Accident Analysis and Prevention 119 (2018) 16–22

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Estimating factors of individual and regional characteristics affecting the drink driving recidivism

Tzu-Ying Chen, Rong-Chang Jou*

Department of Civil Engineering, National Chi Nan University, No. 1, University Rd., Puli, Nantou County 54561, Taiwan

ARTICLE INFO

Keywords:
Traffic violations
Drink driving
Recidivist
Multilevel random effect logistic regression model

ABSTRACT

Traffic violations, particularly drink driving, are a menace to the drivers themselves, and to other road users. Drink driving crashes often cause death or serious injury to the driver. Understanding the recidivism effect factor of drink driving is essential for designing effective countermeasures. This study is based on register-based data from the National Police Agency, Ministry of the Interior of Taiwan and monthly administrative area information from 2012 to 2015 for the entire population. Hence, this study not only focuses on the effect factor and violation differences between recidivists and non-recidivists, but discusses the entire regional characteristics effect for recidivism. The purpose of this study is to offer a comprehensive econometrical framework, using a multilevel random effect logistic model, which highlights important contributors to drink driving recidivism from regional attributes. As the study findings from our empirical results indicate, there are statistically significant differences with drink driving in administrative areas, depending on the number of report on drink driving by police, divorce rate of the population, alcohol consumption, number of community security patrol teams, number of bus trips, and level of education. The results of this study provide suggestions to the government for enhancing community security and developing public transportation, both of which can effectively decrease drink driving recidivism and improve public road safety.

1. Introduction

Yearly statistics done by the National Police Agency, the Ministry of the Interior of Taiwan (2017) indicated that there are more than 120,000 drink driving violation incidents per year and an average of one person dies per day due to drink driving. That is, up to one sixth of deaths are due to drink driving traffic crashes in Taiwan. Authorities in Taiwan amended the Road Traffic Security Rules in 2013, the pecuniary punishment of drunk drivers was increased from NT$[1]15,000-NT$60,000 (US$500-US$2000) to now NT$15,000-NT$90,000 (US$500-US$3000). The fine of NT$90,000, the maximum pecuniary punishment, is applied to offenders who violate the Road Traffic Security Rules and recidivist within 5 years. In the same year, Article 185 of the Criminal Law was re-amended. Drivers whose blood alcohol content (BAC) exceeding 0.55 mg/L in the breathalyser test would face imprisonment up to 2 years, criminal detention, or a fine up to NT$200,000 (US$6666). Those drunk drivers who caused injuries have to face 1–7 years of incarceration, but for those who caused deaths would face a longer incarceration (3–10 years). In addition, the BAC limits were lowered to 0.15 mg/L (for those drivers without license or having a license less than two years) and 0.25 mg/L (for other drivers). In recent year, the police in Taiwan try to avoid drink driving incidents by increasing routine stop checkpoints, especially in the nightclub and drinking establishment areas of the key region.

According to the Kaplan and Prato (2007) study, this is because of higher compliance with the law among females. The vast majority of recidivists (93.9%) were male drivers, and the result was the same in Møller et al. (2015). The Roach (2007) study shows that drink driving related offences may be indicative of involvement in additional and more serious offences. Also, previous studies indicated that the drink driving recidivism rate is 2% per year and they show an increased risk of involvement in road crashes (Bean et al., 2014 and Brewer et al., 1994). The Fell (1995) report, from the United States, found that about one third of the drivers charged with exceeding the BAC legal limit were recidivists. In recent years, Armstrong et al., (2014) and C'de Baca et al., 2001 indicated that the likelihood of drink driving recidivism increases with the increase of high blood alcohol concentration (BAC) at the time of detection. As the Ferrante et al. (2001) study pointed out, numerous criminological studies have shown that an early age of onset foreshadows a longer and more serious criminal career (Loeber and le

* Corresponding author.
E-mail address: rcjou@ncnu.edu.tw (R.-C. Jou).

[1] US$1 = NT$30.

https://doi.org/10.1016/j.aap.2018.06.011
Received 2 August 2017; Received in revised form 2 April 2018; Accepted 15 June 2018

Blanc, 1990; Farrington, 1992).

Similarly, drink driving violations of recidivists may be one of the early predictors of 'high risk' drinking drivers. Some factors related to recidivist, including gender (Armstrong et al., 2014; C'de Baca et al., 2001; Lapham et al., 2000; Møller et al., 2015), age (Møller et al., 2015), education level (C'de Baca et al., 2001; Møller et al., 2015), BAC (Bailey, 1993; Brewer et al., 1994), previous drink driving incidents registered by the police (Brewer et al., 1994; Hedlund and Fell, 1995; Ferrante et al., 2001), and others. Brewer et al. (1994) studied the association between being detected for drink driving and dying in an alcohol-related crash. The results indicated that drivers with a high BAC were more likely to be fatally injured than a driver with a low BAC. More specifically, with a rising BAC, the risk of a fatality or serious injury increases significantly when driving under the influence of alcohol. The Hels et al. (2013) study results imply that when driving under the influence with a BAC of 0.12% or above, the risk of being seriously injured is 78 times higher, compared to driving while not under the influence of alcohol. The prevalence of fatalities involving a BAC over the legal limit was high in the countries of Norway, Finland, Sweden, and Portugal. They also determined that high BACs was a key challenge. (Legrand et al., 2014).

Above, it is important to design effective countermeasures as an understanding of drinking driving recidivist. Such as the Møller et al., (2015) study indicated that a five year prevalence of 17% for drinking driving recidivists, aged 18 or older, of Denmark. However, Most of previous studies of recidivist have only focused on specific sample of drinking driving offenders (e.g., Cavaiola et al., 2007; Ouimet et al., 2007).

The purpose of this study is to not only identify important factors influencing on drink driving recidivism, but also differentiate the effects of spatial characteristics on drink driving recidivism. From a methodological viewpoint, using the traditional logistic regression models can disaggregate the macro social (regional characteristics) and economic data (violation characteristics) into a single-level model. However, this will violate the independence and homogeneity hypothesis because an assumption of logistic regression is that the residuals from the model are independent across subjects (Jones and Jørgensen, 2003). However, a multilevel model relaxes the above assumption and allows the effects of these variables to vary across groups. Using a multilevel logistic regression can account for lack of independence across levels of nested data (i.e. individuals nested within groups), and is appropriate when there is correlation among clusters of subjects. For this reason, in this study, we first construct general modelling (non-hierarchical) methods and determine if the data are appropriately observed. Secondly, we consider using a hierarchical structure that is a statistical description of the data characterised by correlated attributes within hierarchical clusters, which is appropriately justified by the presence of correlation within the clusters.

Considering the sample integrity, this study presents a design by focusing on the data consists of 14,151 drink-driving violation recidivists occurring in 12 administrative areas of Taipei city in Taiwan, while examining a 5-year period between 1/1/2012 and 31/12/2015 to measure the drinking recidivist. Moreover, considering that the regional characteristics may vary in different areas over time due to their demographic and socioeconomic infrastructure (such as annual household disposable income, number of drinking establishment, police enforcement, and public transportation network), this study is the first to look at the differences between the individual level characteristics and area level characteristics, not only between recidivists and non-recidivists, but also with respect to the entire population of area. The violation data hierarchy is postulated as follows: Drink-driving recidivists represent the lowest level of the hierarchy, while the change in administrative areas over time represents the higher-level hierarchy, or cluster. Hence, it is reasonable to claim that correlation exists among violation behaviour occurring in the same administrative area, because these violation behaviours may share unobserved and/or unrecorded characteristics of the administrative area.

2. Model methodology

2.1. Logistic regression model

A traditional logistic model is generally used to analyse the binary response function, as shown below:

$$\text{logit}(P_i) = \log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

where β_0 is the intercept, and X_i represents a vector of independent variables for individual i, with a vector of β_1 as the corresponding coefficients. When $P_i = 1$, represents the recidivists. The possibility of choice is as below:

$$P_i = prob[u_i > -(\beta_0 + \beta_1 X_i)] = 1 - F[-(\beta_0 + \beta_1 X_i)] \tag{2}$$

The maximum likelihood method is employed to measure the associations by constructing the likelihood function as follows ($y_i = 1$, if individual i is a recidivist; $y_i = 0$, otherwise):

$$L = \prod_{y_i=1} P_i \prod_{y_i=0} (1 - P_i) \tag{3}$$

2.2. Multilevel random effects logistic models

In the multilevel model, there have two-level structure data, three different equations can be formulated: individual-level model (level 1 model), administrative area-level model (level 2 model), and combined model. Assuming normally distributed errors, for subject ij of the study (ith individual at level 1 and jth area at level 2), we have a level 1 model as:

Level 1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij},\ Y_{ij} \sim N(\beta_{0j} + \beta_{1j} X_{ij}, \sigma^2);\ \varepsilon_{ij} \sim (0, \sigma^2) \tag{4}$$

where β_{0j} is the intercept, and β_{1j} are the regression coefficients associated with the predictors X_{ij}. β_{1j} are assumed to vary across districts depending on their effects at the administrative area level. ε_{ij} is the error term accounting for random effects of levels 1 and 2. The formulation is similar to a traditional regression model, however, there is an important difference in that both intercept and regression coefficients have subscript j, indicating that the intercept β_{0j} and slope coefficients β_{1j} are permitted to vary across the level 2 administrative area.

According to Yannis et al., (2008) and Kreft and Leeuw (1998), corresponding with a defined random intercept and random slope model, the level 2 model shown below has a subscript because they are assumed to vary across administrative area.

Level 2 model: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_{0j}$

$$\beta_{1j} = \gamma_{10} + \mu_{1j} \tag{5}$$

To consider the binominal variables as continuous variables, a logit transformation is required. In this study, the dependent variable P_{ij} denotes the possibility of drink-driving recidivism and $P_{ij} = (0,1)$. To be more precise, to predict the possibility of recidivism, considering a binomial $Y_{ij} = (0,1)$ outcome, and $P_{ij} = [\exp(Y_{ij})/(1 + \exp(Y_{ij}))]$, substitute this with formula (4) and (5), as shown below:

Multilevel logidtic model:

$$Logit(\theta) = \log\left(\frac{P(Y_{ij})}{1 - P(Y_{ij})}\right) = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}$$

$$= \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_{1j} X_{ij} + \mu_{0j} + \varepsilon_{ij} \tag{6}$$

Y_{ij} is the logit prediction for the ith subject at level 1 and jth unit at level 2. γ_{00} is the intercept denoting the grand mean, W_j, the regional (administrative area) level characteristic, X_{ij}, the individual (recidivists) level characteristic, and γ_{10} are the regression coefficients associated with regional level characteristic and individual level characteristic. μ_{0j}