant area, weekend-overnight	Cluster 2 Bar & restaurant area, other timespans	Cluster 3 Urban area, weekend-overnight	Cluster 4 Urban area, other timespans	Cluster 5 Rural area, weekend-overnight	Cluster 6 Rural area, other timespans
Model	ZIExpon	ZIGamma	ZIExpon	ZIGamma	ZILogNorm	ZIGamma	ZIExpon
Regression-state							
Constant	3.608**	1.858**	3.376	2.397**	0.669	2.599**	6.398**
Season							
- Spring	29.530**	1.910**	24.900**	1.212*	0.711	1.574**	29.990**
- Summer	13.670**	0.904*	20.690*	1.451*	1.031*	0.912**	9.369**
- Autumn	9.018**	0.526	6.505**	0.682	0.968**	0.819**	10.740**
- Winter	(Control)						
Road type							
- Major	-0.167	0.031	0.268	0.022	0.438	-0.175	0.201
- Minor	(Control)						
Gender							
- Male	1.425	0.336	2.171	-0.235	0.232	-0.337	-0.679
- Female	(Control)						
Vehicle class							
- Commercial	-1.758	-0.134	2.447	-0.443	-0.215	0.012	0.120
- Taxi	-2.796**	-1.861**	-3.025*	0.235	-0.877*	-0.143	2.252
- Non-commercial	(Control)						
Cluster							
- Cluster 1	-1.045**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 2	0.399**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 3	-0.570**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 4	0.252*	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 5	-0.069	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 6	(Control)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Zero-state							
Constant	3.565**	2.485**	3.571**	3.368**	3.670**	3.267**	3.654**
Gender							
- Male	-0.187	-0.242	-0.015	-0.458	0.238	-0.596	-0.415
- Female	(Control)						

Vehicle class							
- Commercial	1.091**	81.930	0.840*	1.700	0.989*	80.120	1.202**
- Taxi	1.558**	2.477**	1.028**	82.010	1.041**	81.070	2.082**
- Non-commercial	(Control)						
Cluster							
- Cluster 1	0.887	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 2	-0.268	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 3	2.410	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 4	0.870	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 5	1.533	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
- Cluster 6	(Control)	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Parameters							
- a	N.A.	1.886	N.A.	1.233	N.A.	1.843	N.A.
- b	N.A.	0.102	N.A.	0.062	N.A.	0.088	N.A.
Predicted mean BrAC (μg/100ml)#	0.097	18.460	0.132	19.810	1.849	20.850	0.088
Estimated proportions oftested drivers							
- BrAC = 0 μg/100ml	97.33%	95.00%	97.89%	96.62%	98.35%	93.94%	96.71%
- 0 <BrAC ≤5	1.02%	0.56%	1.02%	0.60%	0.73%	0.58%	1.17%
- 5 <BrAC ≤10	0.63%	0.96%	0.53%	0.60%	0.31%	1.02%	0.75%
- 10 <BrAC ≤15	0.39%	0.92%	0.27%	0.50%	0.17%	1.02%	0.48%
- 15 <BrAC ≤20	0.24%	0.75%	0.14%	0.40%	0.10%	0.87%	0.31%
- 20 <BrAC ≤25	0.15%	0.56%	0.07%	0.31%	0.07%	0.69%	0.20%
- 25 <BrAC ≤30	0.09%	0.40%	0.04%	0.24%	0.05%	0.53%	0.13%
- 30 <BrAC ≤35	0.06%	0.28%	0.02%	0.18%	0.04%	0.39%	0.08%
- 35 <BrAC ≤40	0.03%	0.19%	0.01%	0.14%	0.03%	0.28%	0.05%
- 40 <BrAC ≤45	0.02%	0.13%	0.01%	0.10%	0.02%	0.20%	0.03%
- 45 <BrAC ≤50	0.01%	0.09%	0.00%	0.08%	0.02%	0.14%	0.02%
- 50 <BrAC ≤55	0.01%	0.06%	0.00%	0.06%	0.02%	0.10%	0.01%
- 55 <BrAC ≤60	0.01%	0.04%	0.00%	0.04%	0.01%	0.07%	0.01%
- 60 <BrAC ≤65	0.00%	0.02%	0.00%	0.03%	0.01%	0.05%	0.01%
- 65 <BrAC ≤70	0.00%	0.02%	0.00%	0.02%	0.01%	0.03%	0.00%
- 70 <BrAC ≤75	0.00%	0.01%	0.00%	0.02%	0.01%	0.02%	0.00%
- 75 <BrAC ≤80	0.00%	0.01%	0.00%	0.01%	0.01%	0.02%	0.00%

- 80 <BrAC ≤85	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%
- 85 <BrAC ≤90	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%
- 90 <BrAC ≤95	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
- 95 <BrAC ≤100	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
RMSE	10.392	13.456	7.678	17.847	1.221	15.362	11.376
No. of observations	33,472	1,034	8,201	965	9,469	2,359	11,444

Notes: RMSE – Root mean square error

* $p < 0.05$. ** $p < 0.01$.

The predicted mean BrAC are estimated based on the regression equation of corresponding zero-inflated model.

Figure 1 Geographical Regions in Hong Kong

The New Territories
1. Islands
2. Kwai Tsing
3. North
4. Sai Kung
5. Shatin
6. Tai Po
7. Tsuen Wan
8. Tuen Mun
9. Yuen Long

Kowloon
10. Kowloon City
11. Kwun Tong
12. Sham Shui Po
13. Wong Tai Sin
14. Yau Tsim Mong

Hong Kong Island
15. Central and Western
16. Eastern
17. Southern
18. Wan Chai

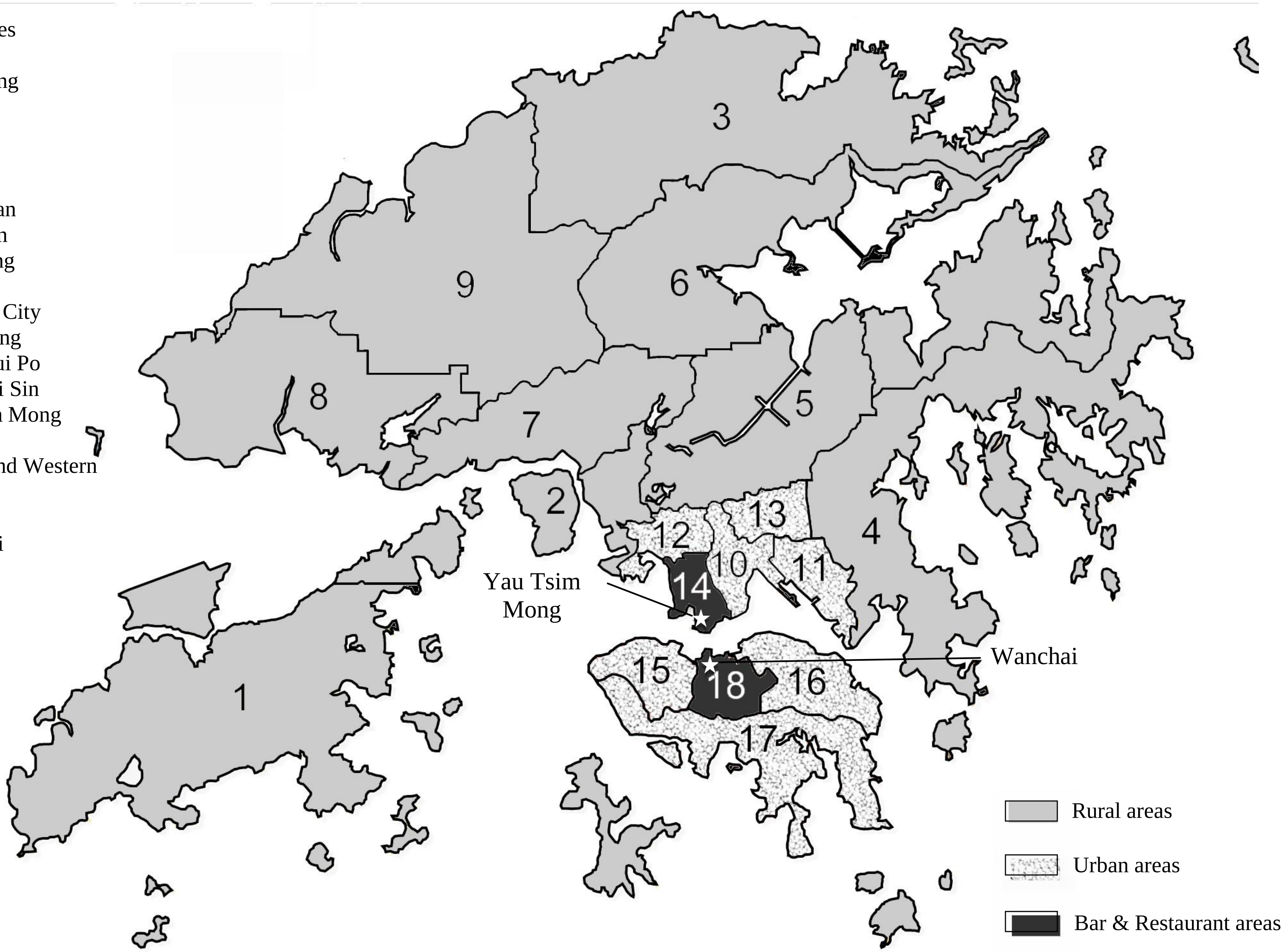

Figure 2 Tree obtained from cluster analysis after pruning