Modelling crash propensity of carshare members

Vinayak Dixit*, Taha Hossein Rashidi

School of Civil and Environmental Engineering, University of New South Wales, UNSW, Gate 11, Botany Street, HI0 CVEN, L1, Room 113, Sydney, NSW 2031, Australia

A R T I C L E   I N F O

Article history:
Received 1 May 2013
Received in revised form 1 March 2014
Accepted 11 March 2014
Available online 12 April 2014

Keywords:
Carshare
Safety
Censored biprobit
Quasi-induced exposure

A B S T R A C T

Carshare systems are considered a promising solution for sustainable development of cities. To promote carsharing it is imperative to make them cost effective, which includes reduction in costs associated to crashes and insurance. To achieve this goal, it is important to characterize carshare users involved in crashes and understand factors that can explain at-fault and not-at-fault drivers. This study utilizes data from GoGet carshare users in Sydney, Australia. Based on this study it was found that carshare users who utilize cars less frequently, own one or more cars, have less number of accidents in the past ten years, have chosen a higher insurance excess and have had a license for a longer period of time are less likely to be involved in a crash. However, if a crash occurs, carshare users not needing a car on the weekend, driving less than 1000 km in the last year, rarely using a car and having an Australian license increases the likelihood to be at-fault. Since the dataset contained information about all members as well as not-at-fault drivers, it provided a unique opportunity to explore some aspects of quasi-induced exposure. The results indicate systematic differences in the distribution between the not-at-fault drivers and the carshare members based on the kilometres driven last year, main mode of travel, car ownership status and how often the car is needed. Finally, based on this study it is recommended that creating an incentive structure based on training and experience (based on kilometres driven), possibly tagged to the insurance excess could improve safety, and reduce costs associated to crashes for carshare systems.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Since the 90s innovators, industries, cities and decision makers (Prettenhalter and Steininger, 1999) have been working to materialize the concept of car sharing in large cities of Australia, Switzerland and Germany, as well as Sweden, the Netherlands, Canada and the United States on a smaller scale (Mont, 2004; Steininger et al., 1996). An efficiently designed carsharing system would provide an ideal alternative to private vehicle ownership, car leasing and renting a car, by providing users the ability to have flexibility and travel large distances while maintaining their predominant choice of mode as public transit, walking, bicycling, taxi etc. Shaheen et al. (2009) provided a comprehensive review of the evolution of carsharing, and its value towards development of a sustainable transport system. Several other studies have also discussed the sustainability benefits of carsharing (Fellow and Pitfield, 2000; Steininger et al., 1996; Cervero et al., 2007; Firnkorn and Muller, 2011).

Recently there have been studies (Habib et al., 2012; Constain et al., 2012) that have investigated choice behaviour of carsharing users regarding: membership duration, the decision to be an active member, and the frequency of monthly usage. It was found that usage costs are the biggest determinants of the usage of carshare systems. Members of a carshare system do not deal with fuel or insurance costs but pay by use. Therefore fuel and insurance costs are embedded in the usage costs, which depend on the time period of use and/or kilometres travelled. To ensure the success of carsharing, it is imperative to lower overall costs, which also includes reducing insurance costs. Therefore, it is critical to understand crash risks of various individuals participating in a carshare scheme.

To the best of the authors’ knowledge, after an extensive review of literature, this study is the first to study crash risk of carshare users. This research studies the discrete outcomes of crashing and being at-fault. Previous studies have assessed crash characteristics using discrete outcome models, such as the binary logit (Haque et al., 2009), ordered probit (Pai and Saleh, 2008), multinomial logit (Shankar and Manering, 1996), nested logit (Savolainen and Manering, 2007), and mixed logit (De Lapparent, 2006). Zhang (2010) identifies the assumptions and strengths of these models.
Unlike most crash datasets in which only information of people involved in accidents are present, a unique aspect of the dataset used in this study is that it contains information regarding all members, including those who are not involved in crashes as well as those involved in crashes and identified as at-fault or not-at-fault drivers. This also provides an opportunity to study attributes of not-at-fault drivers which play an important role in crash studies utilizing quasi-induced exposure methods.

Characterizing the cause of crashes relies heavily on normalizing occurrences of crashes based on risk exposure. Kilometres driven by a particular driver is widely accepted as a representation of this exposure. Unfortunately, crash databases do not typically contain this information making estimation and approximation of the exposure rate a challenging issue. To address this issue the concept of induced exposure was developed by Thorpe (1967). Induced exposure was found to have typical issues associated with assigning responsibility for crash causation, which led to the development of quasi-induced exposure (Haight, 1970).

Quasi-induced exposure method is based on the premise that not-at-fault drivers are a random sample of the driving population, which is utilized to characterize the at-fault drivers by studying the involvement ratio, i.e. the ratio of at-fault drivers to not-at-fault drivers for a certain group. An important advantage of this method is that it does not require other exogenous measures of risk exposure such as vehicle kilometres travelled, traffic volumes at a location etc., and can rely solely on the crash database. Due to these advantages researchers and practitioners (Yan et al., 2005; Chandraratna and Stamatiadis, 2005; Jiang and Lyles, 2010; Mendez and Izquierdo, 2010; Jiang et al., 2011, 2012) have begun to heavily rely on quasi-induced exposure method to characterize crash risks. Due to the increasing popularity of this method, it has become imperative to study the validity of the underlying assumptions, such as that of not-at-fault drivers being a random sample of the driving population, which also requires that crash responsibilities are correctly assigned. Jiang et al. (2011) used the U.S. National Household Survey to validate the Quasi Induced exposure method, and found the results to be promising.

Studies that have undertaken to evaluate quasi-induced exposure method can be categorized into two groups. The first type of these studies have focussed on biases associated to police officer’s judgement in assigning fault to drivers during a crash. These biases in judgements can result in violation of the assumption of quasi-induced exposure regarding not-at-fault drivers being a random sample of all drivers. DeYoung et al. (1997) identified “negative halo effect”, where in an investigating officer might assign fault to drivers’ with suspended/revoked license, alcohol/drug use or based on other negative perceptions about the driver, despite not being objectively responsible for the crash. Kirk and Stamatiadis (2001) as well as Lenguerrand et al. (2008) found such biases in the crash data set. Rather than using the police assigned crash responsibility, the studies used exogenous methods to identify crash responsibility. Chandraratna and Stamatiadis (2009) when analysing multi-vehicle crashes found evidence of “negative halo effects” biasing the representation of not-at-fault drivers. Recently Jiang et al. (2012) found that hit-and run, gender, age, injury severity, and alcohol and illegal drug use significantly impact investigating officers’ decision making.

The second type of studies can be characterized as those which compare quasi-induced exposure with traditionally used exposure metrics such as vehicle miles travelled or traffic volumes based on time of day or other disaggregate characteristics based on environmental, vehicle, roadway or driver characteristics. Lighthizer (1989), in one of the earliest attempts, compared key variables regarding crashes using the quasi-induced exposure method using the Michigan crash data. The study found the assumptions of the quasi-induced exposure method were met. Kirk and Stamatiadis (2001) were also able to qualitatively show the validity of the quasi-induced approach using the Kentucky crash database. On the contrary, there have been several studies that have found not-at-fault drivers to have significant under representation of vehicles with new technologies (Evans, 2004) and drivers belonging to younger age groups (Kahane and Hertz, 1998) that have higher accident avoidance capabilities. Drivers having higher speed have also been found to have a higher representation in not-at-fault drivers (Mendez and Izquierdo, 2010), mainly due to higher number of accident prone interactions (Navon, 2003) and reduced ability to avoid crashes. Jiang and Lyles (2007) also found that differentials in average speed between vehicle types and road users can affect involvement ratio while studying quasi-induced exposure.

In this respect, researchers (Stamatiadis and Deaco, 1997; Jiang and Lyles, 2007; Chin and Haque, 2010; Haque et al., 2012) have suggested the use of ‘clean’ dataset when utilizing quasi-induced exposure methods. As highlighted by Haque et al. (2012).

“Since induced exposure estimation relies on the fault of crash involvement, biased cases of fault assignment need to be removed.”

This new ‘clean’ dataset is then compared with the entire dataset to determine whether the cleaning process resulted in any systematic biases, before being used in studying crash likelihood.

As identified by Chandraratna and Stamatiadis (2009), for the quasi-induced exposure method, the assumption of randomness of the not-at-fault driver sample is critical. They state:

“In statistical terms, a simple random sample is a set of drivers that have been selected from the driver population in such a way that every driver had an equal opportunity to be involved in a crash without being the at-fault driver. In other words, since the driver at-fault does not intentionally select a driver to strike, it can be reasonably assumed that each driver has an equal chance to be included in the not-at-fault driver sample.”

Therefore, under this assumption of randomness, carshare members involved in a crash and not-at-shoud be a random sample of the carshare members. Since every member had an equal opportunity to be involved in a crash and be not-at-fault. However, if there is a non-linear relationship between certain driver, carsharing or exposure characteristics this might not hold true. The unique GoGet dataset contains information about the members as well as drivers involved in crashes but were not-at-fault, therefore providing a unique opportunity to evaluate this assumption associated to quasi-induced exposure.

The study also characterizes carshare users who have the propensity to be involved in a crash as well as be-at-fault. This study utilizes member and crash database from the Sydney GoGet carshare users to evaluate the quasi-induced exposure method. The next section describes the data used for the analysis.

2. Data

This analysis utilizes GoGet member crash data in Sydney, New South Wales (NSW), Australia. The purpose of this study is to evaluate factors that affect the risk propensity of at-fault and not-at-fault drivers. The data was collected during the period from August 2010 to July 2012. GoGet have about 1000 vehicles located at strategic carshare locations around Sydney. The responsibility of the crash was assigned by the insurance company based on vehicle crash characteristics. The insurance company has no information about the personal details of the individual members involved in the crash. During the period of study the data included a total of 25,120 members, of which 161 were at-fault crashes, and 106 were not-at-fault crashes. This is shown in Table 1. The database included
Table 1
Crash classification of members in the data.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>No crash</td>
<td>24,847</td>
<td>98.94</td>
</tr>
<tr>
<td>Not at fault</td>
<td>106</td>
<td>0.42</td>
</tr>
<tr>
<td>At fault</td>
<td>161</td>
<td>0.64</td>
</tr>
</tbody>
</table>

single and multivehicle crashes. Fig. 1 shows the extent of driving in Sydney during the analysis.

A comparison of representation of the different demographics in the carshare crash database and the carshare membership database is shown in Table 1. All the members of the carshare are also referred to as carshare member population. Several individual level characteristics were available and were used in this analysis, however, due to privacy concerns some of the variables associated to demographics, vehicle characteristics and exact location of the crashes were not available for analysis. The variables are shown in Table 2.

A statistical comparison was conducted between the carshare member population and those that crashed. The variables can be classified as categorical and continuous variables. A Chi-Squared test was conducted to compare statistical differences between the two groups for the binary variables. A Mann–Whitney–Wilcoxon test (Mann and Whitney, 1947) was used to compare statistical differences between continuous (ordered) variables. The left most columns of Tables 2 and 3 show the p-values for the statistical tests.

Sixty percent of the crashes were found to consist of at-fault drivers as compared to not-at-fault. Members involved in crashes were found to have statistically higher kilometres driven last year as well as had a statistically lower number of years with a driving license.

The distribution of drivers who crashed and the carshare member population were significantly different for: “How Often You Need a Car”, “Vehicle Ownership”, “Live near a Dedicated Parking” and “Main Mode of Transport”.

The variable living near a dedicated parking identifies whether the member lives near a dedicated carshare parking spot. It was found that members living near a dedicated carshare parking spot had a higher representation in the crash database as compared to the member database (Table 2). Further investigation indicated a significant correlation (−0.45) between lives driven less than 1000 km last year and living near a dedicated parking. To explore these relationships further a systematic statistical model is developed.

To explore the assumption of not-at-fault carshare members being a random sample of the carshare membership, Table 3 compares the different socio-demographic categories for GoGet members involved in not-at-fault crashes and the general membership pool. The assumption of random sample is a key assumption of quasi-induced exposure. The last column of Table 3 reveals that not-at-fault drivers have statistically significant higher kilometres driven last year as compared to the carshare member population. Janke (1991) found a non-linear relationship between kilometres driven and the propensity to be involved in crashes. This non-linearity could potentially breakdown the randomness assumption of not-at-fault drivers, and therefore not-at-fault carshare drivers are found to have higher kilometres driven last year compared to the carshare members.

Statistically significant differences were also observed between not-at-fault drivers and the carshare member population in the case of, “How Often You Need a Car”, “Vehicle Ownership”, “Main Mode of Transport”, “Live near a dedicated parking”, “Kilometers driven last year” and “Number of years with Driver’s License”. This suggests some biases in the quasi-induced exposure method, which need to be further analysed to characterize any systematic differences.

It is observed that occasional drivers (once or twice a month, once or twice a week, and weekends) are over-represented in the not-at-fault carshare database. Initial hypothesis was that occasional drivers would have significantly different kilometres driven last year, however, subsequent analysis showed no correlations with kilometres driven last year (correlation ranged between −0.03 and 0.02), suggesting that the occasional drivers might be driving less frequently but larger distances for trip purposes, during time periods (Lardelli-Claret et al., 2003) and roads which have different exposure to crash risk. This biased exposure has been discussed in numerous studies (Janke, 1991; Hakamies-Blomqvist et al., 2002; Langford and Koppel, 2005; Keall and Frith, 2006 and Langford et al., 2008). A similar trend is found among carshare members who do not use cars but use bicycle, train and bus as their main mode of transport. Though no correlations are observed between the variable and kilometres driven last year (ranging from −0.02 and 0.002), these demographics are found to be over-represented in the not-at-fault database, while members who use cars as their main mode of travel are found to be under-represented in the not-at-fault database.

To control unobserved heterogeneities and correlations, covariates that describe carshare users as well as driver characteristics are used to develop a model. This provides the ability to determine factors that are able to statistically characterize carshare users that are likely to be involved in crashes as well as be at-fault with respect to those who did not crash. Based on the above qualitative analysis, the above demographics were further investigated to characterize any significant impact they might have on crash propensity and being at-fault or not-at-fault.

3. Methodology

It is critical to recognize that studying the propensity to be at-fault or not-at-fault is prone to the classic selection problem, where in subjects at-fault or not-at-fault are only observed among members who have been involved in crashes, and there is unobserved heterogeneities that impact the model.

The basic selection problem arises because the members who were at fault \( (f) \) consist of only members involved in crashes \( (c) \), and these members may differ in important unmeasured ways from those who were not involved in crashes. These unmeasured heterogeneities need to be appropriately controlled for in the models.

Assume that the attributes explaining the propensity to crash \( (c) \) and being at fault \( (f) \) if a crash happens are described by the equations below, where \( Z_i \) and \( X_i \) are the measured attributes.
available in the data, and $u_i$ and $e_i$ are the respective errors for the two models.

$c_i = w_iZ_i + u_i$

$f_i = \beta_iX_i + e_i$

To control for the unobserved heterogeneities we allow the error terms to have the following specification:

$E[u_i|Z_i, X_i] = E[e_i|Z_i, X_i] = 0$

$Var[u_i|Z_i, X_i] = Var[e_i|Z_i, X_i] = 1$

$Cov[u_i, e_i|Z_i, X_i] = \rho$

This is the equivalent of Heckman’s selection model (Heckman, 1979) except that a probit model is used for both the selection equation the outcome equation and the second model also has a discrete dependent variable. This model is referred to as censored biprobit. In the model, we account for the fact that we observe fault ($f_i = 1$) only if crash happens ($c_i = 1$). However, if $c_i = 0$, then we have no information about fault. Thus, the first probit equation is completely observed, but we have only a selected (censored) sample for the second. Note that in terms of the four possible outcomes, two ($c_i = 0, f_i = 1$ and $c_i = 0, f_i = 0$) are indistinguishable. There are three types of observations in a sample with the following probabilities.

$c = 0 \Pr(c = 0) = 1 - \Phi(x_1\beta_1)$

$c = 1; f = 0 \Pr(c = 1; f = 0) = \Phi(x_1\beta_1) - \Phi(x_1\beta_1, x_2\beta_2; \rho)$

$c = 1; f = 1 \Pr(c = 1; f = 1) = \Phi_2(x_1\beta_1, x_2\beta_2, \rho)$

From this, it is easy to generate the log-likelihood function

$\ln L = \sum_{i=1}^{N} \left( c_i f_i \ln[\Phi_2(x_1\beta_1 x_2\beta_2; \rho)] + c_i (1 - f_i) \ln[\Phi(x_1\beta_1)] - \Phi_2(x_1\beta_1 x_2\beta_2; \rho) + (1 - c_i) \ln[1 - \Phi(x_1\beta_1)] \right)$

As with the Heckman model, at least one variable in the selection equation should not appear in the outcome equation. The model was estimated using STATA.

4. Results

To characterize the crash propensity as well as at-fault and not-at-fault several discrete choice models were developed including multinomial logit (MNL), and nested logit models. It was observed
that the Identically Independent Alternative assumption is violated by using the MNL formulation making its usage indefensible. Nested logit formulations seemed to be a suitable candidate for modelling the discrete set of alternatives of crash-fault, crash-no-fault and no-crash, however, by obtaining values of greater than one for the coefficient of inclusive value, it was concluded that nested logit modelling framework for this dataset is inconsistent with utility maximization.\(^2\) Therefore to systematically account for the correlation between the two layers of crash and no-crash and fault and no-fault was formulated using the censored biprobit structure presented in the previous section.

Based on the Wald Chi Square test (22.59) for the model shown in Table 4, the general goodness-of-fit of the model is acceptable. In addition, the overall results observed in Table 4\(^3\) agree with the qualitative findings discussed based on Tables 2 and 3.

---

\(^2\) After developing the nested logit model, it was found that the estimation for the value of correlation parameter does not fall between 0 and 1 meaning that the estimation is not consistent with the maximum utility maximization assumptions and the parameters estimations are not consistent. This might be due to the violation of the IIA assumption within the nests. Which in this study corresponds to not-at-fault drivers are independent of at-fault drivers, however as identified by earlier studies (Stamatiadis and Deaco, 1997; Jiang and Lyles, 2007; Chin and Haque, 2010; Haque et al., 2012) there might be systematic biases that occur in fault assignment that would have violated this assumption.

\(^3\) The final model presented includes only significant variables at an 85% confidence interval. All variables were significant at 90% confidence interval, except for the variable left hand traffic driver's license (p-value of 0.12).

### 4.1. Crash model

Several variables were found to affect the propensity of being involved in a crash. Based on the results shown in Table 4, the “needing a car less often” (represented by variables: Rarely need a car and Need car once or twice a month) or “having driven less than 1000 km last year” decreases the chance of a crash. In other words, if a person drives less, it gets less likely that the person would be involved in a crash. Similarly, the coefficient for owning one or more cars was found to be negative. This trend is attributed to the decrease in the need for carsharing with vehicle ownership, which eventually decreases the exposure to accidents while using the carsharing system.

The duration of holding a driver’s license was found to have a negative coefficient. The duration a member has held a driving license is highly correlated with age and the years of driving experience, therefore implying that more experienced drivers are less likely to be involved in crashes (Kahane and Hertz, 1998; Dixit et al., 2011; Dixit, 2013; Dixit et al., 2014). Living near a dedicated carshare parking was found to increase the propensity of crash. Further investigation indicated a significant correlation (−0.45) between driven less than 1000 km last year and living near a dedicated parking. Therefore, people living near a dedicated parking tend to drive more and therefore have an increased crash exposure.

If the carshare member is using a cycle (bike) as his/her main mode of transport, (s)he is more likely to be involved in a crash. As discussed earlier this could be a result of exposure biases associated to driving during different time periods and roads, further
investigations are warranted to explore relationships between these factors. In addition, previous instance of accidents in the past ten years is also found to increase the likelihood of crashing.

The coefficient for variable representing people having insurance excess of $300 was found to be positive, indicating that people with higher coverage, i.e. needing to pay $300 instead of $1500 as excess in the event of a crash, increases the likelihood of a member being involved in a crash. However, no such statistically significant impact was observed on fault assignment. Therefore, there seems to be no biases in fault assignment based on insurance policies. However, the fact that there is a higher likelihood of members in a higher excess plan to be involved in crashes suggests perhaps that drivers in higher insurance coverage are associated to higher crash risk, which is consistent with previous studies (Cummins et al., 2001; Cohen, 2005; Sandroni and Squintani, 2013). Which has been attributed to drivers learning about their risk type (higher crash risk propensity), and then self-selecting into a higher coverage group (Cohen, 2005).

It is important to recognize that the crash model identifies the likelihood (exposure) of individual members being involved in crashes. The variables that were used in the model were those that were observed, however unobserved factors that affect the likelihood of crash could affect the likelihood of being at-fault, and this is controlled for by properly accounting for these correlations in the censored biprobit model.

### 4.2. Fault model

In the fault model four variables were found to be statistically significant, which are listed in Table 4. It should be noted that the fault model is a conditional model given that a crash occurs. Based on the results, it can be concluded that individuals who use cars less frequently are more likely to be at-fault when a crash occurs. This statement is supported by the positive coefficients for variables: rarely needing a car and driving less than 1000 km last year. On the contrary, if a carshare user needs the service mainly for weekends, then the propensity of being at fault decrease for him/her.

In addition, the variable Australian driver’s license was found to be marginally significant with a p-value of 0.12, the variable is binary variable with members with Australian drivers license having a value of 1, while members having drivers licenses from other countries having a value of 0. A positive coefficient indicates that given a crash occurs, drivers who have Australian licenses are more likely to be at-fault. This implies that there is a difference in at-fault crash risk between immigrants and the population, which is contrary to findings by Redelmeier et al. (2011).

A significant finding of this study pertains to the correlation in the error terms being statistically significant. This is shown in Table 4, where the correlation coefficient is found to be negative and statistically significant at a 95% confidence level. This implying that there are unobserved heterogeneities that affect both the likelihood of crash as well as the likelihood of being at-fault given a crash occurs. Therefore, the model correctly accounts for the correlation between the error terms of these two models. The negative correlation coefficient implies that the unobserved factors that increases the likelihood of crash reduces the likelihood of being at-fault given a crash occurs.

The analysis carried out on a dataset of carshare users used to develop model estimates that provide important insights regarding crash and at-fault likelihood of carshare members. In addition, the unique dataset also provided an opportunity to explore certain assumptions about quasi-induced exposure.

### 5. Conclusion

Carshare systems are seen as useful options to have sustainable cities. Promoting cost effective carshare systems requires reduction of costs associated to crashes and insurance. To achieve this goal, it is imperative to characterize carshare users involved in crashes and understand factors that can explain at-fault and not-at-fault drivers.

This study utilized carshare data from GoGet members in Sydney, NSW, Australia. Characterization of not-at-fault drivers also provided useful insights for the quasi-induced exposure method. As recognized earlier that though this study only uses carshare users, focussing on the driver demographics of not-at-fault drivers provides useful insights about the quasi-induced exposure method. Statistically significant differences were also observed between not-at-fault drivers and the carshare member population in the case of, “Kilometres driven last year”, “How Often You Need a Car”, “Vehicle Ownership”, “Live near a Dedicated Parking”, “Main Mode of Transport”, “Kilometer driven last year” and “Duration of holding a driver’s license”. Therefore, caution needs to be undertaken when using the quasi-induced exposure method. As part of future studies, the robustness of these findings, which are presently based on the limited dataset need to be further explored with larger datasets having more number of not-at-fault crashes, which will help further explore the question of randomness of not-at-fault drivers. Further, larger datasets would also provide an opportunity to utilize ‘clean’ datasets (as defined by: Stamatiadis and Deaco, 1997; Jiang and Lyles, 2007; Chin and Haque, 2010; Haque et al., 2012) to further evaluate the robustness of these findings. However, another key factor that needs to be further explored is that even if the random sample assumption does not work in sub-populations, do not-at-fault drivers represent the overall driving population?

It was found that driving less in the previous year reduces propensity to crash for carshare users. This is in line with the direction of relationship between kilometres driven and crash propensity from previous studies (Yan et al., 2005; Chandraratna and Stamatiadis, 2009) that found a positive relationship which they exploited and used for exposure. However, the fact that not-at-fault drivers were found to have statistically higher kilometres driven last year suggests a non-linear relationship between kilometres driven and likelihood of crash. This has been found by Janke (1991) as well, and therefore there needs to be a sense of caution when using this variable as an exposure.

The duration of holding a driver’s license which is highly correlated with age and the years of driving experience, was found
to reduce the likelihood of being involved in a crash. In addition, higher insurance excess in case of damage due to crash also increases the propensity to crash.

 Owning one or more car and less frequently using a car were also found to reduce the likelihood of being involved in a crash. However, use of bicycle as a main mode of travel was found to increase likelihood of being involved in a crash. Though no systematic differences were observed with respect to kilometres travelled last year across these demographics, it is suspected that these systematic differences in crash risk are associated with exposure biases associated to time of day and road usage (Janke, 1991; Lardelli-Claret et al., 2003).

 Finally, though individuals involved in accidents in the past ten years were found to be highly likely to be involved in crashes, there was no evidence that they are more likely to be at-fault when involved in a crash. It was also found that given a crash occurs, carshare users not needing a car on the weekend, driving less than 1000km, rarely using a car and having and Australian license increases the likelihood of being at-fault.

 Based on this study it is possible to reduce crash propensity by having higher insurance excess for drivers highly prone to accidents, such as those with previous instances of accidents4 and those who have no car.5 Additionally, providing some incentives with regard to reduction in excess based on training and experience (based on kilometres driven) could also improve safety and reduce costs associated to crashes for carshare systems.

 For future research, driving cycle and attributes of drivers as well as trip purposes will be extracted from the GPS data associated with carshare members. Consequently, for each driver, average driving speed, geometry attributes of road on which he/she drives and built environment attributes of areas in which the driver mainly drives can be estimated and used in the risk exposure estimation. When more data of multiple years becomes available, temporal analysis can be conducted to examine how accident likelihood evolves over time as carshare systems are being more recognized.

 Acknowledgements

 The authors would like to thank Mr. Bruce Jeffreys and Ms. Rachel Moore from GoGet for providing data and enormous support in helping us understand it. We greatly appreciate the comments and advice from the three anonymous reviewers, which helped us significantly improve the paper. We would also like to thank the Australian Research Council for their support under Linkage Grant # LP130109983.

 References

 Firnkon, J., Muller, M., 2011. What will be the environmental effects of new free-floating car-sharing systems? The case of car2go in Ulm. Ecological Economy 70, 1519–1528.
 Jiang, X., Qiu, Y., Lyles, R., 2011. U.S. national household travel survey used to validate the exposure estimates by the quasi-induced exposure technique. Transportation Research Record 2237, 152–159.