Cyclist death and distracted driving: Important factors to consider.

SAS Student Symposium submission, SAS Global Forum 2016

Team disaster, number 33

Introduction

Pedalcyclists, riders of bicycles and non-motorized vehicles powered by pedals, account for 2% of the traffic related fatalities in the United States and represented 743 deaths in 2013.[1] While the rates of cycling in the US remain far lower than those in many other area including many eastern European countries and Australia[2,3] there is a current surge, particularly with certain subgroups. Compared to women, men are riding their bikes more, as well as people living in urban centers or near universities[4] and where the built environment is changing to support cycling[5]. The reasons for riding are also changing; the National Household Travel Surveys show a 21% increase in utilitarian cycling and a three-fold increase in cycling for the purpose of accessing public transit from 2001 to 2009[4].

Cycling has many broad benefits both for personal and public health including reducing over-weight and obesity, improving cardiovascular health and reducing all-cause mortality[6,7]. However, there risk of injury or death is the most commonly cited deterrent and outweighs the perceived benefits for many [8,9]. Other concerns include the condition of the roads and interactions with motor vehicles[9], both of which have been shown to influence the risk of bicyclist injury. Major streets with parked cars and no dedicated bike lane were the riskiest for cyclists, while the absence of parked cars or smaller, local streets proved protective [10]. Since 2010 there has been a 19% increase in the number cyclist fatalities in the United States.

Concurrently with the uptick in the numbers of cyclists on the road, particularly in urban areas, and the increase in the number of fatalities, there continues to be staggering rates of drivers who are using phones to talk and text. Ninety-eight percent of young drivers report texting and driving of some sort, with 70% initiating, 81% replying and 92% reading texts [11]. People perceive driving and texting to be very risky yet this perception has little influence on actual texting behavior [11]. In fact, there is evidence to suggest that texting while driving is becoming an automated behavior that occurs without awareness or control [12,13]. It is not surprising that these behaviors are contributing factors in traffic-related injury and death[14]. In 2008 alone, distracted drivers were at the wheel in 1 out of every 6 fatalities and 515,000 injuries [15], proving to be a major public health concern. Furthermore, Wilson and colleagues [16] showed a positive relationship with the percentages of distracted-related fatalities and the number of cell phone subscriptions and the average number of monthly text sent.

Are cyclists dying at the hands of distracted drivers? Yes, and in increasing numbers[17]. What are the factors that are uniquely associated with drivers who are distracted? In other words, what else is happening when distracted drivers fatally hit cyclists? This paper aims to elucidate the factors related with cycling fatalities when drivers are distracted. We use a novel machine learning approach, Adaptive Least Absolute Shrinkage and Selection Operator (aLASSO)[18,19], to select a parsimonious model and estimate the importance of the relevant factors.

Data

We used a national database on all vehicular fatalities occurring on public roads in the United States during 2013 to study which factors are associated with a cyclist fatality when a driver is distracted. The Fatality Analysis Reporting System (FARS) reports detailed information on every fatal injury in which a motor vehicle was involved in all 50 states, the District of Columbia and Puerto Rico. These data come from multiple sources including police reports, medical records, state registration files, and vital statistics.

For each accident in which there was at least one fatality, FARS reports concurring factors. The factors we used for this paper fell into 3 domains: (1) driver attributes, e.g. if they were impaired from alcohol or other drugs, (2) cyclists attributes, e.g. if they were following the laws of the road, and (3) environmental attributes, e.g. if the accident occurred at an intersection. Please see Table 1 for a list of variables included in this analysis.

We defined the driver as being distracted if they confirmed that they were using a mobile phone or other device to call or text, vehicle navigation system or "heads-up" display, or were inattentive due to talking, eating, reading or other activity.

Problem

Distracted driving is becoming more prevalent. At the same time the number of cyclists on the roads and cyclists fatalities are increasing. In order to guide policy and intervention efforts to decrease related mortality and morbidity while still promoting this healthy activity, we need to understand the factors related to cyclist deaths at the hands of distracted drivers.

The objectives of this study are to:

1. Obtain the proportion of cyclist fatalities where the driver was distracted,
2. Understand the relative frequencies of potential contributing factors when distracted drivers fatally hit cyclists, and
3. Model potential predictive factors using an innovative machine learning technique, called aLASSO to:
    a. Identify which factors are related (variable selection), and
    b. Estimate the magnitude of the important variables (coefficient estimation).

Data Cleaning and Validation

In 2013, there were 32,719 fatalities due to vehicular accidents in the United States. Of these, there were 743 fatalities of cyclists. Though the FARS reporting is extensive, 18% of these cases did not include information about the driver being distracted or not, thus our dataset consisted of N=611 pedalcyclists fatalities.

For the predicting variables, we excluded those that were missing in more than 10 percent (missingness of n=61) of the cases as to not significantly reduce our sample size for modeling or bias the results. There was no way of knowing if this information was missing due to factors related to the driver being distracted or not. We also omitted factors that were measuring the same concept and would likely be highly correlated with one another. For example, we chose one item representing the road conditions yet this information could have been recorded in more variables, e.g. wet road and atmospheric conditions - precipitation. Secondly, including variables that are highly correlated could produce unstable estimates.

We excluded variables that were seemingly unrelated to our problem of interest, such as information about additional passengers or damage caused to physical objects. We also excluded variables for which less than 3 of the responses were indicated as a "yes." For example, road rage was a factor in only 1 cyclist fatality thus these data are not powered to detect a relationship.

We investigated 18 variables as possible predicting factors related cyclist death when a driver is distracted. In all cases, these variables were presented as multiple categories. The majority (14) were made binary (0/1) for the purpose of analysis. A "1" indicated a positive response to the variable (e.g. the accident was at an intersection of any type) whereas a "0" indicated a negative response (e.g. absence of an intersection). The remaining 4 variables were collapsed into 3 or 4 relevant categories and coded into the respected number of indicator variables for analysis. We coded variables to missing if it was indicated as "not reported" or "unknown."

The data was extracted from 8 separate files and merged according to the accident number.

<u>Analysis</u>

The purpose of this analysis is to determine which factors out of the 18 possible variables are related to our outcome, and understand that magnitude in which these effect cyclist death when a driver is distracted. We used Adaptive Least Absolute Shrinkage and Selection Operator (aLASSO)[18,19] for variable reduction as well as model estimation. aLASSO is a modified version of LASSO that uses weights to penalize different coefficients. The weights are data-dependent and adaptively chosen from the data. Large coefficients receive small weights (penalties) where as small coefficients receive larger weights. These weights converge to a finite constant as the sample size increases. This reduces bias and retains important predictors in the model. Small coefficients receive large weights and as the sample size increases these weights become inflated (to ∞). aLASSO increases sparsity and thus makes the model simpler. Computationally, aLASSO can be solved using the same algorithm as LASSO[20] making it easy to implement. It also enjoys the oracle properties, which is not true for LASSO[19].

aLASSO requires a tuning parameter that we estimated using 10-fold external cross validation. This method uses the penalized model that is fit 10 times instead of just once by using sub-sampling out 10 different training samples. The statistic found is not based on ordinary least squares regression like the statistic found using cross validation but is based on the penalized regression. We generated the analysis for this paper using SAS Studio software, Version 3.4[21]. In particular, we constructed the aLASSO model using the GLMSELECT procedure.

<u>Results and Generalizations</u>

Of the N=611 cyclist fatalities in 2013, 16.7% involved a driver who was distracted. As for the drivers, 2.0% failed to obey a traffic sign, 8.9% were speeding, and over 8% had either been drinking alcohol at the time of the crash and/or were impaired by drugs or alcohol. Sixteen point three percent had previous speeding violations and 0.5% had a previous drunk driving record.

At the time of the accident, 69.1% of cyclists performed an improper action. Over 40% of the accidents occurred at an intersection and over 50% were during the light of day. Of these accidents, 47% occurred where there was a junction in the road and 2.2% happened with the cyclist in a bike lane. The road was dry for almost 92% of the fatalities.

Males were disproportionately involved in cyclist fatalities: 73.4% of drivers and 86.8% of the cyclists were male. (Table 1).

The aLASSO approach iteratively determined the best fit model, please see Table 2 for details. Three of the 18 predictor variables were retained in the model. These are:

1. Improper action by the cyclist (Improper_action),
2. Failure to obey traffic signs by the driver (Failure), and
3. The accident being speed related (Speed_related)

It's worthy to note the first variable left out, Intersection, is on the edge of significance (p=0.0514) but the statistic found using the tuning method is slightly higher than that of the optimal model, 0.1280 versus 0.1277. Please see Figure 1 for the cutoff point for the model at the third variable. This is where the smallest external cross validation statistic is found producing the best model. Note the closeness in the value of the external cross validation statistic between steps 2 and 7. Any of these would appear to be a reasonable choice. The tuning methods AIC and BIC find the optimal model to be at step 4. (Figure 2.)

When there is a cycling fatality and the cyclist performed an improper action, it is less likely that the driver was distracted ($\beta = -0.167$, $p=0.0009$). However, if the driver was engaged in a speed-related action, or if they failed to obey a traffic sign, the likelihood that they were distracted increases ($\beta = 0.050$, $p=0.006$ and $\beta = 0.130$, $p<0.0001$, respectively). (Table 3).

<u>Future Studies</u>

This study illuminated important factors related to cyclist death when a driver is distracted. However, these types of observational data can only report relationships and cannot speak to causation. We cannot say that distraction, speed, or any other factors did or did not cause a death. Other prospective study designs could speak to causality and could better explain these relationships.

The reporting of driver distraction poses a limitation on these results. It is possible that there is social pressure to declare that the driver was not distracted when they might have been. The numbers of actual distracted drivers is likely underestimated due to the ability of law enforcement to identify the type of distraction and it's role in the crash[15]. Future studies could focus on improving the reporting of distractedness of drivers to better estimate this effect.

The FARS data reports 1,063 fewer fatalities overall in 2013 compared to 2012[22] but yet cyclist fatalities are increasing[1]. Stimpson and colleagues found that the cyclists who died from distracted drivers were more likely to be non-Hispanic, White males[17]. Future studies could ask what demographic characteristics might be having an effect on the increase of cycling fatalities.

<u>Conclusion</u>

Cyclist death due to distracted drivers is a serious public health problem. Not only are the consequences immediate due to injury, but the threat or fear of injury can deter people from riding their bikes. Yet cycling is protective for a myriad of health outcomes including all-cause morbidity and mortality[6,7]. To help protect cyclists and to inform policy and intervention, the factors that influence the likelihood of cyclist death when a driver is distracted are important.

We can conclude that if a cyclist makes an improper action at or just before the time of the crash, the likelihood of the driver of the vehicle being distracted decreases, implying that when the bike rider is at fault, it is less likely that the driver was also at fault due to distracted. At the same time, if the driver is speeding or has failed to obey a traffic sign and fatally hits a cyclist, the likelihood of them also being distracted increases. Being distracted is related to other risky driving practices when cyclists are fatally injured.

None of the environmental factors, such as the road conditions and light conditions had a significant impact on this relationship. Although both the drivers and the cyclists were more likely to be male, gender and age were not important factors related to fatalities with these data. The driver's history of speeding or impairment were also not significant suggesting that it is the current actions of the distracted driver, or the cyclist when the driver is not distracted.

Appendices

References

1. *Bicyclists and other cyclists: 2013 data.* Washington, DC: National Highway Traffic Safety Administration;2015.
2. Bassett Jr DR, Pucher J, Buehler R, Thompson DL, Crouter SE. Walking, cycling, and obesity rates in Europe, North America, and Australia. *J Phys Act Health.* 2008;5(6):795-814.
3. Buehler R, Pucher J. Big City Cycling in Europe, North America, and Australia. *Pucher, J, Buehler, R. eds.* 2012:287-318.
4. Pucher J, Buehler R, Seinen M. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice.* 7// 2011;45(6):451-475.
5. Pucher J, Dill J, Handy S. Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive medicine.* 2010;50:S106-S125.
6. Oja P, Titze S, Bauman A, et al. Health benefits of cycling: a systematic review. *Scandinavian journal of medicine & science in sports.* 2011;21(4):496-509.
7. Hamer M, Chida Y. Active commuting and cardiovascular risk: a meta-analytic review. *Preventive medicine.* 2008;46(1):9-13.
8. De Hartog JJ, Boogaard H, Nijland H, Hoek G. Do the health benefits of cycling outweigh the risks? *Environmental health perspectives.* 2010:1109-1116.
9. Winters M, Davidson G, Kao D, Teschke K. Motivators and deterrents of bicycling: comparing influences on decisions to ride. *Transportation.* 2011;38(1):153-168.
10. Teschke K, Harris MA, Reynolds CC, et al. Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. *American journal of public health.* 2012;102(12):2336-2343.
11. Atchley P, Atwood S, Boulton A. The choice to text and drive in younger drivers: Behavior may shape attitude. *Accident Analysis & Prevention.* 2011;43(1):134-142.
12. Bayer JB, Campbell SW. Texting while driving on automatic: Considering the frequency-independent side of habit. *Computers in Human Behavior.* 2012;28(6):2083-2090.
13. Oulasvirta A, Rattenbury T, Ma L, Raita E. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing.* 2011;16(1):105-114.
14. Organization WH. Mobile phone use: a growing problem of driver distraction. 2011.
15. Ascone D, Lindsey T, Varghese C. *An examination of driver distraction as recorded in NHTSA databases.* 2009.
16. Wilson FA, Stimpson JP. Trends in fatalities from distracted driving in the United States, 1999 to 2008. *American journal of public health.* 2010;100(11):2213-2219.
17. Stimpson JP, Wilson FA, Muelleman RL. Fatalities of pedestrians, bicycle riders, and motorists due to distracted driving motor vehicle crashes in the US, 2005–2010. *Public Health Reports.* 2013;128(6):436.
18. Zhang HH, Lu W. Adaptive Lasso for Cox's proportional hazards model. *Biometrika.* 2007;94(3):691-703.
19. Zou H. The adaptive lasso and its oracle properties. *Journal of the American statistical association.* 2006;101(476):1418-1429.
20. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological).* 1996:267-288.

21. *SAS Studio for Windows* [computer program]. Version Resease 3.4. Cary, NC: SAS Institute; 2015.
22. Administration NHTS. US Department of Transportation. 2013 Motor Vehicle Crashes: Overview. *Traffic Safety Facts.* http://www-nrd*. nhtsa. dot. gov/Pubs/811856. pdf.* 2014.

Appendices

Table 1: Description of predictor variables and relevant frequencies

	Variable name	Description (1=yes, 0=no)	Percent (n)
Outcome			
	Driver_distracted	Driver distracted	16.7% (102/611)
Driver attributes			
	Failure	Failed to obey traffic signs	2.0% (12/594)
	Speed_related	Any speeding, racing, driving too fast for conditions	8.9% (53/599)
	Passing	Driver passed where prohibited	1.4% (8/594)
	Previous_DWI	Previous DWI violations	0.5% (3/583)
	Driver_gender	Gender (male)	73.4% (430/586)
	Driver_drinking	Drinking alcohol at time of crash	8.4% (51/611)
	Driver_impaired	Impaired from drugs/alcohol	8.9% (51/573)
	Prev_speeding_violat	Previous speeding violations	16.3% (95/583)
	Driver_age	<20 21-64 >65	12.3% (75/611) 77.1% (471/611) 10.6% (65/611)
Cyclist attributes			
	Improper_action	Performs improper action	69.1% (400/579)
	Cyclist_gender	Gender (male)	86.8% (527/607)
	Cyclist_age	<10 10-20 21-64 >65	3.1% (19/607) 14.0% (85/607) 68.4% (415/607) 14.5% (88/607)
Environment attributes			
	Intersection	Accident occurred in or near intersection	40.1% (245/611)
	Light_condition	Light Dark Dawn/dusk	52.5% (320/610) 42.3% (258/610) 5.2% (32/610)
	Clear	Weather is clear	79.1% (481/608)
	Road_dry	Road is dry	91.8% (561/611)
	Location	Intersection Non-intersection in roadway Bike lane Other	34.6% (208/601) 57.4% (345/601) 2.2% (13/601) 5.8% (35/601)
	Road_junction	Road junction	47.0% (287/611)

Table 2: aLASSO model selection summary

Step	Effect Entered	Effect Removed	Number Effects In	Model R-Square	Adjusted R-Square	AIC	AICC	BIC	CP	SBC	ASE	CVEX PRESS	F Value	Pr > F
0	Intercept		1	0.0000	0.0000	-511.5498	-511.5264	-1026.6953	39.6310	-1024.3056	0.1352	0.1356	0.00	1.0000
1	Improper_action		2	0.0215	0.0196	-520.7284	-520.6815	-1035.9236	29.7648	-1029.2401	0.1323	0.1330	11.26	0.0009
2	Failure		3	0.0685	0.0648	-544.0767	-543.9982	-1059.0743	5.7929	-1048.3442	0.1259	0.1281	25.83	<.0001
3	Speed_related		4	0.0817	0.0763	-549.4296	-549.3118	-1064.3118	0.4951	-1049.4530*	0.1241	0.1277*	7.35	0.0069
4	Intersection		5	0.0885	0.0813	-551.2671*	-551.1017*	-1066.0443*	-1.2724*	-1047.0462	0.1232	0.1280	3.81	0.0514
5	Passing		6	0.0888	0.0798	-549.4324	-549.2115	-1064.1610	0.5659	-1040.9674	0.1232	0.1281	0.16	0.6862
6	Previous_DWI		7	0.0894	0.0786	-547.7558	-547.4712	-1062.4300	2.2498	-1035.0466	0.1231	0.1281	0.32	0.5724
7	Light_condition_1		8	0.0901	0.0776	-546.1977	-545.8413	-1060.8112	3.8182	-1029.2444	0.1230	0.1282	0.44	0.5097
8	Driver_gender		9	0.0948	0.0804	-546.8211	-546.3846	-1061.2926	3.2634	-1025.6236	0.1224	0.1288	2.58	0.1086
9	Driver_drinking		10	0.1000	0.0839*	-547.7947	-547.2698	-1062.0888	2.3832	-1022.3530	0.1217	0.1300	2.92	0.0879
10	Location_2		11	0.1008	0.0830	-546.2744	-545.6529	-1060.4766	3.9201	-1016.5886	0.1216	0.1305	0.47	0.4935
11	Cyclist_gender		12	0.1016	0.0820	-544.7283	-544.0018	-1058.8360	5.4823	-1010.7983	0.1215	0.1309	0.44	0.5057
12	Driver_impaired		13	0.1026	0.0811	-543.2819	-542.4419	-1057.2861	6.9489	-1005.1078	0.1213	0.1313	0.54	0.4628
13	Clear		14	0.1026	0.0793	-541.2983	-540.3364	-1055.2248	8.9331	-998.8800	0.1213	0.1313	0.02	0.8996
14	Road_dry		15	0.1037	0.0786	-539.9208	-538.8285	-1053.7329	10.3340	-993.2583	0.1212	0.1318	0.60	0.4371
15	Location_1		16	0.1050	0.0781	-538.6867	-537.4553	-1052.3701	11.5980	-987.7800	0.1210	0.1324	0.74	0.3893
16	Location_3		17	0.1057	0.0770	-537.0803	-535.7013	-1050.6543	13.2200	-981.9295	0.1209	0.1327	0.38	0.5375
17	Cyclist_age		18	0.1071	0.0765	-535.8671	-534.3317	-1049.2999	14.4656	-976.4721	0.1207	0.1333	0.76	0.3838
18	Prev_speeding_violat		19	0.1075	0.0751	-534.0957	-532.3953	-1047.4237	16.2466	-970.4566	0.1207	0.1335	0.22	0.6390
19	Driver_age		20	0.1076	0.0734	-532.1996	-530.3253	-1045.4310	18.1471	-964.3162	0.1206	0.1336	0.10	0.7522
20	Road_junction		21	0.1079	0.0718	-530.3394	-528.2825	-1043.4703	20.0133	-958.2119	0.1206	0.1336	0.13	0.7144
21	Light_condition_2		22	0.1079	0.0699	-528.3532	-526.1048	-1041.3937	22.0000	-951.9816	0.1206	0.1337	0.01	0.9083

Table 3: Optimal model parameter estimates, based on external cross-validation

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5.68646	1.89549	15.15	<.0001
Error	511	63.93684	0.12512		
Corrected Total	514	69.62330			

Parameter Estimates			
Parameter	DF	Estimate	Standardized Estimate
Intercept	1	0.246338	0
Improper_action	1	-0.136101	-0.167172
Speed_related	1	0.065437	0.050257
Failure	1	0.364027	0.129733

Figure 1: Coefficient progression during the model selection proves using Adaptive Least Absolute Shrinkage and Selection Operator (aLASSO)

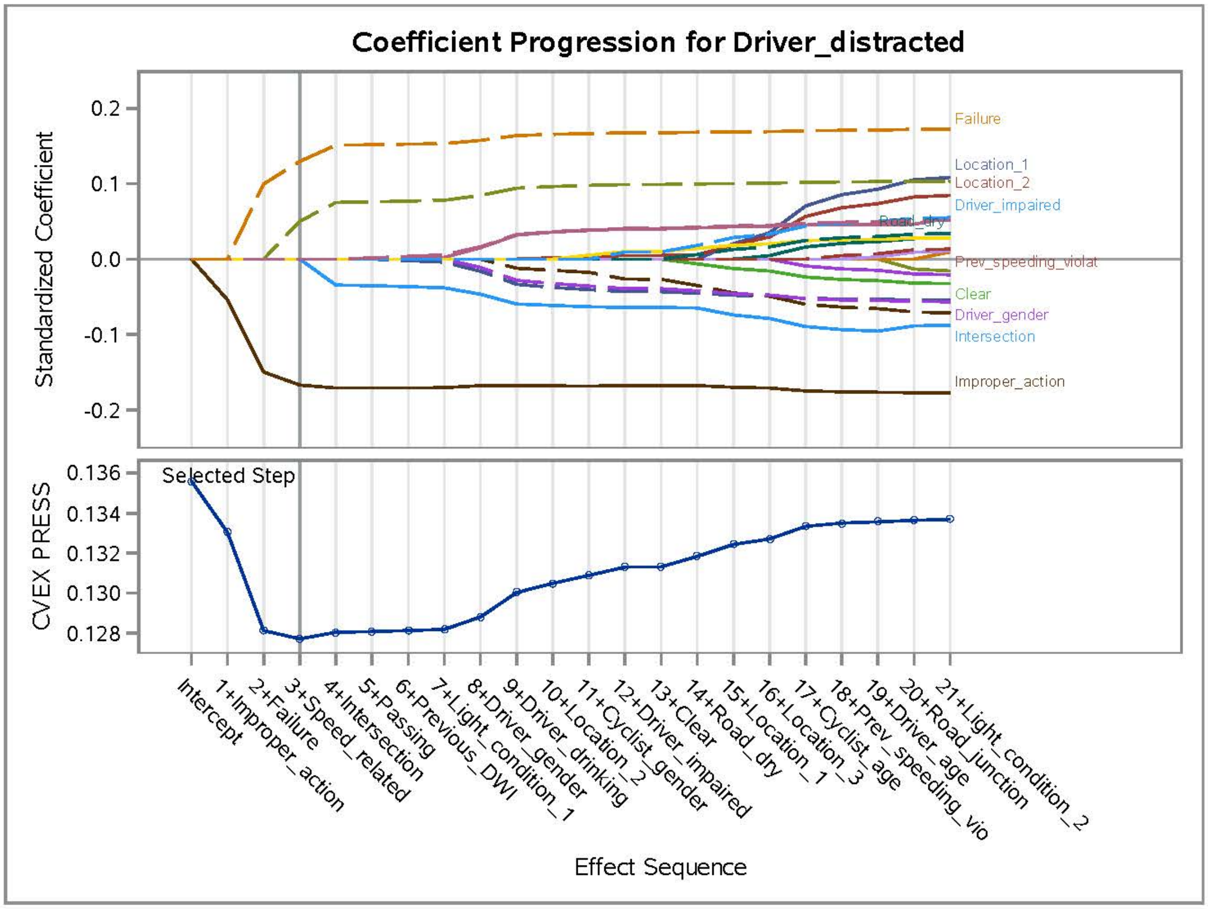

Figure 2: Model fit criteria