The relation between working conditions, aberrant driving behaviour and crash propensity among taxi drivers in China

Yonggang Wang[a], Linchao Li[b], Carlo G. Prato[c,*]

[a] School of Highway, Chang'an University, Middle Section of South 2 Ring Rd., P.O. Box 487, Xi'an, 710064, Shaanxi, China
[b] School of Transportation, Southeast University, 2 Sipailou, Nanjing, 210018, Jiangsu, China
[c] School of Civil Engineering, The University of Queensland, St. Lucia 4072, Brisbane, Australia

ABSTRACT

Although the taxi industry is playing an important role in Chinese everyday life, little attention has been posed towards occupational health issues concerning the taxi drivers' working conditions, driving behaviour and road safety. A cross-sectional survey was administered to 1021 taxi drivers from 21 companies in four Chinese cities and collected information about (i) sociodemographic characteristics, (ii) working conditions, (iii) frequency of daily aberrant driving behaviour, and (iv) involvement in property-damage-only (PDO) and personal injury (PI) crashes over the past two years. A hybrid bivariate model of crash involvement was specified: (i) the hybrid part concerned a latent variable model capturing unobserved traits of the taxi drivers; (ii) the bivariate part modelled jointly both types of crashes while capturing unobserved correlation between error terms. The survey answers paint a gloomy picture in terms of workload, as taxi drivers reported averages of 9.4 working hours per day and 6.7 working days per week that amount on average to about 63.0 working hours per week. Moreover, the estimates of the hybrid bivariate model reveal that increasing levels of fatigue, reckless behaviour and aggressive behaviour are positively related to a higher propensity of crash involvement. Lastly, the heavy workload is also positively correlated with the higher propensity of crashing, not only directly as a predictor of crash involvement, but also indirectly as a covariate of fatigue and aberrant driving behaviour. The findings from this study provide insights into potential strategies for preventive education and taxi industry management to improve the working conditions and hence reduce fatigue and road risk for the taxi drivers.

1. Introduction

Taxi has become recently an important share of the transport market in China because of its comfort, convenience and flexibility. The growth of the taxi market shares has resulted in profitable business, especially in large cities. An example is Xi'an in the Shanxi province, where in 2014 there were 49 licensed taxi companies ranging from less than 100 to over 1000 vehicles. The total number of taxi vehicles was 12,135 and provided about 476 million person-rides that corresponded to a market share of 19% of the total passenger volume, second only to the public bus service covering the largest market share at 62%. Similar examples of the growth of the taxi industry have also been reported for other cities in China such as Nanjing (Routley et al., 2009), Wuhan (Ma et al., 2010), and Beijing (Shi et al., 2014). Although the taxi industry is playing an important role in people's everyday life in China, the literature has paid little attention to occupational health issues concerning the taxi drivers' working conditions, driving behaviour and safety.

Taxi drivers are often working in stressful and hazardous conditions including long working hours, frequent driving tasks and occasional disputes with passengers. These conditions translate into an increase in their physical and mental stress. A survey in Australia revealed that 67% of taxi drivers worked up to 12 hours per day and at least 50 hours per week (Dalziel and Job, 1997). Another survey in China reported that taxi drivers in Beijing spent an average of 11 hours per day driving and worked 27.8 days per month. Moreover, taxi drivers must pay a licensing management fee (commonly known as Fenzi Fee) to the taxi company that can be as much as half of their daily gross income (Shi et al., 2014). Another survey in Taiwan showed that taxi drivers worked approximately 10 hours per day and took only 2.7 rest days per month (Tseng, 2013). Clearly, the daily earnings of taxi drivers correlate heavily with their driving distance within working hours. Consequently, they have to spend long hours driving each day and carry as many passengers as possible to earn more money. Not surprisingly, taxi drivers engage frequently in occupational risky activities while driving.

The literature shows a handful of studies that examined the main

---

effects of exposure to hazardous working conditions on the health of taxi drivers and identified potential risk factors contributing to the occurrence of road crashes. Dalziel and Job (1997) added a different dimension as they indicated that busy or overworked taxi drivers in Sydney (Australia) were more likely to be involved in personal injury (PI) collisions due to physical and mental fatigue. Maag et al. (1997) showed that road crashes involving taxis in the Quebec province were related to the age of the drivers, as taxi drivers less than 30 years old were more likely to be involved in those crashes. La et al. (2013) suggested that the number and severity of road crashes involving taxis are associated with age and license status of the taxi drivers, but also with type of employment (full or part-time), satisfaction with the current income, and seat-belt usage. Zhao et al. (2015) reported that taxi drivers between 26–40 years old and with 3–5 years of job experience were at greater risk of being involved in crashes, especially in property-damage-only (PDO) collisions. Lim and Chia (2015) investigated the health dimension and revealed that, when compared with the general population, taxi drivers smoke more, exercise less, and sleep in the vehicle during breaks. Moreover, they have a higher body mass index (BMI) and more unhealthy eating habits due to the extended hours of driving, thus confirming previous findings about musculoskeletal problems being commonly observed among them (Raanaas and Anderson, 2008).

As taxi drivers work regularly long hours into the late night or early morning, they can be vulnerable to fatigue or sleepiness and hence increased risk while driving (Dalziel and Job, 1997; Meng et al., 2015; Zhao et al., 2015) even though the mechanism underlying sleepiness while driving is still not fully understood (Wang et al., 2017). Moreover, taxi drivers are vulnerable to other occupational exposures that affect the risk of being involved in road crashes, including vision defects (Maag et al., 1997), mental stress (Mann et al., 2010), irregular shifts (Wang et al., 2015), lower income (Lim and Chia, 2015), and traffic congestion conditions (Vayalamkuzhi and Amirthalingam, 2016).

Fatigue is not the only factor that might contribute to the risk of being involved in road crashes, as the literature presents a handful of studies that examined aberrant behaviour of taxi drivers. In fact, as their earnings depend heavily on the number of rides within their scheduled shift, taxi drivers tend to drive faster to save time and transport more passengers, and hence they commit violations to the traffic rules. Taxi drivers in Beijing and Nanjing (China) did not wear seat belts often, although they were not more likely to commit this violation than other drivers (Fleiter et al., 2009; Routley et al., 2009). In Wuhan (China), a survey among taxi drivers revealed that their attitudes towards traffic law violations affected their risky driving behaviour (Ma et al., 2010). A study of road crashes involving taxis in Taiwan showed that 26% of the involved taxi drivers committed at least one speeding violation in the previous year, and the frequency of speeding behaviour was significantly related to their age, years of experience, daily driven kilometres, and number of days off per month (Tseng, 2013). Sullman et al. (2013) reported that taxi drivers exhibiting a passive-aggressive driving behaviour got easily distracted and had an increased risk of crashes. Wang et al. (2014) showed that male taxi drivers in Xining (China) were often displaying aberrant driving behaviour (e.g., unsafe passing, crossing red lights, speeding, following too closely, unsafe lane changes, etc.), and they were frequently involved in road crashes.

Given the literature about the hazardous working conditions, the aberrant driving behaviour and the risk of road crash involvement of taxi drivers, this study presents a comprehensive analysis of the relations between workload of taxi drivers, consequent fatigue and stress, related aberrant driving behaviour, and crash occurrence. A cross-sectional survey captured self-reported information about these different aspects and 1021 taxi drivers in four major Chinese cities participated in this study. Moreover, a hybrid bivariate model of the probability of being involved in both PDO and PI crashes was specified to understand the relation of the aforementioned factors aspects to the occurrence of the crashes. This study aims at answering questions about (i) the current workload of Chinese taxi drivers, (ii) their engagement in aberrant driving behaviour, (iii) their feeling of work related fatigue, and (iv) the correlation between these aspects and road crash involvement. Most relevantly, this study proposes a hybrid model able to look into the association of the occurrence of PDO and PI crashes while considering both observed characteristics (e.g., sociodemographic characteristics, workload conditions) as well as unobserved traits of taxi drivers (e.g., level of fatigue, measure of aberrant behaviour, level of aggressive behaviour).

The remainder of the paper is structured as follows. Section 2 describes the survey design and administration, as well as the model that jointly represents the involvement in PDO and PI crashes as a function of both observed and unobserved traits of the taxi drivers. Section 3 presents then the sample characteristics and the model estimates. Section 4 provides a discussion of the results and draws the conclusions of this study.

2. Methods

2.1. Survey design and administration

Given the literature review about driving behaviour and crash involvement of taxi drivers, the survey design intended to answer the following research questions.

The first question investigated several aspects related to the working conditions of the taxi drivers: (i) the workload, given extremely heavy values reported by taxi drivers in Taiwan and Beijing (Tseng, 2013; Shi et al., 2014); (ii) the frequency of aberrant behaviour, given reports of taxi drivers committing frequent traffic violations (Fleiter et al., 2009; Routley et al., 2009; Ma et al., 2010; Tseng, 2013) and engaging in aggressive behaviour (Sullman et al., 2013; Wang et al., 2014); (iii) the frequency of fatigue perception, given reports of vulnerability to fatigue or sleepiness (Dalziel and Job, 1997; Mann et al., 2010; Meng et al., 2015; Zhao et al., 2015).

The second question explored the intuitive correlation between heavy workload conditions and fatigue, given that working long hours is considered the main determinant of drowsy driving (see, e.g., Valent et al., 2010; Higgins et al., 2017), as well as the possible correlation between heavy workload and aberrant behaviour, given that the latter consists of both violations and errors (see, e.g., Reason et al., 1990; Martinussen et al., 2013). Violations could on the one hand relate to taxi drivers attempting to transport more passengers in a shorter time, while errors could on the other hand relate to taxi drivers being overworked and consequently less attentive.

The third question looked at the relation between working conditions, fatigue effects, aberrant behaviour and road crash involvement, given existing literature mentioning the higher crash involvement of taxi drivers (Dalziel and Job, 1997; Maag et al., 1997; Tseng, 2013; La et al., 2013; Wang et al., 2014; Zhao et al., 2015). In particular, a multivariate analysis allowed controlling for the effect of each factor and examining whether the working conditions relate to crash involvement directly or indirectly.

Accordingly, a cross-sectional survey was designed by collecting information within four sections: (i) demographic characteristics (3 questions), (ii) working conditions (6 questions), (iii) frequency of daily aberrant driving behaviors (7 questions), and (iv) involvement in PDO and PI crashes over the past two years. Table 1 presents the details of the survey as well as the percentages of responses to every option for each question.

The first part collected information about age, gender and education of the taxi drivers. The second part asked them to report their average working conditions over the past two years, including the number of working hours per day, the number of days off per week, the type of work schedule, and the management fee due to the taxi company. Two additional questions gathered additional information that, according to

Table 1
Sample characteristics and survey answers.

Questions	Answers	%	Answers	%
Demographic characteristics				
What is your gender?	Male	91.8	Female	8.2
What is your age?	18–25–years old	9.0	41–50 years old	24.5
	26–30 years old	19.2	> 50 years old	6.5
	31–40 years old	40.8		
What is your education?	Primary school or below	19.5	Senior high school	34.1
	Middle school	33.3	Higher education	13.1
Working conditions				
How any hours per day do you work?	7 or less	6.6	10	29.3
	8	16.4	11	15.9
	9	28.8	12 or more	3.0
How many days off per week do you enjoy?	0	71.0	2	2.1
	1	26.9		
How is your schedule?	Fixed, decided by the company	80.4	Flexible, decided by me	8.4
	Flexible, decided by the company	11.2		
How much do you pay for the company's daily management fee?	0-150 CNY	14.8	181-210 CNY	20.8
	151-180 CNY	64.2	> 210 CNY	0.2
Are you dissatisfied with your current income?	Yes, always	58.9	No, rarely	7.2
	Yes, often	18.9	No, never	2.2
	Yes, sometimes	12.8		
How often do you have sleep problems (e.g., difficulty in falling asleep, difficulty in staying asleep, early morning awakening)	Never	14.0	Frequently	13.5
	Hardly ever	28.2	Nearly all the time	2.6
	Occasionally	23.1	Always	0.0
	Quite often	18.6		
Aberrant driving behaviour				
How often do you cross a junction with a red light?	Never	13.0	Frequently	14.3
	Hardly ever	17.9	Nearly all the time	5.3
	Occasionally	28.8	Always	0.0
	Quite often	20.7		
How often do you disregard the speed limits?	Never	9.9	Frequently	17.6
	Hardly ever	13.2	Nearly all the time	6.2
	Occasionally	24.6	Always	0.0
	Quite often	28.5		
How often do you accelerate to cross an intersection with a yellow light?	Never	9.3	Frequently	19.6
	Hardly ever	14.0	Nearly all the time	4.5
	Occasionally	27.2	Always	0.0
	Quite often	25.4		
How often do you driver aggressively (e.g., sudden acceleration, sudden braking, unsafe tailgating)?	Never	4.9	Frequently	17.0
	Hardly ever	10.6	Nearly all the time	2.4
	Occasionally	29.0	Always	0.0
	Quite often	36.1		
How often do you feel fatigued while driving?	Never	3.2	Frequently	26.3
	Hardly ever	5.8	Nearly all the time	19.6
	Occasionally	16.6	Always	0.3
	Quite often	28.2		
How often do you use your mobile phone while driving?	Never	35.5	Frequently	2.0
	Hardly ever	33.3	Nearly all the time	0.0
	Occasionally	22.8	Always	0.0
	Quite often	6.4		
How often do you overtake a vehicle without using the left/right turn light?	Never	14.1	Frequently	12.5
	Hardly ever	11.8	Nearly all the time	1.3
	Occasionally	31.5	Always	0.0
	Quite often	28.9		
Crash involvement				
How many property-damage-only crashes have you been involved in as a taxi driver over the past two years?	0	1.1	5	17.6
	1	13.0	6	4.3
	2	19.6	7	1.3
	3	22.8	8	0.1
	4	20.2		
How many personal injury crashes have you been involved in as a taxi driver over the past two years?	0	89.0	2	0.7
	1	10.3		

the literature, relates to the workload and the consequent stress: the dissatisfaction with the current income and the frequency of sleep problems (e.g., difficulty in falling asleep, difficulty in staying asleep, early morning awakening).

The third part requested the taxi drivers to assess how often they engaged in daily aberrant driving behaviour (e.g., speeding, crossing a red traffic light, accelerating at a yellow traffic light, using the mobile phone while driving, unsafe overtaking). Drivers reported the frequency of their behaviour on a 7-point Likert scale ranging from 0 (never) to 6 (always). The fourth part collected self-reports of involvement in road crashes as a taxi driver while differentiating between PDO and PI crashes.

The survey was administered between 2014 and 2015 in four cities in China (as illustrated in Fig. 1). The criteria for the selection of survey participants were that the taxi drivers had a minimum of 2-year experience after obtaining their license and a minimum of 50,000 km per year driven in the last two years. Initially, 21 taxi companies were contacted (8 in Xi'an, 5 in Xining, 4 in Jilin and 4 in Shantou) and 1138

Fig. 1. Data survey from four cities in mainland China.

taxi drivers (417 in Xi'an, 274 in Xining, 236 in Jilin and 211 in Shantou) received the questionnaire. After removing 117 incomplete or missing respondents, 1021 complete questionnaires were analysed (374 in Xi'an, 255 in Xining, 209 in Jilin and 183 in Shantou) for a response rate close to 90%.

2.2. Data analysis and hybrid bivariate model of crash involvement

The complete answers to the four parts of the survey allowed to analyse the self-reports of taxi drivers about their working conditions and aberrant driving behaviour. Moreover, the self-reports of crash involvement allowed to specify a bivariate model for PDO and PI crashes in relation to taxi drivers' sociodemographic characteristics, working conditions and aberrant driving behaviour. An examination of the number of PDO and PI crashes revealed that the former varied between 0 and 8 while the latter had most taxi drivers not experiencing crashes resulting in personal injuries, and only a handful being involved in more than one crash.

The selection of the model specification followed extensive testing to have the best explanatory power. Given the testing, the PDO crashes were modelled as an ordered dependent variable that was divided in four categories on the basis of their distribution, while the PI crashes were modelled as a binary dependent variable that captured whether or not the taxi driver was involved in them. The two models were specified at the person-level models (i.e., the unit of analysis was the taxi driver) and possibly shared common unobserved factors that affected the involvement in both types of crashes. Accordingly, a bivariate model was specified for calculating the probabilities of being involved in PDO and PI crashes as well as accommodating correlation that may exist between the error terms. The error terms were assumed to follow a bivariate normal distribution and the parameters of the models were estimated by using maximum likelihood estimation. It should be noted that the multivariate probit modelling methodology has been previously used in

the travel behaviour context (e.g., Dias et al., 2017; Emmerink et al., 1996; Ferdous et al., 2010).

Moreover, the bivariate model considered that both observed and unobserved variables entered the specification of the ordered probit for the PDO crashes and the binary probit for the PI crashes. Observed variables were the sociodemographic characteristics and the working conditions of the taxi drivers, while unobserved variables were the dimensions that were captured by the Likert scale indicators in the survey. Accordingly, a hybrid bivariate model was estimated to include a latent variable model and the original bivariate model of crash involvement models. Notably, the latent variable model consisted of measurement equations that related the collected indicators to unobserved traits of the taxi drivers and structural equations that associated these traits with observed ones. This implies that the sociodemographic characteristics and the working conditions might have both a direct (via being predictors in the crash involvement models) and an indirect (via being predictors in the structural equations of the latent variable model) effect on the probability of crash involvement (Vij and Walker, 2016).

Exploratory factor analysis was performed initially to elicit the unobserved traits and then the measurement equations were expressed accordingly as follows:

$$I_n = \alpha X_n^* + \upsilon_n \tag{1}$$

where I_n is a vector of R indicators corresponding to the survey answers of taxi driver n, X_n^* is a vector of L unobserved traits of taxi driver n, α is a vector of R parameters to be estimated, and υ_n is a vector of normally distributed error terms with a mean vector of zeros and covariance matrix Ψ_υ.

The structural equations of the unobserved traits X_{ln}^* were expressed as follows:

$$X_n^* = \gamma X_n + \omega_n \tag{2}$$

where X_n^* is a vector of L unobserved traits of taxi driver n, X_n is a vector of K observed characteristics of taxi driver n, γ is a $(L \times K)$ matrix of parameters to be estimated, and ω_n is a vector of L normally distributed error terms with a mean vector of zeros and covariance matrix Ψ_ω.

Given the vector X_n of observed characteristics and the vector X_n^* of unobserved traits of taxi driver n, the model assumed that there exists a partitioning of underlying continuous latent variables that relates directly to the involvement of taxi driver n in PDO and PI crashes (Dias et al., 2017). Consider p and q as the indices for the discrete outcomes corresponding respectively to the crash involvement in PDO and PI crashes: p takes different values for the PDO crash involvement ordered probit, and q takes two values for the PI crash involvement binary probit. The model assumes the following form:

$$f_n^* = \beta_f^* X_n^* + \beta_f X_n + \varepsilon_n \text{ where } f_n = p \text{ if } \delta_{p-1} < f_n^* < \delta_p, \ \delta_0 = -\infty, \ \delta_p = +\infty$$

$$g_n^* = \beta_g^* X_n^* + \beta_g X_n + g_0 + \eta_n \text{ where } g_n = q \text{ and } q = 1 \text{ if } g_n^* > 0 \text{ or } q = 0 \text{ if } g_n^* \le 0 \tag{3}$$

The latent variables f_n^* and g_n^* indicate the propensity of a taxi driver n to be involved in respectively PDO and PI road crashes, and the larger their values, the higher the propensity. The variables f_n and g_n are the observed categories that are retrieved from the self-report of involvement in PDO and PI crashes. The vectors β_f^*, β_f, β_g^* and β_g consist of parameters to be estimated, and so are the thresholds δ_p that partition the latent variable f_n^* in P categories of PDO crashes as well as the constant g_0 for the binary model of PI crashes. Last, the vectors ε_n and η_n are the random error terms of the two models.

The marginal distributions of the error terms ε_n and η_n are assumed to be normally distributed, and both ε_n and η_n are assumed to be independently and identically distributed (i.i.d.) across taxi drivers n. For identification purposes, the normal distribution of the error terms is standardised (Dias et al., 2017) and a joint bivariate standard normal distribution is assumed for the error terms ε_n and η_n to capture their potential correlation because of unobserved factors related to the taxi drivers n. The parameters to be estimated include the elements of the vectors β_f^*, β_f, β_g^*, and β_g, the thresholds δ_p, and the constant g_0. The parameters are estimated by maximising the log-likelihood function and a binary indicator $I_n(p,q)$ is defined: $I_n(p,q)$ is equal to 1 if taxi driver n was involved in a number of PDO crashes in category p and a number of PI crashes in category q, and 0 otherwise.

The joint probability of taxi driver n being involved in PDO and PI crashes is expressed as:

$$Pr(f_n = p, g_n = q) = Pr(\delta_{p-1} < f_n^* < \delta_p \text{ and } g_n = q)$$
$$= Pr(d_{n,p-1} < \varepsilon_n < d_{n,p} \text{ and } b_{n,0} < \eta_n < b_{n,1})$$
$$= \Phi_2(d_{n,p}, b_{n,1}; \theta) - \Phi_2(d_{n,p-1}, b_{n,1}; \theta) - \Phi_2(d_{n,p}, b_{n,0}; \theta)$$
$$+ \Phi_2(d_{n,p-1}, b_{n,0}; \theta) \tag{4}$$

where $d_{n,p} = \delta_p - \beta_f^* X_n^* + \beta_f X_n$, $b_{n,q} = q - \beta_g^* X_n^* + \beta_g X_n$, Φ_2 is the bivariate cumulative normal distribution function, and θ is a correlation parameter to be estimated. As there exist unobserved traits X_n^* from the latent variable model, the joint probability requires integration over their distribution (Vij and Walker, 2016; Walker and Ben-Akiva, 2002):

$$Pr(f_n = p, g_n = q)$$
$$= \int_{X_n^*} \Phi_2 \bigg(f_n = p, g_n$$
$$= q | X_n, X_n^*; \beta_f^*, \beta_f, \beta_g^*, \beta_g, \delta_p, g_0, \theta \bigg) h_I (I_n | X_n, X_n^*; \alpha, \psi_v) h_{X_n^*}$$
$$(X_n^* | X_n; \gamma, \psi_\omega) dX_n^* \tag{5}$$

where $\Phi_2(f_n = p, g_n = q | X_n, X_n^*; \beta_f^*, \beta_f, \beta_g^*, \beta_g, \delta_p, g_0, \theta)$ is the bivariate probability in eq. (4), $h_I(I_n | X_n, X_n^*; \alpha, \Psi_v)$ is the distrution of the indicators, and $h_{X_{n*}}(X_n^* | X_n, X_n^*; \gamma, \Psi_\omega)$ is the distribution of the

latent variables. The log-likelihood function takes then the following form:

$$logL = \sum_{n=1}^{N} \sum_{p=1}^{P} \sum_{q=0}^{1} I_n(p, q) Pr(f_n = p, g_n = q) \tag{6}$$

The log-likelihood function was coded in PythonBiogeme 2.5 (Bierlaire, 2016) and the maximisation of the function was performed by simulation via 1000 draws from the Modified Latin Hypercube Sampling algorithm (Hess et al., 2006).

3. Results

3.1. Sample characteristics

The 1021 taxi drivers who completed the questionnaire were predominantly males and were on average 36 years old, a good representation of a male-dominated and quite young profession in China. The education level of the taxi drivers was quite heterogeneous, with almost one fifth of drivers having only primary school education, but about one sixth of drivers holding a university degree. Table 1 presents a summary of the sample characteristics and the answers to the survey questions.

The analysis of the answers paints a gloomy picture of the working conditions of taxi drivers. Heavy workload was reported across the board because of the very long working hours (mean = 9.4, st.dev. = 1.2) combined with the very little days off during the week (mean = 0.3, st.dev. = 0.5) that make for an average of about 63.0 working hours per week. Notably, almost half of the taxi drivers reported to work 10 hours or more per day and over 70% of them added that they do not take days off during the week. Not surprisingly, about one third of the taxi drivers mentioned to have often problems with their sleep. If the heavy workload were not enough, the daily management fee affected heavily the taxi drivers, as over 85% of them paid over 150 CHY per day to their company. Relatedly, almost 60% of the sample was very dissatisfied with the current income while less than 10% mentioned satisfaction in the same sentence with income.

The analysis of the answers shows a tendency towards daily aberrant driving behaviour by taxi drivers. When looking at the raw data on the Likert scale from 0 (never) to 6 (always), sleepiness and fatigue while driving was the most frequently reported trait (mean = 3.3, st.dev. = 1.3), followed by self-reported aggressiveness (mean = 2.6, st.dev. = 1.1). The taxi drivers provided information also about disregarding speed limits (mean = 2.5, st.dev. = 1.4), accelerating when seeing a yellow light (mean = 2.5, st.dev. = 1.3), crossing a junction with a red light (mean = 2.2, st.dev. = 1.4), and overtaking without notifying other drivers (mean = 2.2, st.dev. = 1.3). Interestingly, the taxi drivers mentioned to use on average their mobile phone more than "hardly ever" (mean = 1.1, st.dev. = 1.0).

Last, the involvement in crashes during the past two years was quite high. Taxi drivers reported on average to have been involved in 3.3 PDO crashes (st. dev. = 1.5) and 0.1 PI crashes (st. dev. = 0.3). Interestingly, almost all taxi drivers were involved in at least one collision that resulted only in property damage, but only slightly more than one tenth was involved in a crash where persons were injured and only a handful of drivers was involved in two injury crashes.

3.2. Estimates of the hybrid bivariate model

The estimation of the hybrid bivariate model required extensive specification testing in order to obtain the best model for explaining the correlated variation of the involvement in PDO and PI crashes in relation to sociodemographic characteristics, working conditions and aberrant driving behaviour of the taxi drivers.

Initially, an exploratory factor analysis investigated how many unobservable traits could be identified. The items revealed a good internal

reliability, as the Cronbach's Alpha was 0.829, and a good sampling adequacy towards factor analysis, as the Kaiser-Meyer-Olkin measure was 0.824 at the overall level and 0.741–0.903 at the item level. The exploratory analysis revealed the existence of three factors: (i) fatigue, (ii) reckless behaviour, and (iii) aggressive behaviour. Consequently, the hybrid bivariate model included nine measurement equations according to the uncovered relations between the items and the unobserved traits, and three structural equations exploring the association between the unobserved traits and the observed characteristics of the taxi drivers. Variables were retained in the structural equations if the related parameters were significant at the 0.05 level.

Then, several models were estimated with a different number of categories for the ordered probit model of PDO road crashes. When estimating the hybrid bivariate model, the best specification was obtained for four categories (i.e., 0–1, 2–3, 4–5, more than 5 PDO crashes). Last, observed and unobserved traits entered both the ordered probit and the binary probit models and variables were retained in the hybrid bivariate model if the associated parameters were significant at the 0.05 level. Tables 2 through Tables 4 present the estimates of respectively the measurement equations, the structural equations and the hybrid bivariate model. The log-likelihood of the presented model is equal to −509.87 and the adjusted McFadden rho-square is equal to 0.630. Likelihood ratio tests reveal that the model is to be preferred to two separate models for the PDO and the PI road crashes (LRT = 130.92, df = 1, p = 0.000) as well to a bivariate model without unobserved traits (LRT = 41.32, df = 6, p = 0.000). The correlation between the error terms shows also that the bivariate model is the preferred specification as it is equal to 0.451 and it is significantly different from zero at the 0.01 confidence level.

The measurement equations in Table 2 used the exploratory analysis for the specification. Fatigue relates to the feeling of fatigue, the perception of sleep problems, and the dissatisfaction with the salary that probably captures stress that in turn relates to the other two items. Reckless behaviour correlates with violating the traffic laws concerning the use of mobile phones, the respect of red lights and the regard of speed limits. Aggressive behaviour associates with situations where aggression plays a part, either generic or specific such as accelerating in presence of a yellow light and overtaking without notifying other drivers.

The structural equations in Tables 3 reveal the variation of the three unobserved traits with the observed characteristics of the taxi drivers. Fatigue is associated with older age of the drivers, higher amount of working hours daily, lack of off days during the week, and a management fee over 150 CHY per day (i.e., the parameters for the three categories of management fees were statistically equal). Male and older drivers are linked to more reckless and aggressive behaviour, and so is a

heavier workload in terms of working hours per day. Moreover, a heavier workload in terms of a lower number of days off is also related to more reckless behaviour.

The hybrid bivariate model (in Table 4) reveals that all three unobserved traits are associated with a higher propensity of being involved in both PDO and PI crashes. Intuitively, fatigue is positively related to the likelihood of crashing and so are the reckless and aggressive behaviour. Interestingly, the number of days off per week has a direct effect on the probability of being involved in either a PDO or a PI crash that has to be added to the indirect effect on the tendency to be fatigued and to be more reckless while driving. Interestingly as well, the workload in terms of number of working hours per day has also a direct effect on the likelihood of being involved in a PI crash that has to be summed to the indirect effect on all three unobserved traits. Overall, a heavy workload contributes to the propensity of crashing (and in particular the propensity of being involved in a crash with injured road users) both directly and indirectly.

4. Discussion and conclusions

This study aimed at investigating three research questions focusing on the workload and the driving behaviour of taxi drivers, as well as understanding the relation between the involvement in PDO and PI road crashes and both observed characteristics (e.g., sociodemographic characteristics, workload conditions) as well as unobserved traits of taxi drivers (e.g., level of fatigue, measure of aberrant behaviour). A

Table 3
Estimates of the structural equations of the unobserved traits.

Fatigue	estimate	t-stat	p
Age between 31 and 40 years old	0.141	4.55	< 0.01
Age over 40 years old	0.189	4.97	< 0.01
Over 10 working hours daily	0.158	2.08	< 0.05
Number of days off duty in a week	−0.173	−3.53	< 0.01
Daily management fee over 150 CNY	0.080	2.16	< 0.05
Constant	0.235	3.92	< 0.01
Reckless behaviour	**estimate**	**t-stat**	**p**
Male	0.205	2.44	< 0.05
Age over 40 years old	0.294	4.32	< 0.01
Over 10 working hours daily	0.320	3.52	< 0.01
Number of days off duty in a week	−0.162	−2.66	< 0.05
Constant	0.333	7.00	< 0.01
Aggressive behaviour	**estimate**	**t-stat**	**p**
Male	0.135	2.60	< 0.01
Age over 40 years old	0.134	3.05	< 0.01
Over 10 working hours daily	0.141	2.07	< 0.05
Constant	0.257	5.35	< 0.01

Table 2
Estimates of the measurement equations of the unobserved traits.

Fatigue	estimate	t-stat	p
How often do you have sleep problems?	1.000	–	–
How often do you feel fatigued while driving?	1.164	16.87	< 0.01
Are you unsatisfied with your current salary?	0.956	12.10	< 0.01
Reckless behaviour	**estimate**	**t-stat**	**p**
How often do you cross a junction with a red light?	1.000	–	–
How often do you disregard the speed limits?	0.953	50.16	< 0.01
How often do you use your mobile phone while driving?	1.319	19.98	< 0.01
Aggressive behaviour	**estimate**	**t-stat**	**p**
How often do you overtake a vehicle without using the left/right turn light?	1.000	–	–
How often do you drive aggressively?	1.365	20.68	< 0.01
How often do you accelerate to cross a junction with a yellow light	1.014	56.33	< 0.01

Note: the estimates of the constants and the standard deviations are omitted for brevity.

Table 4
Estimates of the joint hybrid bivariate model.

Property damage crashes	estimate	t-stat	p
Number of days off duty in a week	−0.389	−2.43	< 0.05
Fatigue	1.111	5.07	< 0.01
Reckless behaviour	0.607	3.03	< 0.01
Aggressive behaviour	1.023	4.92	< 0.01
Threshold (0–1 / 2–3 crashes)	0.537	3.14	< 0.01
Threshold (2–3 / 3–4 crashes)	1.516	8.87	< 0.01
Threshold (3–4 / 5 or more crashes)	2.236	12.29	< 0.01
Injury crashes	**estimate**	**t-stat**	**p**
Over 10 working hours daily	0.328	2.28	< 0.05
Number of days off duty in a week	−0.475	−2.68	< 0.05
Fatigue	2.103	9.35	< 0.01
Reckless behaviour	1.169	3.54	< 0.01
Aggressive behaviour	1.354	6.57	< 0.01
Constant	0.800	2.89	< 0.05
Error terms	**estimate**	**t-stat**	**p**
Correlation	0.451	4.14	< 0.01

multivariate analysis was performed via the specification of a joint hybrid bivariate model of crash involvement propensity.

The answer to the first research question is that self-reports of 1021 taxi drivers in four Chinese cities suggest that taxi drivers have to endure a very heavy workload. Over 75% of the drivers declared to work 9 hours or more per day, and over 70% revealed not to have even one day per week of rest. Alongside the stress because of the workload, the taxi drivers have to manage the stress because of the income, given that the management fee to the taxi companies amount on average to a sum comparable (if not higher) to their net income. The information about the workload portrays a picture even more worrisome than the heavy workload already reported in Australia and China (Dalziel and Job, 1997; Shi et al., 2014), while the stress about the income reflects previous findings showing that a sizeable percentage of taxi drivers admits being under extreme economic pressure (Lim and Chia, 2015; Shi et al., 2014).

Moreover, self-reports of the 1021 taxi drivers suggest also that they are engaging frequently in aberrant driving behaviour. In particular, the reported levels of disregarding traffic laws about mobile phone use, red light stop, and speed limits appear in line with attitudes of taxi drivers towards driving law violations (Ma et al., 2010). Aggressive behaviour emerged also from the self-reports and the finding from the joint hybrid bivariate model that males and older taxi drivers engage more in this type of behaviour is in line with previous findings from Taiwan (Tseng, 2013) and China (Wang et al., 2014).

The answers to the second and third research question are found in the estimates of the hybrid bivariate model of the propensity of being involved in a PDO or PI road crash. Specifically, the model uncovered the relations with observed characteristics and unobserved traits of the taxi drivers while controlling for unobserved correlation between the probability of being in either type of crash. Although previously employed in the travel behaviour literature (e.g., Dias et al., 2017; Emmerink et al., 1996; Ferdous et al., 2010), the model specification in this study introduced the latent variable dimension and hence required the integration of the probability over the distribution of the latent variables.

The findings from this study agree with previous studies that found taxi drivers being vulnerable to fatigue or sleepiness (Dalziel and Job, 1997; Meng et al., 2015; Zhao et al., 2015), but also answer positively to the research questions about the working conditions relating to the fatigue and in turn to the propensity of being involved in a crash. Moreover, the findings indicate that fatigue was related to income dissatisfaction in the measurement equations and to the licensing fee in the structural equations. This suggests that the financial pressure from both an actual (the management fee) and an attitudinal (the income dissatisfaction) perspective is associated indirectly with the propensity of being part of a crash.

Relevantly, the model estimates show how the relation between working conditions and propensity to be involved in crashes is not only a direct one, but also an indirect one because the working conditions increase the level of fatigue and the manifestation of reckless and aggressive behaviour, thus significantly relating to the risk on the road. When considering in particular the direct effects, the lack of days off is associated with a higher propensity of being in a PDO or PI crash, while the amount of working hours per day is associated only with the probability of being in a crash where road users were injured. The finding about the number of days off is even more worrisome when considering that their average number in the sample is equal to only 0.3 days per week.

A few limitations of this study are worth noting. Firstly, the study is based on a sample of taxi drivers that might not be representative of all the professional taxi drivers in China. However, the multivariate analysis surely captures significant relations that could be investigated in the future on a larger scale. Secondly, the study is based on self-reported information that might imply biases in the reporting of crashes and the aberrant behaviour. However, the fact that almost the entire sample declared to have been involved in at least one PDO crash suggests that the taxi drivers were not refraining from declaring to have been in road crashes. Moreover, recent literature has showed that self-reports are good proxies for objective measures of risk retrieved via driving simulators and in-vehicle data recorders (Taubman-Ben-Ari et al., 2016).

The findings from this study suggest the measures should be in place to reduce fatigue, stress, and aberrant driving behaviour. A solution to the fatigue and stress problem is the release of the heavy economic burden (Rahmani et al., 2013) via different measures: (i) a decrease in the management fee that amounts for nearly half of the taxi drivers' gross income; (ii) an increase in the possibility to decide for flexible working hours that would allow drivers to schedule shifts and rest more effectively; (iii) a reform in the management of the taxi industry with stricter regulations about working hour upper limits, management fee reductions, and possibly employees' benefits that would reduce the stress. A solution to the reckless and aggressive behaviour problem is the education of taxi drivers about the need for attention to traffic laws and the introduction of the aforementioned measures that would limit fatigue and in turn curtail risky behaviour (Wang et al., 2014; Wang et al., 2017). While a reform in the management of the taxi industry appears a government problem, taxi companies could improve working conditions by promoting training and safety education, negotiating flexible schedules, and establishing trade unions able to tackle the management fee issue.

Overall, the findings from this study shed considerable light on the factors affecting road crashes involving taxi drivers in China. Further research could look into the company and the government perspective by verifying what affects the actionability and effectiveness of the aforementioned measures with the aim of improving dramatically the safety performance of the taxi industry. Moreover, further research could investigate the possible effect of a help line where both passengers and drivers could complain about service issues and solve disputes as proposed for improving the safety of taxi drivers in Beijing (Shi et al., 2014). Lastly, further research could investigate whether other markets with a different management structure experience the same issues, and in the case that they do not, observe which measures were successful in limiting fatigue and aberrant behaviour of taxi drivers.

Acknowledgements

References

Bierlaire, M., 2016. PythonBiogeme: a Short Introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Dalziel, J.R., Job, R., 1997. Motor vehicle accidents, fatigue and optimism bias in taxi drivers. Accid. Anal. Prev. 29, 489–494.

Dias, F.F., Lavieri, P.S., Garikapati, V.M., Astroza, S., Pendyala, R.M., Bhat, C.R., 2017. A behavioral choice model of the use of car-sharing and ride-sourcing services Washington, D.C. Presented at the 96th Annual Meeting of the Transportation Research Board.

Emmerink, R.H., Nijkamp, P., Rietveld, P., van Ommeren, J.N., 1996. Variable message signs and radio traffic information: an integrated empirical analysis of drivers' route choice behaviour. Transp. Res. Part A Policy Pract. 30, 135–153.

Ferdous, N., Eluru, N., Bhat, C.R., Meloni, I., 2010. A multivariate ordered-response model system for adults' weekday activity episode generation by activity purpose and social context. Transp. Res. Part B Methodol. 44, 922–943.

Fleiter, J.J., Gao, L., Qiu, C., Shi, K., 2009. Availability, functionality, and use of seat belts in Beijing taxis prior to the 2008 Beijing olympic games. Accid. Anal. Prev. 41, 342–344.

Hess, S., Train, K.E., Polak, J.W., 2006. On the use of a modified latin hypercube sampling (MLHS) method in the estimation of a mixed logit model for vehicle choice. Transportation Part B Methodol. 40, 147–163.

Higgins, J.S., Michael, J., Austin, R., Åkerstedt, T., van Dongen, H.P.A., Watson, N.,

Czeisler, C., Pack, A.I., Rosekind, M.R., 2017. Asleep at the wheel - the road to addressing drowsy driving. Sleep 40 zsx001.

La, Q.N., Lee, A.H., Meuleners, L.B., Duongc, D.V., 2013. Prevalence and factors associated with road traffc crash among taxi drivers in Hanoi, Vietnam. Accid. Anal. Prev. 50, 451–455.

Lim, M., Chia, S.E., 2015. The prevalence of fatigue and associated health and safety risk factors among taxi drivers in Singapore. Singapore Med. J. 56, 92–97.

Ma, M., Yan, X., Huang, H., Abdel-Aty, M., 2010. Safety of public transportation occupational drivers risk perception, attitudes, and driving behavior. Transp. Res. Rec. 2145, 72–79.

Maag, U., Vanasse, C., Dionne, G., Laberge, N.C., 1997. Taxi drivers' accidents: how binocular vision problems are related to their rate and severity in terms of the number of victims. Accid. Anal. Prev. 29, 217–224.

Mann, R.E., Stoduto, G., Ialomiteanu, A., Asbridge, M., Smart, R.G., Wickens, C.M., 2010. Self-reported collision risk associated with cannabis use and driving after cannabis use among Ontario adults. Traffic Inj. Prev. 11, 115–122.

Martinussen, L.M., Lajunen, T., Møller, M., Özkan, T., 2013. Short and user-friendly: the development and validation of the mini-DBQ. Accid. Anal. Prev. 50, 1259–1265.

Meng, F., Li, S., Cao, L., Li, M., Peng, Q., Wang, C., Zhang, W., 2015. Driving fatigue in professional drivers: a survey of truck and taxi drivers. Traffic Inj. Prev. 16, 474–483.

Raanaas, R.K., Anderson, D., 2008. A questionnaire survey of Norwegian taxi drivers' musculoskeletal health, and work-related risk factors. Int. J. Ind. Ergon. 38, 280–290.

Rahmani, A., Khodaei, R., Farjami, A., Mahmodkhani, S., Gharagozlou, F., Ahmadnezhad, I., Vatani, J., 2013. Determination of job stresses and their consequences in drivers in Ilam. Electron. Phys. 5, 594–598.

Reason, J.T., Manstead, A.S.R., Stradling, S.G., Baxter, J., Campbell, K., 1990. Errors and violations on the road – a real distinction. Ergonomics 33, 1315–1332.

Routley, V., Ozanne-Smith, J., Qin, Y., Wux, M., 2009. Taxi driver seat belt wearing in Nanjing, China. J. Saf. Res. 40, 449–454.

Shi, J., Tao, L., Li, X., Atchley, P., 2014. A survey of taxi drivers' aberrant driving behavior in Beijing. J. Transp. Saf. Secur. 6, 34–43.

Sullman, M.J.M., Stephens, A.N., Kuzu, D., 2013. The expression of anger amongst Turkish taxi drivers. Accid. Anal. Prev. 56, 42–50.

Taubman-Ben-Ari, O., Eherenfreund-Hager, A., Prato, C.G., 2016. The value of self-report measures as indicators of driving behaviors among young drivers. Transp. Res. Part F Traffic Psychol. Behav. 39, 33–42.

Tseng, C.M., 2013. Operating styles, working time and daily driving distance in relation to a taxi driver's speeding offenses in Taiwan. Accid. Anal. Prev. 52, 1–8.

Valent, F., Di Bartolomeo, S., Marchetti, R., Sbrojavacca, R., Barbone, F., 2010. A case-crossover study of sleep and work hours and the risk of road traffic accidents. Sleep 33, 349–354.

Vayalamkuzhi, P., Amirthalingam, V., 2016. Influence of geometric design characteristics on safety under heterogeneous traffic flow. J. Traffic Transp. Eng. (Engl. Ed.) 3, 559–570.

Vij, A., Walker, J., 2016. How, when and why integrated choice and latent variable models are latently useful. Transp. Res. Part B Methodol. 90, 192–217.

Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Math. Soc. Sci. 43, 303–343.

Wang, Y., Li, M., Du, J., Mao, C., 2015. Prevention of taxi accidents in Xi'an, China: what matters most? Cent. Eur. J. Public Health 23, 77–83.

Wang, Y., Li, L., Feng, L., Peng, H., 2014. Professional drivers' views on risky driving behaviors and accident liability: a questionnaire survey in Xining, China. Transp. Lett. 6, 126–135.

Wang, Y., Xin, M., Bai, H., Zhao, Y., 2017. Can variations in visual behavior measures be good predictors of driver sleepiness? A real driving test study. Traffic Inj. Prev. 18, 132–138.

Zhao, Y., Zhang, J., He, X., 2015. Risk factors contributing to taxi involved crashes: a case study in Xi'an, China. Period. Polytech. Transp. Eng. 43, 189–198.