PAPER 2021–2017

DRIVING BELOW TRAFFIC FLOW

SAS® Global Forum Student Symposium 2017

Team Flow Riders

Zhongjie Cai, MS
Student, PhD in Biostatistics
zhongjic@usc.edu

Khang Chau, MS
Student, PhD in Biostatistics
khang.chau@usc.edu

Dixin Shen, MS
Student, MS in Biostatistics
dixin.shen@usc.edu

Faculty Advisor:
Meredith Franklin, PhD
Assistant Professor
323–442–2703

Department of Preventive Medicine
Keck School of Medicine
University of Southern California
Los Angeles, CA

Introduction

Drivers in the United States are often instructed in driving schools to follow the flow of traffic when driving. Various studies have shown positive association between driving too fast and the risk of being involved in an accident. But what about driving too slowly? According to the California Driver Handbook, "if you block the normal and reasonable movement of traffic by driving too slowly, you may be cited."[1] Clearly driving too slowly is not recommended. However, is driving below traffic flow that problematic?

One study in 1964 concluded that "the greater the driver's variation from this average speed, the greater his chance of being involved in an accident,"[2] while another in 2001 found that "travelling speeds below the mean traffic speed were associated with a lower risk of being involved in a casualty crash."[3] These conflicting findings only increase our confusion. To the best of our knowledge, existing research has primarily been to examine whether fatal crashes are associated with driving too fast. Therefore this study serves as a needed complement of what has been previously examined.

Problem

Although slow driving is often proscribed, its association with risk of dying in a severe accident (defined as an accident where at least one person died) remains unknown. We define "slow driving" as driving at least 15% below the speed of average traffic flow. For comparison, we also define "fast driving" as driving at least 15% above traffic flow. In this study, we evaluate the association between the risk of dying when involved in severe accidents and driving too slowly right before the accident occurs, and adjust for potential confounders.

Methods

Data

We obtained data from the Fatality Analysis Reporting System (FARS)[4] which contain a census of fatal motor vehicle crashes nationwide within the 50 U.S. states, the District of Columbia, and Puerto Rico from 1975 to 2015. Observations in these datasets contain information about motor vehicle accidents happening on public traffic ways that resulted in the death of one or more persons within 30 days of the accident. The data are in two parts:

(1) A set of three main datasets that contain variables related to the accidents, the people, and the vehicles involved.

(2) Auxiliary datasets that simplifies several categorical variables in the three main datasets in (1).

For example, in 2004 air bag usage had 13 categories in the main datasets (1) but only 3 categories (deployed, not deployed, and not applicable) in the auxiliary datasets (2), which simplifies the data structure. In this study we limited our scope to records from 2006 to 2015 for the state of California.

To determine whether the driver was driving slower or faster than traffic flow, we acquired traffic speed data that the California Department of Transportation (Caltrans) collects using their speed monitoring stations (SMS) on most interstate and U.S. highways.[5] These stations report the average speed of traffic in 5-minute intervals daily. To reduce the tremendous size of the datasets without sacrificing too much accuracy, we took the following measures:

(1) Keep speed records in 15-minute intervals, i.e., reducing from 12 to 4 speed records per hour per station per day.
(2) Use the speed data on the 5th, 15th, and 25th day of each month in 2015 and calculate monthly average speed for each 15-minute interval.

FARS records the highways where the accidents occur (e.g., I-5, US-101, etc.) but not the direction of the highway, so we obtained the line shapefiles of the highways from Caltrans.[6]

We used ArcMap and R to match each accident to the appropriate SMS, which is the nearest station on the same side of the highways. We loaded the highway shapefile in ArcMap as well as the FARS accidents and speed stations using their coordinates (**Figure 1**). We first matched each accident to the nearest highway to incorporate the road direction to each accident, using the Join tool in ArcMap. We identified a few cases where the accidents were matched to the wrong highway, often near overpasses and underpasses where the highway lines intersect, and manually matched these to their stations.

In some cases, an accident was spatially closer to a station on the wrong highway or the wrong side of the highway. To overcome this, we performed three matching processes:

(1) We matched each accident to the nearest station using the Join tool (i.e., there are stations that are not matched). We then filtered out cases where the accidents were matched to stations on the wrong highway or wrong side of the highway.
(2) We matched each station to the nearest accident using the Join tool (i.e., there are accidents that are not matched). We then filtered out cases where the highway or the side of the highway were incorrectly matched. We then kept only the stations with the shortest distances to the same accidents.
(3) Combining and removing duplicates from the above two steps covered about 76% of the accidents. We manually matched the rest of the accidents in ArcMap.

Finally, we matched the accidents with the speed data from the SMS's by station ID and month, hour, and minute of the hour of each accident. If speed data was not available in that month, we used speed data from the next closest month available. Due to their low frequency, we excluded vehicle that are vans (3), pick-up trucks (1), buses (13), and unknown type (74), as well as vehicles with no damage (41) and unknown damage (71). The following accidents were also excluded:

- Driverless and one-vehicle accidents
- Accidents with no injury severity (outcome variable)
- Accidents with one or more vehicles with no reported traveling speed
- Accidents with no coordinates
- Accidents with missing traffic speed data from monitoring stations

Statistical Methods

We explored 16 variables that might affect the association between severe accidents and driving speed. These were variables about the accidents, the vehicles (e.g., manufacturer, type, extent of damage, etc.), and the drivers (e.g., age, gender, alcohol use, etc.). Most variables were classified into 3 groups: yes, no, and unknown. Some variables such as the type of vehicle (e.g. sedan, light truck, motorcycle, etc.) were categorized into multiple groups. Age of driver and age of vehicle were ordered categories, so these were treated as continuous covariates in the model. As all covariates are categorical, we performed preliminary examinations via the Pearson's chi-square test of homogeneity as well as the Hantel-Maenszel test when appropriate.

To control for environmental factors such as road and weather conditions, we used a conditional logistic regression model, conditioning on the accidents. We fitted the univariate model for the association between driving speed (slow, with traffic flow, or fast) and severe accidents. In variable selection, we adjusted each covariate individually to the univariate model and kept those with significant changes ($> 15\%$) in the adjusted odds ratios (OR) for driving speed. We then checked for interaction among the significant covariates, using significance level $\alpha = 0.05$.

Spatial matching of accidents and SMS's were done using a combination of tools in ArcMap 10.4.1 and R 3.3.1. Further data manipulations and statistical analyses were performed in SAS 9.4, notably PROC LOGISTIC. Both ArcMap and SAS softwares were used under free student licenses. The programming language and software R is free.

Results

2673 vehicles (1131 unique accidents) were included in our final analysis, including 896 (33.5%) with fatal injuries and 1777 (66.5%) without fatal injuries. There were 772 (28.9%) slow drivers, 829 (31%) fast drivers, and 1072 (40.1%) traffic-flow drivers. For the drivers, we looked at age, gender, alcohol and drug use, prior involvement in other accidents, and prior suspension of driver's license. For the accidents, we looked at restrain system use (e.g., seat belts, etc.), airbag deployment, driver ejected from vehicle, fire occurance, initial contact point, and extent of damage. For the vehicles, we looked at vehicle manufacturer, type, age, and trailing unit (towing). These are summarized in **Table 1**.

In the univariate model stratified by accidents, the association between driving speed and the driver's mortality in severe accident was statistically significant ($p < 0.0001$). Individually adjusting for the covariates of interest, we found vehicle type, trailing unit, extent of damage, ejection, and alcohol use to be possible confounders, with adjusted OR changes greater than 15%. However, in the full model with all five covariates, trailing unit and ejection were no longer statistically significant, so we excluded them. None of the two-way interactions were statistically significant. The final conditional logistic model included three covariates: vehicle type, extent of damage, and alcohol use. In the full model, the association between severe accidents and driving was statistically significant ($p = 0.004$). See **Table 2** for the ORs corresponding with the covariates in the final model.

Discussion

In our final analysis, we find that slow drivers are 2.4 times as likely to die in a severe accident as traffic-flow drivers (OR = 2.41, 95% CI: 1.39, 4.18) and 1.26 times as likely as fast drivers (OR = 1.26, 95% CI: 0.67, 2.35), although only the first odds ratio is statistically significant. Among the included covariates, larger vehicles are less likely to have fatal injuries. For example, drivers in light trucks are 0.62 times as likely to die as sedan drivers (OR = 0.62, 95% CI: 0.39, 0.99), which is not surprising. Similarly, smaller extents of damage are associated with lower odds of having fatal injuries: drivers in damage-but-functional vehicles are 0.007 times as likely to have fatal injuries as drivers in vehicles disabled by damages (OR = 0.007, 95% CI: 0.001, 0.056). Lastly, drivers with reported alcohol use are 1.35 times as likely to die as those without (OR = 1.35, 95% CI: 0.76, 2.40). Although a large number of drivers recorded 'unknown' alcohol use, the high OR when comparing these to those with no alcohol use (15.18, 95% CI: 7.69, 29.98) suggested that some of these might have consumed alcohol without reporting, for obvious reasons.

Suggestions for Future Studies

One limitation of this study is that we used average traffic speed of only 36 days of year 2015 (the 5th, 15th, and 25th of each month) to determine if each vehicle was travelling slower than traffic flow. This limitation is largely due to the large quantity of data sets (individual datasets for each day for each of 10 districts over 10 years) as well as Caltrans' prohibition of automated data downloads. As Caltrans has traffic data dating back to 2002, future studies could incorporate more precise traffic data to more accurately determine slow driving.

Another limitation is that we used the average traffic speed of all lanes. Typically on highways, inner lanes are faster than outer lanes, hence the recommendation that slower vehicles stay to the right (outer) lanes. If in the future we are able to acquire the actual lane where the accident happens, and has more refined spatial data for the lanes on each highway, along with the corresponding speed data, then the analysis could be improved.

The small sample problem in our study is two-fold: (1) California currently only has speed monitoring stations on interstate and U.S. highways, while there are numerous state highways with similar speed limits yet without speed stations. Should the Department of Transportation begin monitoring state highways, a similar study can include accidents on state highways. (2) We examine FARS accidents only in California. Future studies can utilize highway speed monitoring stations in other states to evaluate similar hypotheses as the present study.

Due to its limited scope in only California, generalization of our results to other states is not always appropriate, especially for those with lower population density and lower volume of vehicles. Further, it is likely that people living in urban, suburban, and rural areas have different driving habits derived from how frequently they drive and the kind of traffic they often encounter on the road.

Conclusion

In our study, after adjusting for vehicle type, extent of vehicle damage, and alcohol use, we find that slow drivers (driving more than 15% below traffic flow speed) are more likely to die in a severe motor accidents compared to traffic-flow drivers, with statistical significance. Given our findings and existing research about the risk of driving too fast, our recommendation is that drivers follow traffic flow as much as possible.

Appendix

Table 1. Summary of Variables

Variable	n	**Fatal accident ($n = 896$)**	**Non-fatal accident ($n = 1777$)**	p**-value***
Driving speed				**< 0.0001**
Slow	772	179 (23%)	593 (77%)	
Fast	829	433 (52%)	396 (48%)	
With traffic flow	1072	284 (26%)	788 (74%)	
Gender				**$< 0.0001^{\dagger}$**
Male	2031	727 (36%)	1304 (64%)	
Female	635	169 (27%)	466 (73%)	
Unknown	7	0 (0%)	7 (100%)	
Age of driver (years)				$0.38^{\dagger}$
< 20	112	32 (29%)	80 (71%)	
20–29	697	249 (36%)	448 (64%)	
30–39	560	186 (33%)	374 (67%)	
40–49	560	163 (29%)	397 (71%)	
50–59	425	131 (31%)	294 (69%)	
$\geq$60	313	135 (43%)	178 (57%)	
Unknown	6	0 (0%)	6 (100%)	
Alcohol use				**< 0.0001**
Yes	292	143 (49%)	149 (51%)	
No	1842	314 (17%)	1528 (83%)	
Unknown	539	439 (81%)	100 (19%)	
Drug use				**< 0.0001**
Yes	54	27 (50%)	27 (50%)	
No	851	176 (21%)	675 (79%)	
Unknown	1768	896 (34%)	1777 (66%)	
Prior accident				**< 0.0001**
Yes	446	179 (40%)	267 (60%)	
No	2227	717 (32%)	1510 (68%)	

*Pearson's chi-square test of homogeneity, unless otherwise indicated.

†Mantel-Haenszel chi-square test

Table 1. Summary of Variables (cont.)

Variable	***n***	**Fatal accident ($n = 896$)**	**Non-fatal accident ($n = 1777$)**	***p*-value***
Prior suspension of license				**0.0012**
Yes	311	149 (48%)	162 (52%)	
No	2362	747 (32%)	1615 (68%)	
Restrain system use				**< 0.0001**
Yes	2296	621 (27%)	1675 (73%)	
No	183	150 (82%)	33 (18%)	
Unknown	194	125 (64%)	69 (36%)	
Airbag deployment				**< 0.0001**
Yes	814	302 (37%)	512 (63%)	
No	1812	564 (31%)	1248 (69%)	
Unknown	47	30 (64%)	17 (36%)	
Driver/passenger ejected				**< 0.0001**
Yes	112	103 (91%)	9 (9%)	
No	2306	558 (24%)	1748 (76%)	
Unknown	255	235 (92%)	20 (8%)	
Fire or explosion				**< 0.0001**†
None	2558	823 (32%)	1735 (68%)	
Fire in vehicle	113	71 (63%)	42 (37%)	
Fire in other vehicle	2	2 (100%)	0 (0%)	
Initial point of impact				**< 0.0001**
No collision	40	36 (90%)	4 (10%)	
Clock points	2262	698 (31%)	1564 (69%)	
Other direction	315	127 (40%)	188 (60%)	
Unknown	56	35 (63%)	21 (37%)	
Extent of vehicle damage				**< 0.0001**
Disabling damage	1654	825 (50%)	829 (50%)	
Functional damage	514	51 (10%)	463 (90%)	
Other/minor damage	505	20 (4%)	485 (96%)	

*Pearson's chi-square test of homogeneity, unless otherwise indicated

†Mantel-Haenszel chi-square test

Table 1. Summary of Variables (cont.)

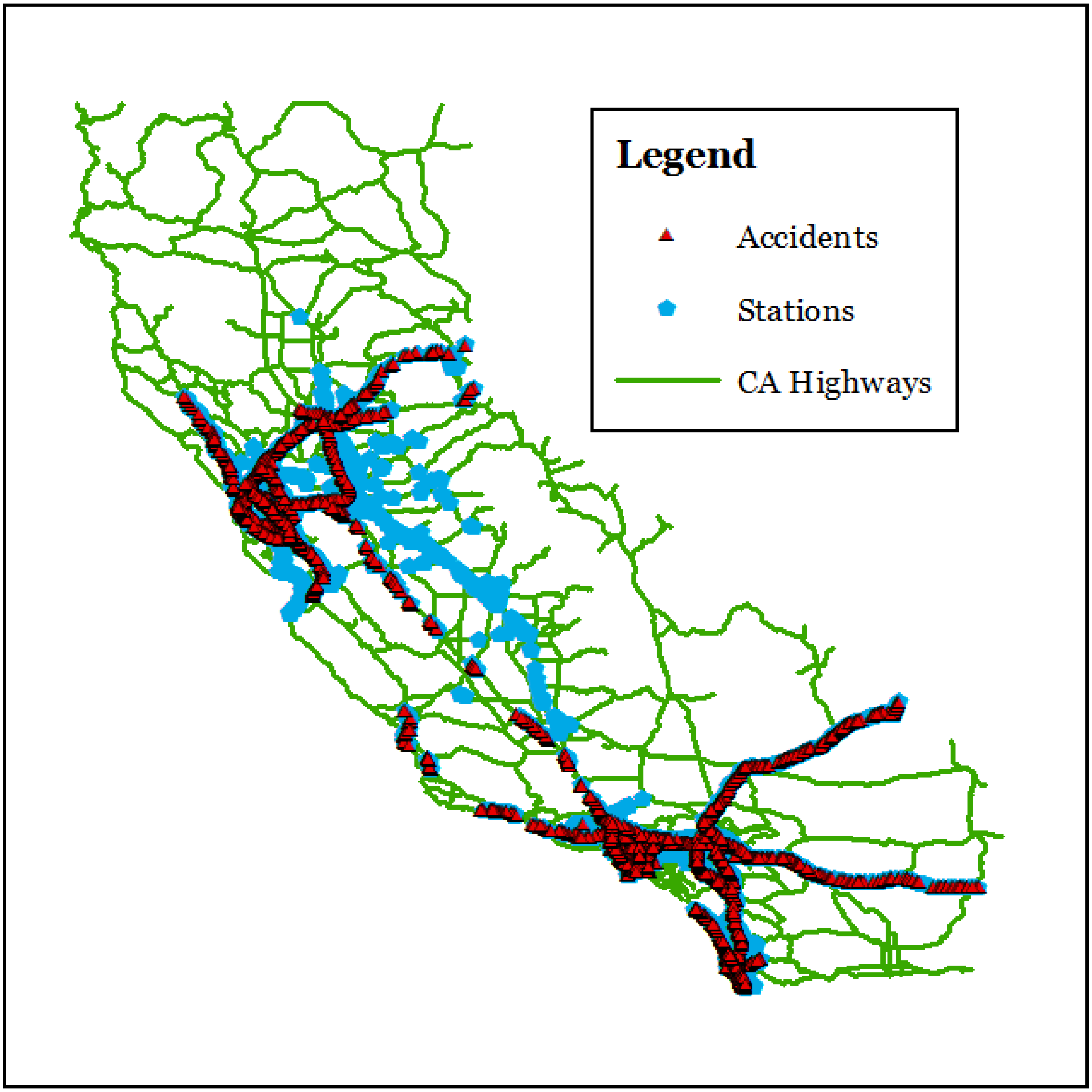
Variable	***n***	**Fatal accident ($n = 896$)**	**Non-fatal accident ($n = 1777$)**	***p*-value***
Vehicle manufacturer				**< 0.0001**
Domestic	1198	320 (27%)	878 (73%)	
European	225	61 (27%)	164 (73%)	
Japanese	1226	506 (41%)	720 (59%)	
Unknown	24	9 (38%)	15 (62%)	
Vehicle type				**< 0.0001**
Sedan	1208	416 (34%)	792 (66%)	
Light truck	803	187 (23%)	616 (77%)	
Motorcycle	281	261 (93%)	20 (7%)	
Medium/heavy truck	381	32 (8%)	349 (92%)	
Age of vehicle (years)				**< 0.0001**
< 4	548	163 (30%)	385 (70%)	
4–7	661	201 (30%)	460 (70%)	
8–11	621	201 (32%)	420 (68%)	
12–14	327	115 (35%)	212 (65%)	
≥ 15	509	211 (41%)	298 (59%)	
Unknown	7	5 (71%)	2 (29%)	
Trailing unit				**< 0.0001**
Yes	310	24 (8%)	286 (92%)	
No	2363	872 (37%)	1491 (63%)	

*Pearson's chi-square test of homogeneity

Table 2. Odds Ratios in the Final Model

Variable	OR (95% CI)	***p***
Driving speed		0.004
Slow	2.41 (1.39, 4.18)	
Fast	1.92 (1.08, 3.42)	
With traffic flow	1.00	
Vehicle type		< 0.0001
Light truck	0.62 (0.39, 0.99)	
Motorcycle	569.51 (25.85, $>$ 999.99)	
Medium/heavy truck	0.26 (0.11 0.64)	
Sedan	1.00	
Extent of vehicle damage		< 0.0001
Functional damage	0.007 (0.001,0.056)	
Other/minor damage	0.001 (0.001,0.010)	
Disabling damage	1.00	
Alcohol use		< 0.0001
Yes	1.35 (0.76, 2.40)	
Unknown	15.18 (7.69, 29.98)	
No	1.00	

Figure 1. FARS accidents, speed monitoring stations, and highways



References

[1] State of California Department of Motor Vehicles. California Driver Handbook - Laws and Rules of the Road. https://www.dmv.ca.gov/portal/dmv/detail/pubs/hdbk/speed_limits. Accessed January 3, 2017.

[2] Solomon DH. Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle. Washington, D.C.: U.S. Government Printing Office; 1964.

[3] Kloeden CN, Ponte G, McLean AJ. Travelling Speed and the Risk of Crash Involvement on Rural Roads. Canberra, A.C.T.: Australian Transport Safety Board; 2001.

[4] National Highway Traffic Safety Administration. Index of /fars/. ftp://ftp.nhtsa.dot.gov/fars/. Accessed January 3, 2017.

[5] Caltrans - California Department of Transportation. Caltrans Performance Measurement System. http://pems.dot.ca.gov/. Accessed January 3, 2017.

[6] Caltrans - California Department of Transportation. Caltrans GIS Data Library. http://www.dot.ca.gov/hq/tsip/gis/datalibrary/index.php. Accessed October 20, 2016.