University of Michigan, Department of Statistics

Applied QR 2020

Chicago Traffic Crashes: Comparing the Safety of Passenger Cars to SUVs and Pickup Trucks

Simon Fontaine

University of Michigan, simfont@umich.edu

Contents

1 Introduction

While larger vehicles are safer for their occupants than smaller ones [University at Buffalo, 2013], some studies show that they are more dangerous to the occupants of other cars. Ross and Wenzel [2002, Figure 2] find that SUVs and pickup trucks have similar risks to their drivers as passenger cars while exhibiting increased risk to the drivers of other cars; the difference is particularly large for pickup trucks. White [2004] show that passenger car occupants face an increased likelihood of fatality or serious injury when involved in a crash where the other vehicle is a *light truck*, that is, a SUV, a pickup truck or a van. These results induce a *collective action problem* in the sense that each individual is better off having a light truck for their own safety, but the population, taken as a whole, might be less safe with higher proportion of light trucks. Indeed, White [2004] further calculate that "for each fatal crash that occupants of large vehicles avoid, at least 4.3 additional fatal crashes involving others occur."

These two studies consider US-wide crash data and thus consider both urban and rural traffic crashes. In this analysis, we focus on the urban setting: in particular, we investigate if similar results holds for Chicago traffic crashes between 2015 to 2020.

Additionally, Ross and Wenzel [2002] does not control for any crash-, vehicle- or driver-related information and White [2004] control mostly for driver-related information. We include more details about the circumstances of each crash in our analysis to control for sampling biases such as different driving behaviors between smaller and larger car drivers [National Highway Traffic Safety Administration, 1998, Gladwell, 2004].

It is also worthwhile to mention that White [2004] consider the worst injury among the occupants: larger vehicles contain more occupants on average[1] so it increases the likelihood of observing more extreme injuries. In this analysis, we consider only drivers' injuries in order to avoid this over-sampling.

1.1 Data Overview

The data set we consider comes from The City of Chicago [2020] and contains over 400,000 reports of traffic crashes in Chicago between 2015 and 2020. For each crash, the data set provides information about environmental circumstance (time, location, road layout and condition, weather, visibility, etc.) and crash causes. Additionally, each vehicle involved in the crash has in entry in a related data set describing the vehicle itself, its status and damages resulting from the crash, its number of occupants along with numerous other information. Finally, each individual involved in a crash is described in a third data set containing, among others, demographic information, vehicle occupancy, injuries suffered and blood alcohol content testing. Furthermore, individuals identified as drivers have additional information regarding their driving actions prior to the crash.

1.2 Research Questions

Our main research question is to study the frequency of injuries suffered by drivers of passenger cars or light trucks (SUVs and pickup trucks) when involved in a crash with another car of these categories.

[1] In our sample, we find that passenger cars contain 1.24 occupants on average while SUVs and pickup trucks contain 1.28 occupants on average.

Injury Frequency by Vehicle Type

Vehicle	Passenger				SUV/Pickup			
Vehicle (other)	Passenger		SUV/Pickup		Passenger		SUV/Pickup	
Injury classification	Count	% Col.	Count	% Col.	Count	% Col.	Count	% Col.
Fatal	7	0.01	3	0.03	3	0.03	1	0.02
Incapacitating	238	0.47	43	0.41	62	0.59	22	0.44
Severe	245	0.48	46	0.44	65	0.62	23	0.46
Non-incapacitated	1277	2.52	254	2.43	271	2.60	131	2.65
Reported, not evident	937	1.85	188	1.80	192	1.84	86	1.74
Injury	2459	4.84	488	4.67	528	5.06	240	4.85
No Injury	48303	95.16	9955	95.33	9915	94.94	4710	95.15
Total	50762	100.00	10443	100.00	10443	100.00	4950	100.00

Table 1: *Frequency of reported injuries by occupying vehicle type and other vehicle type.*

In particular, we wish to compare the injury prevalence in the natural four cases of two-car incidents: passenger car v. passenger car, passenger car v. light truck, light truck v. passenger car and light truck v. light truck. We will consider two types of outcome: *Any injury* in the case some injury was reported and *Severe injury* for incapacitating injuries and fatalities. We do not consider fatalities directly as the data set contains too few occurrences.

Table 1 shows the frequencies of injuries for each combination of vehicle type (details on the sample are in Section 2). At first glance, it seems that there is no clear difference in the relative frequencies with respect to the type of vehicle involved in a crash. However, as [White, 2004] shows, there are multiple factors associated with the injury outcome of a crash. Hence, to achieve an accurate comparison of injury frequency by vehicle type, we must control these factors. This naturally leads to our secondary research question: determining which information about crashes, vehicles and drivers are good predictors of the injury outcome. As well as improving the quality of our main analysis, this process can also provide some insight on other important predictive factors of crash-related injuries.

2 Methodology

2.1 Data Pre-processing

The first step in our data processing pipeline is to identify crashes relevant to our research question. We consider crashes satisfying the following conditions:

- The crash report identifies that exactly two vehicles were involved;
- The vehicle report contains exactly two vehicle attached to those crashes;
- Both vehicles are either a passenger car or a light truck (SUV or pickup truck);
- Both vehicles contains at least one occupant, one of which is a driver.

Upon merging vehicle and driver information, we collapse the two entries of a given crash (one for each vehicle-driver report) into a single report by randomly assigning each vehicle-driver report to position 1 or position 2. Position 1 will correspond to our point-of-view, that is, the driver whose injuries we try to model; the vehicle-driver in position 2 will thus only intervene through its covariates. Without ambiguity, we refer to covaraites of vehicles and drivers in position 2 with the mention *(other)* and those in position 1 without any mention.

Then, we exclude some reports based on inspection of their values. For example, the reported speed limits and lane counts are sometimes out of the ordinary: we drop any observation with speed limit of lane count values that occur less than 50 times in the data set. Similarly, some reports contains drivers' age below 15: while this is a real possibility (e.g. stolen car), we choose to remove those reports. After selection of crashes and removal of some reports, we obtain a set of 38,299 crashes.

Most of the potential covariates included in the reports are categorical and a large proportion of these contain large amounts of possible values. To avoid having too many levels per variable, we merge low-frequency values (less than a few hundred observations) into a single category. Additionnally, missing values or entries reported as some equivalent of missing or unknown are also assigned to that cemetery category that is systematically called *Unknown/other*. Among others variables, *Primary cause*, *Sex*, *Safety equipment*, *Action and Maneuver* are treated as such.

The data set contains also a few numerical variables. Since we suspect potential non-linear relationship between *Month*, *Day of week*, *Hour*, *Age* and injuries, we calculate new variables by binning their values into categories. The *Month* variable is binned into *Season*; the *Day of week* variable into *Weekday/Weekend*; *Hour* into *Day/Night* and *Age* into three categories (20-, 21-60, 60+). The *Age* variable also contained some missing values which are labeled as *Unknown* in the binned age variable.

Finally, we compute our two response variable using the *Injury classification* of the driver in position 1. First, whenever some injury was reported, the *Any injury* response is set to 1; it is set to 0 otherwise. Second, whenever the driver was incapacitated or killed, we set the *Severe injury* variable to 1 and to 0 otherwise.

A detailed summary of all variables used in this analysis can be found in Table 4.

2.2 Preliminary Variable Selection

The first step in the analysis it to select control variates for the main analysis of injuries by vehicle types. To achieve this, we consider three different opinions.

First, the study by White [2004] identifies some factors relevant for predicting injuries. Since the setting is fairly similar and we have access to relatively equivalent information, we will include any feature they found significant.

Second, we will consider a naive Random Forest classifier model for each response including all available information. Then, we compute the variable importance of all covariates to identify a collection of features exhibiting predictive power.

Third, since the Random Forest model is intrinsically non-linear, we also consider a penalized logistic regression model using all variables. This model enables us to identify features exhibiting linear predictive power. We choose an elastic net penalty (95% ℓ_1, 5% ℓ_2) with a regularization parameter inducing some sparsity. We compute a variable importance statistic as follows. Given an estimate $\beta_j \in \mathbb{R}^{d_j}$ for some feature j (d_j denotes the number of coefficients associated with that feature), we compute the normalized group norm $\|\beta_j/d_j\|_2$ and standardize these norms so that they sum to 1 across all groups of features.

Finally, we arbitrarily select potentially relevant features from these three subsets as our control variates. Note that, for the last two methods, we only consider information which can temporally predict the injuries and exclude any information that can be seen as resulting from a crash (e.g. vehicle damage).

2.3 Logistic Regression Analysis of Injuries

With our set of control variates in hand, we proceed to a logistic regression model for both responses. We include the main effect of all control variates as well as the two-way interaction between vehicle types. For numerical control variates (*Month*, *Day of week* and *Hour*), we consider three modeling variations: either we fit cubic splines with 5 degrees of freedom, either we use the binned version of these variables or we use it as a categorical variable directly if the number of levels is small. Note that the *Age* variable contained some missing values which prevents us from using splines, so we resort to using the binned variable only. To identify a final model for each response, we compare the AIC statistic and select the lowest value.

To answer our main research question, we estimate three contrasts for each response. Our reference value will be crashes with two passenger cars and the three contrasts are given by varying vehicle types. We then examine the estimated coefficients for the control variates to study their association with injuries.

3 Results

All `Python` code for data processing, analysis, producing tables and figures can be found at:

https://github.com/fontaine618/crime.

3.1 Preliminary Variable Selection

White [2004] results. The author finds that "occupants of v1 are more likely to be killed or seriously injured in crashes that occur at night or on weekends, in cities rather than rural areas, when either vehicle's speed was more than 10 miles per hour above the limit, when the driver of v1 was more than 60 years old or the driver of v2 was male, and when v1 contained more occupants. Occupants of v1 are less likely to be killed or seriously injured if the driver of v1 wore a seat belt." They also find some mild effect from driving under the influence of drugs or alcohol. From this, we identify that *Hour*, *Day of week*, *Action*, *Age*, *Sex, Safety equipment, Blood alcohol content* and *Occupant count* are potential control variates.

Random Forest. Figure 3.1 (in red) contains the variable importance of all available information for both of our responses. Using a threshold of 0.05 for at least one response, we find that *Age*, *Action*, *Maneuver*, *Primary cause* and *Travel direction* are have strong non-linear predictive power.

Penalized Logistic Regression. Figure 3.1 (in blue) contains the standardized and normalized group norms of the estimated coefficient where the regularization parameter is set to 0.001. We find that *Blood alcohol content*, *Action*, *Vision*, *Maneuver*, *Safety equipment*, *Sex* and *Weather condition* all exhibit strong linear effects on the responses.

Comparison. Since the logistic regression model only uses main linear effect, it is not surprising that the numerical variables *Age*, *Month*, *Day of week* and *Hour*, which intuitively should not act linearly, were not selected. Therefore, since [White, 2004] and the non-linear model finds them relevant, we include these covariates, but model them using non-linear terms (cubic spline and binning). All three agents find *Action* to be important and our two models identify *Maneuver* as important, so we add them to our pool of features. *Occupant count* was only considered useful by White [2004] so we exclude it. Now, we choose to include *Primary cause* as well because of its clear interpretability as well as being selected by the Random Forest model. Finally, since the logistic regression model indicates particularly strong effects coming from Blood alcohol content and White [2004] indicates that they are associated with driver's injuries, we include those two features to our set of control variates. In total, we include ten control variates. For crash-related information, we have *Month*, *Day of week*, *Hour* and *Primary cause*. For vehicle-related information, we have *Action* and *Maneuver*. For driver-related information, we have *Age*, *Sex*, *Blood alcohol content* and *Safety equipement*.

3.2 Model Selection

Our three numerical control variates, *Month*, *Day of week* and *Hour*, require non-linear modeling. The *Month* variable is modeled using either cubic splines (GAM) or binned into seasons; the *Day of week*

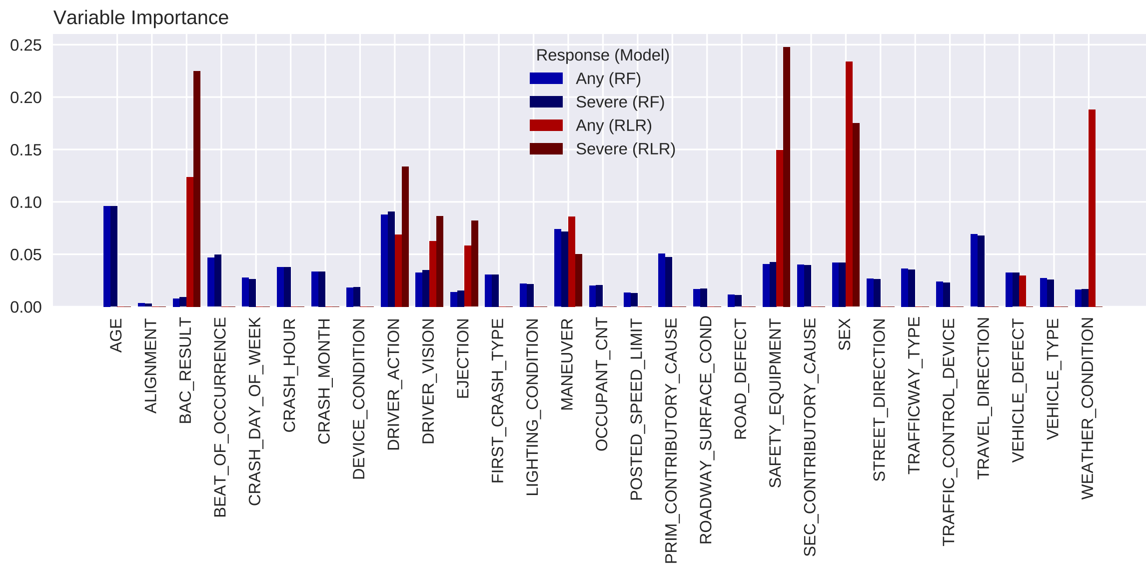

Figure 3.1: *Variable importance using a Random Forest classifier (RF) and regularized logistic regression (RLR). For logistic regression, the variable importance is calculated as the dimension-normalized group norm of the estimates standardized to sum to 1.*

variable is modeled using all seven days as categories (*All*) or binned into *Weekday*/*Weekend*; the *Hour* variable is modeled using either cubic splines (GAM) or binned into *Day*/*Night*. Table 2 shows modeling results for all such combinations and for both responses.

In all cases, we do not observe a particularly large difference between models which indicates that either these covariates have small association with the responses or that these modeling schemes are relatively equivalent. For both responses, binning these numerical variables results in the best log-likelihood, AIC and BIC so we choose these models for the analysis.

3.3 Comparison of Injuries by Vehicle Types

Using the logistic regression models selected in the previous section, we compute contrasts between all combinations of pairs of vehicle type: results are displayed in Table 3.

We find that there is no significant difference in suffering any injury no matter what are the two vehicles involved in the crash. All estimated differences are small in size and do not show any association between injuries and vehicle type.

In the case of severe injuries, we find a significant decrease in log-odds of suffering a sever injury when a driver facing a passenger car occupies a light truck instead of a passenger car. All other comparisons do not yield significant differences, but some estimated differences are generally negative when compared to the two passenger cars case:

- When the driver occupies a passenger car, we observe an insignificant decrease in log-odds of severe injury when the other vehicle changes from a passenger car to a light truck.
- When facing a passenger car, a driver's log-odds of severe injury decreases insignificantly when

Logistic Regression Models						
Month	Day of Week	Hour	Log-lik.	D.f.	AIC	BIC
Any injury						
GAM	All	GAM	-7155	72	14456	-389096
		Dark	-7156	69	14453	-389125
	Weekday	GAM	-7157	67	14449	-389145
		Dark	-7158	64	14446	-389174
Season	All	GAM	-7157	71	14459	-389101
		Dark	-7159	68	14456	-389130
	Weekday	GAM	-7159	66	14452	-389151
		Dark	-7161	63	14449	-389180
Severe injury						
GAM	All	GAM	-1098	72	2342	-401210
		Dark	-1099	69	2337	-401240
	Weekday	GAM	-1101	67	2337	-401257
		Dark	-1101	64	2332	-401288
Season	All	GAM	-1099	71	2342	-401218
		Dark	-1100	68	2337	-401249
	Weekday	GAM	-1102	66	2337	-401266
		Dark	-1102	63	2332	-401297

Table 2: *Estimated contrasts between pairs of vehicule types.*

conducting a light truck instead of a passenger car.

- When the two vehicles are light trucks, we observe an insignificant decrease in log-odds of sever injury for one of two the drivers when compared to when both vehicules are passenger cars.

3.4 Control Variates Analysis

Crash-related information. The estimated coefficient for both responses can be found in Table 5. We find no variable significantly associated with increased or decreased log-odds of injury or severe injury.

Vehicle-related information. The estimated coefficient for both responses can be found in Table 6. We find that the *Action* variable is associated with observing any injury. In particular, all actions, except *Too fast for conditions*, show increased log-odds of injury when compared to *Disregarded control device*. This imply that *Too fast for conditions* and *Disregarded control device*, performed by any of the two drivers, leads to crashes with less frequent injuries than all other actions.

Crash-related information. The estimated coefficient for both responses can be found in Table 7.

We find that the *Unknown/other* age group of the driver is associated with increased log-odds of any and severe injury when compared to all three other age groups. All age groups other than 20- for

Logistic Regression Vehicule Type Contrasts							
						95 % C.I.	
Vehicle	Vehicle (other)	Est.	Std err.	z-value	p-value	Lower	Upper
Any injury							
Passenger	Passenger	0.000	–	–	–	–	–
SUV/Pickup	Passenger	-0.076	0.071	-1.075	0.282	-0.215	0.063
Passenger	SUV/Pickup	-0.027	0.071	-0.380	0.704	-0.167	0.113
SUV/Pickup	SUV/Pickup	-0.009	0.099	-0.089	0.929	-0.204	0.186
Severe injury							
Passenger	Passenger	0.000	–	–	–	–	–
SUV/Pickup	Passenger	-0.583	0.194	-3.006	0.003	-0.963	-0.203
Passenger	SUV/Pickup	-0.273	0.215	-1.270	0.204	-0.695	0.148
SUV/Pickup	SUV/Pickup	-0.449	0.279	-1.608	0.108	-0.997	0.098

Table 3: *Estimated contrasts between pairs of vehicule types.*

the other driver show increased log-odds of any injury and only the *Unknown/other* age group show increased log-odds of sever injury.

Interestingly, and reminiscent of the findings of White [2004], when any of the two drivers was offered a Blood alcohol content test (i.e., was suspected of driving under the influence), we find decreased log-odds of any and sever injury.

We find that female drivers are associated with decreased log-odds of any injure when compared to male and unknown/other drivers; we find similar observations for the other driver as well.

4 Discussion

4.1 On the Analysis

Among the set of control variates, we find that *Primary cause*, *Action* and *Maneuver* are highly correlated both statistically and intuitively. Including all three variables in our main analysis potentially decreases the power of the logistic regression model. Indeed, we find that only *Action* had a significant association with injuries and excluding the other two could decrease the uncertainty around all estimates.

When computing confidence intervals for contrasts and for all control variates, we use unadjusted 95% multipliers. This obviously leads to multiple testing issues as the family-wise Type I error is no longer of level 95%. Applying a Bonferroni correction to the three estimated contrasts would lower a 5% significance level to 1.33% which is still above the observed p-value for the only contrast identified as significant.

4.2 On the Results

In the Chicago urban setting, we find evidence showing light trucks (SUVs and pickup trucks) drivers have lower odds of suffering severe injuries when compared to passenger car drivers when involved in a crash a passenger car. This result seem to agree with studies [University at Buffalo, 2013] and intuition that light truck are safer than passenger cars.

We have also found some weaker evidence that drivers involved in two-light-trucks crashes are less likely to suffer severe injuries than drivers involved in two-passenger-cars crashes. In particular, it does not seem likely that the opposite is true, at least under the situations covered by this analysis. On the surface, this result seem to disagree with those found in Ross and Wenzel [2002] and White [2004]. However, it is important to reiterate that we only consider urban crashes where the conditions are typically different than rural crashes (e.g. different speeds, traffic way types, etc.). Indeed, White [2004] find a significant difference in severe injuries between rural crashes and large city crashes.

As for the association between injuries and the control variates we included in the models, we obtain results mostly agreeing with those in White [2004]. We find that age acts in a similar fashion as younger drivers are associated with fewer injuries. We find the same unintuitive result that drivers suspected of being under the influence of alcohol tend to suffer fewer injuries. We find that male drivers are associated with increased of injury while White [2004] find that to be true only for the other driver. We do not find associations between time (month, day of week, hour) and injuries.

4.3 Further comments

This would not be in the report, but here are some things I would change:

- Some sort of variable selection in the main logistic regression (step-wise possibly) to remove spurious variables
- There is a few predictors that I forgot to include and didn't want to redo the analysis (e.g. exceeds speed limit)
- The variable description table does not include all variables, only those selected. All considered for selection should appear there
- The variable importance plot has the variable codes: I should decode the names into legible values
- Ejection should not be part of the selection process as it is a response variable;
- I should actually exclude Month from the control variates based on the three agents.
- There are many categories and I binned some partly, but observing the estimates indicates I should bin more since they seem to have only 1-3 different values.

References

M. Gladwell. Big and bad. *Gladwell.com*, 2004. URL https://web.archive.org/web/20160219222512/http://gladwell.com/big-and-bad/.

National Highway Traffic Safety Administration. Overview of vehicle compatibility. 1998. URL http://www.nhtsa.dot.gov/cars/problems/studies/LTV/.

M. Ross and T. Wenzel. An analysis of traffic deaths by vehicle type and model. 2002. URL https://www2.lbl.gov/Science-Articles/Archive/assets/images/2002/Aug-26-2002/SUV-report.pdf.

The City of Chicago. Traffic crashes. Online Data set, 2020. URL https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if?fbclid=IwAR25N2kCthVSdzUEz1lq3VJN3diGt1yYZ2pkPcxqvq_ulJfFRw3Fey9XuIg.

University at Buffalo. Passenger car drivers are more likely to die in crashes with SUVs, regardless of crash ratings. *Science Daily*, 2013. URL www.sciencedaily.com/releases/2013/05/130514135417.htm.

M. White. The “arms race” on american roads: The effect of sport utility vehicles and pickup trucks on traffic safety. *The Journal of Law and Economics*, 47(2):333–355, 2004. doi: 10.1086/422979. URL https://doi.org/10.1086/422979.

A Supplemental Tables

Variable Description

Variable	Values (Count, count (other))	Comments
Crash		
Primary cause	Unknown/other (17229)	Missing and low frequency
	Following too closely (6294)	levels assigned to
	Failing to yield right-of-way (5899)	"Unknown/other"
	Improper overtaking/passing (2158)	
	Improper lane usage (1855)	
	Failing to reduce speed to avoid crash (1766)	
	Improper turning/no signal (1625)	
	Improper backing (1473)	
Month	1-12	
Hour	0-23	
Day of week	Sunday-Saturday	
Season	Fall (11005)	
	Spring (9205)	
	Summer (9113)	
	Winter (8976)	
Weekday	Weekday (28178)	"Weekend" defined as
	Weekend (10121)	Saturday and Sunday
Dark	Day (32937)	"Day" defined between 6AM
	Night (5362)	and 8PM
Any injury	False (36434)	True if not "No indication of
	True (1865)	injury"
Severe injury	False (38114)	True if "Fatal" or
	True (185)	"Incapacitating injury"
Drivers		
Sex	M (20915, 20972)	
	F (13890, 13816)	
	Unknown/other (3494, 3511)	
Safety equipment	Safety belt used (20358, 20268)	Missing and low frequency
	Unknown/other (17097, 17141)	levels assigned to
	Not used/none present (844, 890)	"Unknown/other"
Blood alcohol content	Test not offered (37785, 37730)	All but "Test not offered" are
	Test offered (514, 569)	assigned to "Test offered"
Age (binned)	21-60 (23684, 23574)	
	Unknown (9614, 9703)	
	60+ (3531, 3485)	
	20- (1470, 1537)	
Vehicle		
Type	Passenger (30617, 30588)	
	SUV/Pickup (7682, 7711)	
Action	None (14375, 14494)	Missing and low frequency
	Unknown/other (11952, 11908)	levels assigned to
	Failed to yield (3829, 3756)	"Unknown/other"
	Followed too closely (2698, 2687)	
	Improper backing (1282, 1255)	
	Improper lane change (1153, 1087)	
	Improper turn (1026, 1017)	
	Improper passing (856, 918)	
	Too fast for conditions (607, 606)	
	Disregarded control devices (521, 571)	
Maneuver	Straight ahead (21904, 21939)	Missing and low frequency
	Slow/stop in traffic (4316, 4372)	levels assigned to
	Unknown/other (4152, 4126)	"Unknown/other"
	Turning left (2784, 2783)	
	Backing (1488, 1470)	
	Turning right (1326, 1296)	
	Changing lanes (1254, 1218)	
	Passing/overtaking (1075, 1095)	

Table 4: *Estimated contrasts between pairs of vehicule types.*

Logistic Regression: Crash Control Variates 95 % C.I.				
	Any injury		Severe injury	
Term	Lower	Upper	Lower	Upper
Dark (ref. Day)				
Night	-0.11	0.16	-0.38	0.48
Season (ref. Fall)				
Spring	-0.11	0.15	-0.58	0.26
Summer	-0.20	0.06	-0.67	0.13
Winter	-0.08	0.18	-0.64	0.17
Weekday (ref. Weekday)				
Weekend	-0.17	0.05	-0.35	0.31
Primary cause (ref. None)				
Failing to yield right-of-way	-0.14	0.35	-0.39	1.11
Following too closely	-0.16	0.32	-0.61	0.83
Improper backing	-0.03	0.75	-0.29	2.10
Improper lane usage	-0.01	0.64	-1.14	0.61
Improper overtaking/passing	-0.33	0.29	-0.52	1.50
Improper turning/no signal	-0.26	0.37	-0.92	0.94
Unknown/other	-0.10	0.35	-0.40	0.94

Table 5: *Estimated contrasts between pairs of vehicule types.*

Logistic Regression: Vehicle Control Variates Estimates				
	Any injury		Severe injury	
Term	Lower	Upper	Lower	Upper
Action (ref. Disregarded control devices)				
Failed to yield	0.23	0.96	-0.55	1.66
Followed too closely	0.50	1.29	-0.33	2.09
Improper backing	0.86	1.90	-0.43	2.58
Improper lane change	0.47	1.41	-0.96	1.59
Improper passing	0.31	1.32	-0.67	2.75
Improper turn	0.08	0.96	-0.96	1.69
None	0.02	0.69	-0.39	1.66
Too fast for conditions	-0.68	0.18	-2.14	0.15
Unknown/other	0.02	0.69	-0.83	1.20
Action (other) (ref. Disregarded control devices)				
Failed to yield	0.28	0.98	-0.49	1.76
Followed too closely	0.59	1.36	-0.53	1.81
Improper backing	0.53	1.45	-0.81	1.84
Improper lane change	0.40	1.32	-1.22	1.26
Improper passing	0.62	1.69	-0.61	2.81
Improper turn	0.31	1.21	-0.79	2.01
None	0.07	0.71	-0.58	1.47
Too fast for conditions	-0.48	0.38	-1.88	0.51
Unknown/other	0.10	0.74	-0.66	1.39
Maneuver (ref. Backing)				
Changing lanes	-0.49	0.27	-0.75	1.55
Passing/overtaking	-0.29	0.55	-0.85	1.74
Slow/stop in traffic	-0.16	0.46	-0.06	1.84
Straight ahead	-0.22	0.35	-0.53	1.07
Turning left	-0.20	0.47	-0.53	1.42
Turning right	-0.24	0.55	-0.77	1.41
Unknown/other	-0.38	0.24	-0.78	0.96
Maneuver (other) (ref. Backing)				
Changing lanes	-0.41	0.38	-0.31	2.24
Passing/overtaking	-0.28	0.56	-1.01	1.30
Slow/stop in traffic	-0.21	0.43	-0.35	1.42
Straight ahead	-0.29	0.28	-0.40	1.17
Turning left	-0.48	0.18	-0.30	1.69
Turning right	-0.39	0.39	-1.16	0.86
Unknown/other	-0.44	0.18	-0.54	1.19

Table 6: *Estimated contrasts between pairs of vehicule types.*

Logistic Regression: Driver Control Variates 95 % C.I.				
	Any injury		Severe injury	
Term	Lower	Upper	Lower	Upper
Age (binned) (ref. 20-)				
21-60	-0.14	0.30	-0.84	0.60
60+	-0.12	0.40	-0.98	0.68
Unknown/other	0.61	1.13	0.32	2.06
Age (binned) (other) (ref. 20-)				
21-60	0.07	0.48	-0.45	0.86
60+	0.10	0.60	-0.90	0.60
Unknown/other	0.75	1.23	0.16	1.69
Blood alcohol content (ref. Test not offered)				
Test offered	-0.79	-0.15	-1.82	-0.44
Blood alcohol content (other) (ref. Test not offered)				
Test offered	-0.71	-0.08	-1.78	-0.48
Safety equipment (ref. Not used/none present)				
Safety belt used	-0.38	0.27	-0.50	1.32
Unknown/other	-0.61	0.05	-1.06	0.76
Safety equipment (other) (ref. Not used/none present)				
Safety belt used	-0.34	0.29	-0.36	1.18
Unknown/other	-0.61	0.03	-0.33	1.25
Sex (ref. F)				
M	0.07	0.27	-0.38	0.24
Unknown/other	0.20	0.71	-1.21	0.13
Sex (other) (ref. F)				
M	0.08	0.27	-0.18	0.42
Unknown/other	0.01	0.49	-0.82	0.65

Table 7: *Estimated contrasts between pairs of vehicule types.*